



www.openaire.eu



Aija Uzula

Nacionālais atvērtās piekļuves dienests, LU Bibliotēka

Datu pārvaldības plāni (DMP) ievads

”

*Mērķis bez plāna ir
tikai vēlēšanās.”*

- *Antuāns de Sent-Ekziperī*
(1900-1944)

Prezentācijā

- Kas ir DMP?
- Kāpēc DMP nepieciešams?
 - Ko ietver DMP?
 - DMP rīku piemēri

When it comes to data ...



Neefektīva datu pārvaldība

Kas ir datu pārvaldības plāns (DMP)?

*Tas ir instruments, kas palīdz pētniekiem strādāt ar datiem **organizētā** veidā, **pārvaldot** tos, daloties ar tiem un arhivējot datus **efektīvā** veidā.*

DMP palīdz:

- **ietaupīt** laiku un citus resursus;
- strādāt **efektīvāk** ar sadarbības partneriem;
- **nodrošināt** datu atkārtotu izmantošanu;
- **ievērot** finansētāju prasības;
- **nezaudēt** datus;
- vienkāršāk **sameklēt** vajadzīgos datus;
- vienkāršāk **analizēt** organizētus un dokumentētus datus;
- **izvairīties** no datu viltošanas iespējām;
- saņemt **atsauksmes** par datiem.

DMP apraksta:

- **Ko?** projekta gaitā radušos datus.
- **Kā?** dati tiks organizēti, dokumentēti, standartizēti, uzglabāti, aizsargāti, arhivēti un nodrošināta iespēja dalīties ar tiem.
- **Kurš?** Ir atbildīgs par aktivitātēm, kas iekļautas DMP.
- **Kad?** aktivitātes notiks.

DMP galvenie pieturpunkti:

- datu vākšana / atkārtota datu izmantošana;
- datu dokumentēšana un metadatu izveide;
- ētiskie un juridiskie jautājumi;
- datu saglabāšana un dublēšana (*back up*);
- datu izvērtēšana un ilgtermiņa glabāšana;
- dalīšanās ar datiem;
- atbildības un citi resursi.



DMP ietver atbildes uz jautājumiem:

- Kādi dati tiks radīti konkrētā projekta gaitā?
- Kurš būs atbildīgs par datiem katrā projekta posmā?
- Kādi datu formāti tiks izmantoti?
- Kā datus organizēt un nosaukt?
- Kādu informāciju par datiem nepieciešams ietvert, lai tie būtu saprotami citiem ? (*metadati*)
- Kur datus glabāt?
- Kuram nepieciešams piešķirt atļauju piekļūt datiem?
- Kādā veidā dati tiks publiskoti vai arhivēti projekta noslēgumā?

Datu vākšana / atkārtota datu izmantošana

Datu tipi:

- **Novērojumu** (*sensorie, aptaujas dati, paraugu dati, attēli*);
- **Eksperimentālie** (*gēnu paraugi, hromatogrammas, magnētiskā lauka dati*);
- **Simulācijas** (*klimata modeļi, ekonomikas modeļi*);
- **Atvasinātie vai kompilētie** (*teksta un datu ieguve, 3D modeļi, dati, kas apkopoti no publiskie dokumentiem*);
- **Atsauču** (*gēnu sekvences datubankas, ķīmijas struktūras, skaitīšanas dati, telpisko datu portāli*).

Datu vākšana / atkārtota datu izmantošana

*Datiem iedala tekstuālus, skaitliskus, multimediju, moduļu, programmatūras valodas, nozarei specifiskus un rīkiem specifiskus **formātus**.*

Dokumenti (Text, PDF, Word)
Izklājlapas (Excel)
Mājaslapas
Piezīmes / dienasgrāmatas
Datubāzes (Access, MySQL, Oracle)
Aptaujas, transkripcijas, kodi
Audio un video ieraksti
Filmas, fotogrāfijas
Artefakti, diapozitīvi, paraugi
Digitālo objektu kopums, kas iegūts pētniecības gaitā
Neapstrādātie dati (raw data), ko ģenerē programmatūras, sensori vai rīki
Modeļi, algoritmi
Datorprogrammas (ievades, izvades, pieslēgšanās faili programmatūras analīzei, programmatūras simulācija, shēmas)

