

# Andmehaldusplaan

Teadusuuringu kavandamise käigus on oluline läbi mõelda ja kirja panna, kuidas toimub andmete kogumine ja nendega töötamine uuringu vältel, kellel on ligipääs andmetele ja vastutus andmete eest, mis saab andmetest pärast projekti lõppemist jne. Selleks tuleb luua andmehaldusplaan ning edasise töö käigus seda järgida.

---

## DATA COLLECTION

### ANDMETE KOGUMINE

#### Andmekirjeldus

- millised andmetüübid (eksperiment, vaatlus, küsitlus, audio-video jne)
- kuidas uued andmed integreeruvad juba olemasolevate andmetega
- millised andmed väärivad pikaajalist säilitamist
- kas rakendub autoriõigus või intellektuaalomandi õigus (luba nende andmete kasutamiseks)

#### Andmevormingud

- milliseid andmevormingud kasutatakse (soovitus kasutada avatud, standardseid ja masinloetavaid vorminguid )
- kas vorming võimaldab automaatset metaandmete lisamist
- kas repositooriumid toetavad valitud vorminguid
- Soovitavad andmevormingud
  - [File Formats. Open Data Handbook](#)
  - [File Formats. Data Archiving and Networked Services](#)

#### Andmemaht

- Hinda andmemahtu projekti lõpus. Sellest sõltuvad paljud järgmised otsustused ja kulud andmehaldusele
  - säilitamine
  - ligipääs
  - varundamine
  - andmevahetus
  - riist- ja tarkvara
  - tehniline tugi

#### Tarkvara

- võimalusel kasuta koostöövõimelist avatud lähtekoodiga tarkvara, see hoiab riist- ja tarkvara kulud madalad
- eelista tarkvara, mida arendab ja toetab suur kogukond (kõrgem kvaliteet, turvalisus ja kaasajastamine; paraku ka vähene dokumentatsioon ja tugi)
- tarkvara peab võimaldama kõiki analüüse uuesti läbi viia
- kui tarkvara luuakse ise, siis dokumentatsioon, tehniline tugi ja jätkusuutlikus
- versioonihaldussüsteem [git](#)
- pilvepõhine koodirepositoorium [GitHub](#)

## Kuidas andmed saadakse

- kogun ise
- (taas)kasutan oma varem kogutud andmeid
- kasutan avalikke avaandmeid ([Eesti avaandmete portaal](#))
- (taas)kasutan teiste poolt kogutud andmeid, vt repositooriumide register [Re3data](#), andmeregistrid ([Mendeley Data](#), [DataCite Search](#) jm)
- ostan andmed
- mida tähele panna?
  - kui taaskasutatakse või ostetakse andmeid, siis millist versiooni?
  - mida tehakse, kui andmete autor laeb üles uue versiooni?
  - säilita kasutatav versioon ja selle dokumentatsioon ka oma serveris.
  - kontrolli autoriõigusi, litsentse, piiranguid (ligipääs, taaskasutus)
  - kontrolli andmete masinloetavust ja koostöövõimet planeeritava infosüsteemiga

## Andmete organiseerimine

- failide nimetamine: lihtne, loogiline, lühenditeta või standardsete lühenditega (riigid, keeled, mõõtühikud, meetodid)
- lühendid läbivalt ühes keeles
- failide organiseerimine (variandid: projekti nimi, aeg, koht, koguja, materjali tüüp, vorming, versioon)
- kaustade struktuur olgu hierarhiline, lihtne, loogiline, lühike
- failide kopeerimine mitmesse kohta ei ole hea, hoia ühes kohas, loo otseteed
- versioonihaldussüsteem [git](#)
- pilvepõhine koodirepositoorium [GitHub](#)
- metaandmete lisamine (kes vastutab, millal lisatakse)
- Rohkem infot [Data Organization in Spreadsheets](#)

