## Andmehaldusplaani soovitatav vorm

## Tõlke allikas: DCC. (2013). Checklist for a Data Management Plan. V.4.0. Edinburgh: Digital Curation Centre. [http://www.dcc.ac.uk/resources/data-management-plans](http://www.dcc.ac.uk/resources/data-management-plans%20)

# Administratiivsed andmed

**ID**

Identifikaator, väljastatud rahastaja/ülikooli poolt (kui on)

**Rahastaja**

Uuringu rahastaja (kui olemas)

**Grandi number**

Grandi number/url (projekti viitenumber)

**Projekti nimi**

Projekti nimi (sama uurimistaotluses olevaga)

**Projekti kirjeldus**

Tehke lühike kokkuvõte uuringu (või uuringute) liigist, et aidata teistel aru saada, mis eesmärgil andmeid kogutakse ja luuakse.

Lähtuge järgnevast:

* Mis on uurimisprojekti olemus?
* Milliste uurimisküsimustega tegeletakse?
* Mis eesmärgil andmeid kogutakse või luuakse?

**Uuringu juht / Teadur**

Uuringu juhi ja/või peamiste osalevate teadurite nimed

**Uuringu juhi / Teaduri ID**

Teadlaste identifikaatorid, näiteks ORCHID <http://orcid.org/>

**Projekti andmehaldur**

Nimi (kui erinev eelnevatest), telefon ja email

**Esmase versiooni kuupäev**

Andmehaldusplaani esimese versiooni valmimise kuupäev

**Viimane uuenduste kuupäev**

Andmehaldusplaani viimase muutmise kuupäev

**Seotud eeskirjad**

Loetlege kõik muud asjakohased rahastajate, institutsioonide, osakondade või rühmade andmehalduse, andmete jagamise ja andmete turvalisuse põhimõtted. Kui mingi osa andmehaldusplaanis esitatud teabest määratakse teiste asutuste/kolmandate osapoolte eeskirjade järgi, siis lisage viited/lingid.

Lähtuge järgnevast:

* Kas on olemasolevaid protseduure, millele tuginete oma lähenemisviisis?
* Kas teie osakonnal / grupil on juba andmehalduse juhised?
* Kas teie asutusel on andmekaitse- või turbepoliitika, mida järgite?
* Kas teie asutusel on teadusuuringute andmete haldamise poliitika?
* Kas teie rahastajal on uurimisandmete haldamise poliitika?

**Andmete kogumine**

**Milliseid andmeid kogutakse või luuakse?**

Kirjeldage lühidalt andmeid, sealhulgas ka juba olemasolevaid või kolmandate osapoolte andmeid, mida planeerite kasutada ning tooge välja nende andmete sisu, tüüp ja katvus. Nimetage ja põhjendage valitud andmete formaati ning kirjeldage sellest tulenevat andmeformaadi ja -mahu mõju andmete säilitamisele, varundusele ja juurdepääsetavusele.

Lähtuge järgnevast:

* Millist tüüpi, mis formaadis ja millises mahus andmeid kogutakse?
* Kas teie valitud formaat ja tarkvara võimaldavad andmete jagamist ja pikaajalist juurdepääsu?
* Kas on olemas andmeid, mida saate uuesti kasutada?

Andmete maht:

* Milline on andmemahu suurus MB/GB/TB-na. Märkige algandmete, töödeldud andmete ja muude teisejärguliste väljundite (nt aruanded) proportsioonid.
* Kaaluge mahtude mõju andmete salvestamisele, juurdepääsule ja säilitamisele. Kas peate arvestama lisakuludega?
* Kaaluge, kas andmemaht võib tekitada probleeme andmete jagamisel või edastamisel; kui jah, siis kuidas seda lahendada?