Datu vākšana / atkārtota datu izmantošana

Datu tipi	Ieteicamie formāti	Pieņemamie formāti
Tabulārie (tabulas) dati ar apjomīgiem metadatiem <i>Dažādi apzīmējumi, kodu apzīmējumi, definētas trūkstošas vērtības</i>	SPSS pārnēsājamais formāts (.por) Norobežots teksts un komandas ('setup') fails (SPSS, Stata, SAS, u.c.) Strukturēts teksts vai atzīmēts (<i>hashtag</i> jeb tēmturis) fails metadata informācijā, piemēram, DDI XML file	Statistisko pakotņu patentētie (proprietary) formāti: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabulārie dati ar minimāliem metadatiem <i>Kolonnu nosaukumi, dažādi vārdi</i>	Ar komatu atdalītas vērtības (.csv) Tab-norobežots fails (.tab) Norobežots teksts ar SQL datu definīcijas paziņojumiem	Norobežots teksts (.txt) ar apzīmējumiem, kas nav atrodamī dati kā norobežotāji Plaši izmantoti formāti: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geotelpiskie dati <i>Vektoru un rastru dati</i>	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-references TIFF (.tif, .tfw) CAD dati (.dwg) tabulārie GIS atribūtdati Geography Markup Language (.gml)	ESRI geo-datubāzes formāts (.mdb) MapInfo Interchange Format (.mif) vektoru datiem Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) Binārie formāti GIS un CAD pakotnes

Datu vākšana / atkārtota datu izmantošana

Datu tipi	Ieteicamie formāti	Pieņemamie formāti
Tekstuālie dati	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) atbilstoši DTD vai shēmai	Hypertext Mark-up Language (.html) Bieži izmantoti formāti: MS Word (.doc/.docx) Daži uz programmatūru balstīti formāti: NUD*IST, NVivo un ATLAS.ti
Attēlu dati	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) ja oriģināls ir radīts šajā pašā formātā GIF (.gif); TIFF citas versijas (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd); BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio dati	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) ja oriģināls ir radīts šajā pašā formātā Audio Interchange File Format (.aif) Waveform Audio Format (.wav)

Datu vākšana / atkārtota datu izmantošana

Datu tipi	Ieteicamie formāti	Pieņemamie formāti
Video dati	MPEG-4 (.mp4) OGG video (.ogv, .ogg) kustības JPEG 2000 (.mj2)	AVCHD video (.avchd)
Dokumentēšana un skripti	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) bieži izmantoti formāti: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) atbilstoši DTD vai shēmai, e.g. XHTML 1.0

Avots: <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>

Datu vākšana / atkārtota datu izmantošana

Datu failu nosaukumi:

Piemērs:

- 20181120_BIOProjekts_Idejas_BiteOtto_v001.doc
- 20181127_BIOProjekts_Apraksts_BiteOtto_v001.doc
- 20181214_BIOProjekts_Dizains_BiteOtto_v001.pdf
- 20181214_BIOProjekts_Dizains_BiteOtto_v002.pdf
- 20181215_BIOProjekts_Dizains_BiteOtto_v003_gala_versija.pdf
- 20190201_BIOProjekts_Apt_rezult_BiteOtto_v001.xlsx

Datu vākšana / atkārtota datu izmantošana

Nebūtu vēlams dalīties ar:

- provizoriskiem analīžu datiem;
- zinātnisko darbu uzmetumiem;
- nākotnes pētniecības plāniem;
- komunikāciju ar kolēģiem recenzijas tapšanas gaitā;
- fiziskiem paraugiem.

Datu vākšana / atkārtota datu izmantošana

Ko darīt, ja (digitālu) datu nav?

- Tāpat ieteicams veidot DMP.
- DMP obligāti nepieciešams to pateikt un skaidrot.

Ko darīt, ja atkārtoti izmanto citu radītus datus?

- DMP obligāti nepieciešams skaidrot to izcelsmi un lietderību pētījumā.

Datu dokumentēšana un metadatu izveide

Metodes:

- *Kā dati vākti?*
- *Kā datus interpretēt?*
- *Kādi ierobežojumi izmantoti?*

Ieteicams dokumentēt: kodu; aptauju; kodu grāmatu; ontoloģijas; jebko, kas ļauj citam reproducēt rezultātus.