---

## DOCUMENTATION AND METADATA

### ANDMETE DOKUMENTEERIMINE JA METAANDMED

#### Andmete dokumenteerimine

- väga hea andmete dokumenteerimise juhend:
  - Siiri Fuchs, & Mari Elisa Kuusniemi. (2018, December 4). Making a research project understandable - Guide for data documentation (Version 1.2). Zenodo.  
DOI: <http://doi.org/10.5281/zenodo.1914401>
- README tekstifail esitatakse koos andmefailidega. README.txt fail annab teavet andmekogu kohta ja võimaldab andmeid õigesti tõlgendada nii endal kui ka teistel teadlastel pärast andmete jagamist või avaldamist
  - loo üks README.txt fail iga andmekogu kohta
  - nimeta see alati kujul README.txt või README.md (Markdown), mitte LOEMIND, readme, ABOUT jne
- README.txt fail peaks kindlasti sisaldama järgmist infot:
  - andmekogu pealkiri
  - andmekogu lühitutvustus (abstract)

- failide struktuur ja omavahelised seosed
- andmete kogumise meetodid
- kasutatud tarkvara (versioonid)
- kasutatud standardid
- spetsiifiline info andmete kohta (mõõtühikud, lühendite ja koodide selgitused jne)
- andmete taaskasutuse võimalused ja piirangud
- andmekogu üleslaadija kontaktandmed
- [README.txt faili loomise juhend](#)

## Metaandmed

- administratiivsed metaandmed projekti kohta (ID, rahastaja, PI, õigused ja litsentsid)
- tehnilised metaandmed (riist- ja tarkvara kohta, instrumendid, tööriistad, ligipääsuõigused)
- kirjeldavad metaandmed (autorid, pealkiri, lühikirjeldus, sisukirjeldus)
- DataCite Metaandmete raamistik (kohustuslikud, soovitatavad, valikulised) [DataCite Metadata Schema 4.4](#)
- metaandmete standardid on standardid selle kohta, milliseid välju oleks vaja täita
  - [Directory of Metadata standards](#)
  - Universaalsed metaandmete standardid:
    - [Dublin Core](#) (kasutusel [DataDOI](#)-s)
    - [Schema.org](#)
    - [DCAT](#)
- metaandmete kontrollitud sõnastikud ja klassifikatsioonid ütlevad, mida nendele väljadele kirjutada, kasutades standardset terminoloogiat
  - Näiteks:
    - [Eesti märksõnastik](#)
    - [AGROVOC Multilingual Thesaurus](#)
    - [Mammal Species of the World](#)
    - [GeoNames](#)

---

## ETHICS AND LEGAL COMPLIANCE

### TEADUSEETIKA. VASTAVUS ÕIGUSNORMIDELE

#### Teaduseetika

- Eesti Teadusagentuur: "[Juhised eetikaküsimuste lahendamiseks personaalse uurimistoetuse taotluses](#)" 2022
- juhul, kui projektil ei ole kokkupuudet juhendis mainitud eetiliste küsimustega, siis tuleb ka seda taotluses kirjeldada
- [Hea teadustava](#)
- [Hea akadeemiline tava ja akadeemilise eetika põhimõtete rakendamine Eesti Maaülikoolis](#)  
[Lisa 1. Akadeemilise eetika komisjoni moodustamine, tegevuse korraldus ja põhimõtted ning kahtluste ja kaebuste menetlemise kord](#)  
[Lisa 2. Hea akadeemilise tava ja akadeemilise eetika põhimõtete rakendamine](#)

## Autoriõigused ja intellektuaalomandi õigus

- kellele kuuluvad andmed (isiklikud õigused ja varalised õigused)
- andmetel on alati omanik, ka siis, kui need on avaandmed
- kuidas on andmed litsentseeritud
- [Eesti Maaülikoolis loodud intellektuaalomandi käsitlemise kord](#) 2021
- [Creative Commons litsentsid](#)
- kas andmete taaskasutamisel on mingeid piiranguid
- kas andmete jagamist lükatakse edasi/piiratakse, nt avaldamiseks või patentide taotlemiseks