Andmete vorming:

* Tooge välja, millises vormingus (vormingutes) andmed on, nt lihttekst (.txt), komaeraldusega väärtused (.csv), geograafiliselt viidatud TIFF (.tif, .tfw).
* Selgitage, miks olete valinud konkreetsed vormingud. Otsused võivad põhineda uurijate asjatundlikkusel, avatud vormingute eelistamisel, andmekeskuste aktsepteeritud standarditel või antud kogukonnas/erialal levinud kasutamisel.
* Standardiseeritud, vastastiku vahetatavate või avatud vormingute kasutamine tagab andmete pikaajalise kasutatavuse; selliseid vorminguid on soovitatav kasutada ka andmete jagamisel ja arhiveerimisel.
* Kirjeldage ja põhjendage oma vormingu valikut ning arvestage andmevormingu ja andmemahtude mõju salvestamisel, varundamisel ja juurdepääsetavusel.

Erinevatest failivormingutest lisaks: UK Data Service’i [soovitatud formaate](https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats)

 DataONE [failivormingute jaoks](https://dataoneorg.github.io/Education/bestpractices/document-and-store)

Dutch National Centre of Expertise and Repository for Research Data [formaatidest](https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats).

Andmete kirjeldused:

* Tehke kokkuvõte kogutavatest või loodavatest andmetest, märkides teabe sisu, katvuse ja tüübi, nt tabeliandmed, uuringuandmed, eksperimentaalsed mõõtmised, mudelid, tarkvara, audiovisuaalsed andmed, füüsilised proovid jne.
* Mõelge, kuidas andmed saaksid olemasolevaid andmeid täiendada ja nendega integreeruda või kas on olemas praeguseid andmeid või meetodeid, mida saaksite uuesti kasutada.
* Märkige, millised andmed on pikaajalise väärtusega ja mida tuleks jagada ja/või säilitada.
* Kui ostate või taaskasutate olemasolevaid andmeid, selgitage, kuidas on lahendatud selliseid küsimused nagu autoriõigused ja intellektuaalomandi õigused. Peaksite püüdma minimeerida piiranguid kolmandate osapoolte andmete taaskasutamisel ja edasisel jagamisel.

**Kuidas andmeid kogutakse ja luuakse?**

Andke ülevaade, kuidas andmeid kogutakse ja/või luuakse ning milliseid andmestandardeid kasutatakse (kui kasutatakse). Mõelge, kuidas andmeid projekti ajal hallatakse, tuues näiteks kas failide nimetamise tavasid, versioonihaldust ja kaustastruktuuri vm. Kirjeldage, kuidas kontrollitakse ja dokumenteeritakse andmete kogumise järjepidevust ja kvaliteeti. See võib sisaldada protsesse nagu kalibreerimine, proovide või mõõtmiste kordamine, standardiseeritud andmekogumine või salvestamine, andmesisestuse valideerimine, andmete eelretsenseerimine või märksõnastike kasutamine andmete esitamisel.

Lähtuge järgnevast:

* Milliseid standardeid või metodoloogiaid kasutatakse?
* Kuidas struktureeritakse ja nimetatakse kaustad ja failid?
* Kuidas toimub versioonihaldus?
* Milliseid kvaliteedi tagamise protsesse kasutatakse?

**Näide 1:**

Kogutakse kursuse vaatlusandmeid, õppejõudude intervjuude andmeid ja üliõpilaste küsitluste andmeid. Andmeid kogutakse uurimisperioodil (jaanuar 2022 – detsember 2022). Enamik andmeid esitatakse tekstivormingus (küsitlused paberil, märkmed).

Iga faili nimetamisel kasutatakse lühikest kirjeldust/akronüümi, mis kajastab faili sisu, millele järgneb loomise kuupäev. Erinevate versioonide markeerimiseks lisame failinimesse versiooni numbri. Näiteks failinimi GSC\_20220608\_v01.xls tähistab 8. juunil 2022 hangitud 1. versiooni andmeid.

Loome dokumendi, milles täpsustame failide nimetamise tavasid ja esitame failinimedes kasutatud lühendite/akronüümide selgituste loendi.

**Näide 2**:

Laboratoorsed katseandmed kogutakse kasutades mikroskoopi. Aja- ja asukohatempliga pildifailid luuakse *Asukoha Nimi* loodusvaradest. Antud kujutisi kasutatakse olendite, looduslike artefaktide ja valitud paikade olude dokumenteerimiseks konkreetsetes kohtades ja kellaaegadel ajavahemikul 2021–2036. Enamike fotode kohta luuakse ka taksonoomiline teave ja metaandmed. Andmed nähtuste kohta on vaadeldavad ja kvalitatiivsed. Andmete taaskasutuse hõlbustamiseks säilitatakse metaandmete failid.