<https://bartoc.org/>

Metadati:

- *Kas tas ir?*
- *Kas to izveidoja un kad?*
- *Kā tas šeit nokļuva?*
- *Ko tas nozīmē?*
- *Ko ar to var darīt?*

Izpēti metadatu shēmu, pirms vāc datus; **Konsultējies** ar nozares repozitorija pārstāvi un/vai speciālistu datu pārvaldības jautājumos.

Datu dokumentēšana un metadatu izveide

Kādus standartus es izmantošu, lai dokumentētu savus datus?

- <https://rdamsc.dcc.ac.uk/> (*metadatu standartu katalogs*)
- <http://rd-alliance.github.io/metadata-directory/> (*metadatu standartu katalogs*)
- <http://www.dcc.ac.uk/resources/metadata-standards> (*atbilstoši nozarei*)
- <https://dom.lndb.lv/data/obj/file/162615> (*par Dublin Core standartu*)

Datu dokumentēšana un metadatu izveide

Metadata Example: Dublin Core

- Contributor
 - Jane Collaborator
- Creator
 - Kristin Briney
- Date
 - 2013 Apr 15
- Description
 - A microscopy image of cancerous breast tissues under 20x zoom. This image is my control, so it has only the standard staining describe on 2013 Feb 2 in my notebook.
- Format
 - JPEG
- Identifier
 - IMG00057.jpg
- Relation
 - Same sample as images IMG00056.jpg and IMG00055.jpg
- Subject
 - Breast cancer
- Title
 - Cancerous breast tissue control

Ētiskie un juridiskie jautājumi

Datu drošība:

- **Datu veids:** vai **sensitīvi**? (*veselības, institūciju u.c. informācija*)
Datu anonimizēšanas iespējas
- **Drošības risku aspekti:** *drošs digitālais arhīvs, kontrolēta piekļuve, personas informācijas identificēšana, drošības apmācība.*
- **Drošības jautājumi:** *Kuram ir piekļuve? Vai vispār jādalās? Vai tas ir iespējams? Kurš pārbaudīs, vai dati ir drošībā?*
- Vebināru ieraksti par juridiskajiem aspektiem pētniecības datu pārvaldībā: <https://www.napd.lu.lv/par-mums/zinas/zina/t/59124/> (2020. g.)



Datu saglabāšana un dublēšana

- **Datu dublēšana** ir datu rezerves kopija.
- Datu nezūdamības aspekts.

Kāpēc nevar saglabāt visus datus?

- visu datu saglabāšana ir dārga un darbietilpīga;
- liela apjoma lieko datu glabāšana apgrūtina noderīgu datu atrašanu.

Datu saglabāšana un dublēšana

Saglabājot apsver:

- kurus datus finansētājs vai institūcija pieprasa saglabāt;
- vai dati ir nozīmīgi projektam vai institūcijai;
- vai pastāv juridiskas un intelektuālā īpašuma tiesības glabāt un atkārtoti izmantot šos datus;
- vai ir pietiekama dokumentācija datu paskaidrošanai un atrašanai;
- vai var atļauties maksāt par datu glabāšanu.
- **Pārlicinies**, vai izvēlētais datu pakalpojuma sniedzējs nodrošina datu dublēšanu!

Datu izvērtēšana un ilgtermiņa glabāšana

- **Ievēro** nosacījumus (piemēram, finansētāju prasības) un dokumentē visus pieņemtos lēmumus un to iemeslu!
- **Centies ievērot FAIR datu principus!**
- **Liela apjoma datiem** – atbilstošas infrastruktūras izvēle.
- **Apsver glabāšanas izmaksas!**
- Atvērts failu formāts nodrošina atbilstošu ilgtermiņa glabāšanu.
- **Glabāšanas riski:** *nolietošanās; novecošana; nepārdomāta datu glabāšanas vieta*
- www.re3data.org (*Datu repozitoriju reģistrs*)
- <https://www.coretrustseal.org/> (*Sertificētu datu repozitoriju saraksts*)