## Isikuandmete kaitse

- Projektis vajalike isikuandmete kogumise ja töötlemise kirjeldamisel tugineda vastavalt [Isikuandmete kaitse üldmäärusele](#) ning [Eesti Isikuandmete kaitse seadusele](#)
- [Isikuandmete töötlemise põhimõtted Eesti Maaülikoolis](#)

---

## STORAGE AND BACKUP

### ANDMETE TURVALINE SÄILITAMINE TEADUSTÖÖ VÄLTEL

#### Säilitamine, varundamine, ülekanne ja taaste

- eesmärk on säilitada andmete tehniline ja sisuline kvaliteet:
  - käideldavus (kättesaadavus ja juurdepääsetavus)
  - konfidentsiaalsus (kättesaadav ainult selleks volitatud isikutele või süsteemidele, võtmehaldus, logifailide säilitamine)
- säilitus:
  - pilvekeskkonnad
  - kesksed serverid
  - tundlike andmete serverid
  - arvuti kõvaketas
  - väline kõvaketas
  - mobiilsed seadmed
- varundamine: andmete ja/või programmide hetkeseisu koopia loomine, mis pärast turvaintsidenti võimaldab ennistamist selle teadaoleva hetkeseisuni
  - kui sageli varundatakse, mitu koopiat, kas tööprotsess on automatiseeritud
  - masterfaili säilitamine ja varundamine
  - 3-2-1 reegel: kolm koopiat, hoitakse kahes erinevas kohas, millest üks asub kaugel
  - kes vastutab varundamise eest, eriti mobiilsete seadmete puhul
- soovitatav on riskianalüüs: mis saab siis kui...
  - IT-süsteemid ei tööta
  - juhtuvad elektrikatkestused, vee- ja tuleõnnetused
  - seade kaob või varastatakse
  - avastatakse kahjurvara seadmetes
  - töögrupi liige lahkub või sureb jne
- riskide hindamine: ohud ja nende esinemise tõenäosus, nõrkused, meetmed

- infoturbe standard peab vastama ISO 27001 ja 2025 a alguseks E-ITSile (Eesti infoturbestandardile)

### Ligipääs andmetele, infoturve

- ligipääsuõiguste haldamine (kas kõigil ühtmoodi, lepingupartnerite õigused, ajutise töötajate õigused)
- logifailide säilitamine
- pseudonümiseerimine, krüpteerimine, võtmehaldus
- andmevahetus, isikuandmed, kolmandad riigid
- organisatoorne ja füüsiline turve: uue töötaja koolitus, lahkuva töötajaga võimalikud kaasnevad probleemid, töösisekorraeeskirjad, tuleohutus, uste lukustamine
- vastutajad

---

## SELECTION AND PRESERVATION ANDMETE PIKAAJALINE SÄILITAMINE

### FAIR andmed

- millistel andmetel on pikaajaline väärtus? Nende säilitamine ja jagamine taaskasutamiseks
- andmete ettevalmistamine jagamiseks, FAIR andmed
- repositooriumi valik

### (F) Kuidas tehakse andmed leitavaks

- andmetel on püsidentifikaator DOI. Vaata [DataCite Eesti](#)
- metaandmed on DataCite registris
- standardsed metaandmeid nt [Dublin Core](#)
- masinloetavad metaandmed
- võtmesõnad ja märksõnad
- versioonihaldus

### (A) Kuidas tehakse andmed juurdepääsetavaks

- repositoorium, kus andmed säilitatakse
- millised andmed on avatud juurdepääsuga ehk avaandmed
- millised andmed jäävad suletuks ja mis põhjusel
- metaandmed peavad olema avatud ka juhul, kui andmed ei ole avatud (erandid nt haruldaste liikide asukohaandmed)
- tehnilised metaandmed: vajalik tarkvara (versioon), instrumentide spetsifikatsioonid, tarkvaratööriistad
- krüpteeritud andmed
- autentimine, kellelt küsida juurdepääsuõigusi
- kas on vaja luua kasutajakonto, mis seotakse teatavate tingimustega