**Näide 3:**

**Peamiselt avalikud andmed aastatest 2000–2015 hangitakse *asutusest XXX*. Ostetakse ja kogutakse ka teatud hulgal esialgseid (mitteavalikke) rahvaloenduse andmeid ja muid andmeid teatud allikatest, nagu näiteks *XXX Riiklikust Statistikaametist* ja *XXX Riigi Terviseametist*.**

**Näide 4:**

Algandmetena kogutakse eesti- ja ingliskeelseid audiofaile. Pärast failide transkribeerimist luuakse tekstifailid. Intervjuude ja ümberkirjutiste kogumiseks kasutatakse krüpteeritud helisalvestusseadmeid. Intervjuude ja fookusgrupi intervjuude audiofaile ei säilitata helisalvestusseadmetes, pärast kogumist kantakse need turvaliselt projekti pilvepõhisesse virtuaalsesse uurimiskeskkonda läbi turvalise failiedastusprotokolli (FTP – File Transfer Protocol).

**Näide 5:**

Hinnanguliselt kogume umbes 800 küsitlust, 20 intervjuud (igaüks umbes 30 min pikkune) ja 2 fookusgrupi intervjuud (kumbki umbes 90 min). Andmete, sealhulgas versioonide (alg-, põhi-, analüütilised andmed) arvestus kokku on hinnanguliselt alla 30 GB.

**Näide 6:**

Meie failivormingud saavad olema nii vabas kui ka patenditud andmeformaadis. Vabad vormingud tagavad, et neid andmeid saavad kasutada kõik soovijad, kui need on hoiule antud ja avalikult kättesaadavaks tehtud.

Uuringud eksisteerivad .csv (vabavara), MS Exceli ja SPSS (mõlemad patenditud) vormingus. Lisateavet SPSS-i kohta leiate aadressilt: SPSS Wikipedia <https://et.wikipedia.org/wiki/SPSS>

Intervjuude ja fookusgruppide andmed on .mp3 (tasuta), MS Word ja NVivo (mõlemad omanduslikud) vormingutes. NVivo kohta leiate lisateavet: NVivo Wikipedia <https://en.wikipedia.org/wiki/NVivo>

Kõik jagamise ja pikaajalise juurdepääsu jaoks hoiule antud uuringuandmed on .csv-vormingus, seega saab igaüks neid kasutada ilma tasulist tarkvara omamata.

Intervjuude ja fookusgruppide ümberkirjutiste lõplikud tuvastamata versioonid eksporditakse hoiustamise, pikaajalise säilitamise ja juurdepääsuks tavalisse vabasse tekstivormingusse.

**Näide 7:**.

Kogutakse anduriandmeid, pilte ja võimalikke kolmanda osapoole andmeid (ilma- ja teeolud). Andmed salvestatakse Exceli arvutustabelitena ja SQL-i andmebaasis.

**Näide 8:**

Kvantitatiivsed andmed kogutakse *motion capture* süsteemi abil. Töödeldud andmetüübid hõlmavad Matlabi faile, MS Exceli faile, koodiraamatu tekste ja graafilisi faile.

**Näide 9:**

Eeldatavalt toodetavad andmed, proovid ja materjalid koosnevad laborimärkmetest, katsete käigus saadud algandmete failidest, katseanalüüsi andmefailidest, simulatsiooniandmetest, mikroskoobipiltidest, optilistest piltidest, mille andmeid kirjeldatakse allpool:

A. Laborimärkmed: kraadiõppur ja uuringu juht registreerivad käsitsi kõik uurimistöö käigus tekkinud vaatlused, protseduurid ja ideed.

B. Katsete käigus saadud algandmete failid: need failid koosnevad ASCII tekstist, milles sisalduvad otse erinevatest ülivõre nanotraadi termoelektriliste seadmete termoelektriliste omaduste mõõtmiseks kasutatavatest elektrilistest instrumentidest kogutud andmeid.