Dalīšanās ar datiem

- vadmotīvs **“tik atvērti, cik iespējams; tik slēgti, cik nepieciešams”**
(as open as possible, as closed as necessary)
- **Dati ir zinātnisks resurss** (dalīšanās ar datiem ir **līdzvērtīga** zinātniskajai publicēšanai).
- **Svarīga ir atbilstošas licences izvēle** (piemēram, **Creative Commons** <https://creativecommons.org/share-your-work/licensing-examples/> vai **EUDAT** izstrādātais licencēšanas rīks <https://ufal.github.io/public-license-selector/>)

Dalīšanās ar datiem

- Lielāka **rezonanse** pētniecības darbam (arī **citējamība**), papildu **publicitāte** vēl nepublicētiem rezultātiem.
- **Apmierināts** finansētājs;
- Palīdz **atrast** un **atkārtoti** izmantot savus datus **vēlāk**; apliecina, ka dati nav viltoti; stimulē jaunus pētniecības virzienus.
- Ieteicams izmantot **plaši pieejamus** dalīšanās veidus: **zinātnisks žurnāls**, nozares **repozitorijs**.

Atceries!

- **Aktuālās DMP prasības variē atbilstoši finansētājam.**
- **Obligāti pārbaudi DMP prasības pirms DMP veidošanas!**
- **Projektā nepieciešams paredzēt finansējumu DMP izstrādei un izpildei.**
- **Veic regulārus DMP atjauninājumus!**

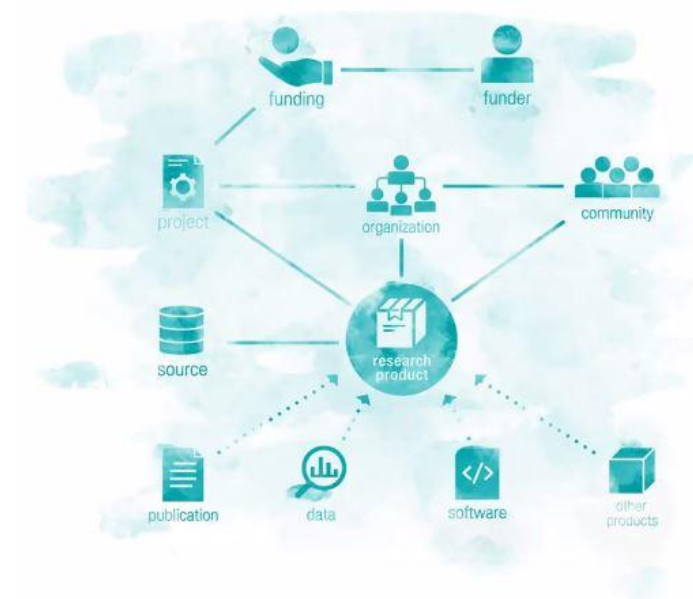
DMP rīki

- Kalifornijas Universitātes pārvaldības centra (UC3) rīks
<https://dmptool.org/>
- Digitālās pārvaldības centra (Apvienotā Karaliste) rīks
<https://dmponline.dcc.ac.uk/>
- Organizācijas OpenAIRE rīks
<https://argos.openaire.eu/splash/>



ARGOS

- **Tiešsaistē** izveido mašīnlasāmu DMP nozarei atbilstošā veidnē;
- iespējama DMP **koprakstīšana** ar sadarbības partneriem;
- iespēja **publicēt** DMP versijas dažādos formātos *Zenodo*;
- Veido datu kopu aprakstus institūciju veidnēs!
- Papildu publicitāte / citējamība;
- Sasaiste ar EOSC.



Interesanti!



**Katra plānošanas minūte
projekta sākumposmā ietaupa
10 minūtes pēc tam**

Noderīgas saites:

- https://dmponline.dcc.ac.uk/files/DMP_Checklist_2013.pdf (*DMP kontROLSARAKSTS*)
- <http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples> (*DCC DMP piemēri*)
- https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm (*H2020 DMP vadlīnijas*)
- <https://www.CESSDA.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide> (*CESSDA DMP vadlīnijas*)

The background features a large white circle in the center, surrounded by dark blue and light blue curved shapes. The light blue areas are decorated with a pattern of small white diamonds.

**Paldies par
uzmanību!**