### (I) Kuidas tehakse andmed koostöövõimelisteks teiste arvutisüsteemidega

- peamiselt repositooriumi ülesanne

- milliseid andmete ja metaandmete standardeid, kontrollitud sõnastikke ja taksonoomiaid kasutatakse
- andmetüüpide ja andmevormingute kirjeldused: kui ei ole standardsed, kuidas tagatakse koostöövõime
- linkimine teiste andmete, metaandmete ja spetsifikatsioonidega
- korrektne viitamine kasutatud andmekogudele
- andmevahetusstandardid

#### **(R) Kuidas tagatakse andmete taaskasutatavus**

- osaliselt repositooriumi ülesanne
- lisa [README.txt](#) fail
- kas tegemist on toorandmete, puhastatud andmete või töödeldud andmetega
- embargoperiood, põhjendus
- litsentsid: [Creative Commons litsentsid](#)
- viitamine: [DataCite viitevormindaja](#)
- standardsed metaandmed, milliseid (erialaseid) standardeid on kasutatud
- andmete päritolu tuvastamine (kes, kus, milleks kogus, kus on avaldatud, DOI-d)
- millist tarkvara versiooni on kasutatud
- kui kaua on tagatud andmete kättesaadavus taaskasutamiseks
- andmete kvaliteedi tagamine (käideldavus, terviklus, konfidentsiaalsus)
- soovitusel, kellele võiks need andmed vajalikud olla (README.txt failis)

---

## **DATA SHARING** **ANDMETE JAGAMINE**

### **Jagamine**

- kas andmed jagatakse repositooriumis või kui *supplementary data* artikli juures või eraldi artiklina andmeajakirjas
- millises repositooriumis andmeid säilitatakse
- kellele võiksid need andmed kasulikud olla
- kuidas jagad oma andmeid (kas on avaandmed või peab küsima, mis tingimusel saab)
- millal jagad (jooksvalt, pärast publikatsiooni ilmumist, embargo lõppemisel)
- kas andmed on lingitud publikatsiooniga
- lingi andmed oma [ORCID](#) kontoga

### **Juurdepääsupiirangud**

- millised andmed on avatud juurdepääsuga ehk avaandmed
  - millised andmed jäävad suletuks ja mis põhjusel
  - kas on krüpteeritud andmeid
  - kuidas toimub autentimine
  - andmete omaniku kontaktandmed
-

## RESPONSIBILITIES AND RESOURCES

### VASTUTUSALAD JA ANDMEHALDUSE MAKSUMUS

#### Kes vastutab andmehalduse eest

- ametikohtade järgi
  - juhtivteadur (PI): andmehalduse poliitika, andmehaldusplaani koostamine, lepingud, kulud, koolitused
  - teadlased: andmehaldusplaani järgimine ja täiendamine, andmehaldus, probleemide lahendamine
  - andmehaldur: koolitused, nõustamine, infoturve, säilitamine, varundamine, riist- ja tarkvara
  - laborant, abipersonal vastavalt nendele antud ülesannetele
- töövoos järgi
  - kes vastutab andmete kogumise, dokumenteerimise, metaandmete, infoturbe jne eest
- näide
  - [TU Delft RD Policy](#)

#### Planeeritavad kulud

- kulud on seotud peamiselt tööjõu, riist- ja tarkvaraga (arvutid, serverid, instrumendid, välitööde seadmed)
- juhendid, koolitused, ümberõpe, juristi ja/või DPO konsultatsioon, tõlketeenus
- andmete ost, salvestatud intervjuude transkribeerimine
- kasutuslitsentsid
- riistvara: arvutid, serverid, instrumendid, välitööde seadmed
- andmeanalüüs
- andmete vormindamine, anonümiseerimine, säilitamine repositooriumis
- partnerite kohtumised, konverentsid
- projekti andmehaldur
- üldine seisukoht: 5% projekti eelarvest