C. Katseanalüüsi andmefailid: need failid koosnevad A-osas nimetatud algandmete tabelitest ja hajuvusgraafikutest. Nendes failides olevaid andmeid on töödeldud, et anda seadme tõhususe kohta sisukad ja kvantitatiivsed väärtuseid. Analüüs viiakse läbi kasutades seadmete tõhususe arvutamise parimaid tavasid ja üldtunnustatud meetodeid.

D. Simulatsiooniandmed: need andmed esindavad turul olemasoleva simulatsiooni- ja modelleerimistarkvara tulemusi kvantvangistuse modelleerimiseks.

E. Mikroskoobipildid: kavandatud räni nanostruktuuride kujutised genereeritakse skaneeriva elektronmikroskoopia (SEM) abil, ülekandelektronmikroskoopia (TEM) abil suure eraldusvõimega, et kvantifitseerida traadi läbimõõt ja karedus, ja aatomijõumikroskoopia (AFM) abil.

F. Optilised pildid: nanostruktuuriga seadmete pildid kogutakse optilise mikroskoobi abil erinevatel suurendussätetel.

G. Ülivõre nanotraadi proovid: nanostruktuuriga proovid koosnevad räni kvantpunkti ülivõre nanotraatidest. Katse läbiviija kasutab neid proove seadme tõhususe mõõtmiseks.

**Dokumentatsioon ja metaandmed**

**Milline dokumentatsioon ja metaandmed on teadusandmetega seotud?**

Kirjeldage andmetega kaasnevat dokumentatsiooni, mis aitaks andmeid taaskasutada ja nendest aru saada. See peaks sisaldama vähemalt kõige lihtsamaid üksikasju selle kohta, kes andmed lõi või neid täiendas, pealkiri, loomise kuupäev ja millistel tingimustel on andmed ligipääsetavad. Samuti võib kirjeldada kasutatud metodoloogiat, analüütilist ja protseduurilist informatsiooni, muutujate määratlust, märksõnastikke, mõõtühikuid, tehtud eeldusi, andmete kogumise ja/või töötlemise tarkvara ja failiformaati ja -tüüpi. Mõelge, kuidas ja kuhu selle informatsiooni salvestate. Soovituslik on kasutada asutuse/eriala metaandmete standardeid kui need on välja kujunenud. Andmehoidlad võivad samuti suunata asjakohastele metaandmete standarditele.

Lähtuge järgnevast:

* Millist teavet on vaja andmete edaspidiseks taaskasutamiseks ja tõlgendamiseks?
* Kuidas jäädvustate/loote vajaliku dokumentatsiooni ja metaandmed?
* Milliseid metaandmete standardeid kasutate ja miks?
* Milline metaandmestik luuakse, et teadusandmed oleks järgnevate kasutajate jaoks identifitseeritavad ja leitavad?

Valides „Metaandmete standardid“:

**Näide 1:**

Projekti käigus kogutud kliinilised andmed dokumenteeritakse CDASH v1.1 standardit kasutades. Standard on kättesaadav [CDISC-i veebilehel](http://www.cdisc.org/cdash).

**Näide 2:**

Kasutades elektroonilist laborimärkmikku genereerime iga märkmiku ja postitusega metaandmeid. Metaandmed hõlmaksid jaotisi, kategooriaid ja võtmeid, mille määraksid kaastöötajad, et säiliks taaskasutamiseks vajalik järjepidevus terminoloogia kasutamisel. Samuti kasutaksime kemikaalide ja materjalide omaduste kirjeldamisel ontoloogiat (ChemAxiomProp).

**Näide 3:**

Kasutame andmete kirjeldamisel [TEI metaandmete standardite](http://www.tei-c.org/index.xml) põhielemente. Lisame metaandmetele ka kohandatud elemente, et kirjeldada paremini õiguste haldust.

Valides „Ei kasutata metaandmete standardeid“:

**Näide 1:**

Selle projekti jaoks kogutud ja loodud andmete jaoks ei kasuta rahvusvahelisi metaandmete standardeid. Kuid on tagatud, et igal Microsoft Wordi, Microsoft Exceli ja Microsoft PowerPointi keskkonnas loodud dokumendil on dokumendi atribuutides piisavalt põhiteavet, nagu näiteks autori nimi, pealkiri, teema, märksõnad ja muu. Lisaks sellele koostatakse eraldi *readme*-fail, mis kirjeldab andmete üksikasju. *Readme*-faili (failide) loomisel rakendatakse *XXX ülikooli* soovitusi. Põhielemendid võiksid olla järgmised: sissejuhatav teave andmete kohta, metodoloogiline, kuupäevapõhine ja jagamise/juurdepääsuga seotud teave.

**Näide 2:**

Üksikute piltide ajastamise ja särituse metaandmed genereeritakse automaatselt kaamera poolt. Seejärel lisatakse järeltöötluses GPS-i asukohad jagatud ajatemplitel põhineva GPS-i jälgimise põhjal. Pildiandmekogumi kui terviku metaandmed genereerib pildihaldustarkvara (iMatch) ning see sisaldab ajavahemikke, asukohti ja taksonite loendit. Need metaandmed tõlgitakse ökoloogiliste metaandmete keelde (*Ecological Metadata Language*, EML), mis on loodud Morpho tarkvaratööriista abil ja mis sisaldavad asukoha ja taksonoomilisi kokkuvõtteid.

Andmekogumiga kaasneb *README*-fail, mis kirjeldab kataloogihierarhiat ja failinimede kasutamist.

Iga kataloog sisaldab INFO.txt-faili, mis kirjeldab selles katses kasutatud katseprotokolli. Samuti registreeritakse kõrvalekalded protokollist ja muu kasulik kontekstuaalne teave. Mikroskoobipildid hõivavad ja salvestavad iga pildiga erinevaid metaandmeid (välja suurus, suurendus, objektiivi faas, suum, võimendus, ava läbimõõt jne). See peaks võimaldama meie uurimisrühma teistel liikmetel andmeid mõista ja lisama andmestikule kontekstipõhist väärtust, kui seda tulevikus uuesti kasutatakse.

**Eetika ja õiguslikud aspektid**

**Kuidas lahendatakse võimalikud eetilised probleemid?**

Eetilised probleemid mõjutavad seda, kuidas andmeid salvestatakse, kes neid kasutada/näha saab ning kui kaua neid säilitatakse. Eetiliste probleemide haldamine võib hõlmata järgmist: andmete anonümiseerimine; suunamine osakondade või institutsioonide eetikakomisjonidesse; ametlikud nõusolekulepingud. Peaksite näitama, et olete teadlik kõikidest võimalikest probleemidest ja olete nendeks ette valmistunud. Kui viite läbi uuringuid, milles osalevad inimesed, peate tagama ka andmete jagamise ja taaskasutuse võimaldamise kohta nõusoleku küsimise.

Lähtuge järgnevast:

* Kas olete saanud nõusoleku andmete säilitamiseks ja jagamiseks?
* Kuidas kaitsete vajadusel osalejate identiteeti? Nt. anonümiseerimise kaudu
* Kuidas käsitletakse tundlikke andmeid, et tagada nende turvaline salvestamine ja edastamine?

Vaata UK Data Service’i juhtnööre [andmete jagamise nõusoleku kohta.](https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing)

[Teaduseetika](https://www.etag.ee/teaduseetika/) ETAG kodulehel

 **Näide 1:**

Uuring hõlmab tundlikke andmeid, kuna need sisaldavad isiku tuvastamist võimaldavaid andmeid.

Uuring hõlmab andmeid suguhaiguste suhtes skriinitud subjektidelt. Lõplik andmekogum sisaldab intervjuude käigus subjektide poolt enda kohta teatatud demograafilisi ja käitumuslike andmeid ning uriiniproovide laboriandmeid. Kuna uuritavad suguhaigused on raporteeritavad haigused, kogume identifitseerivat teavet. Ehkki lõplikust andmekogumist eemaldatakse identifikaatorid, jääb võimalus ebatavaliste omadustega subjektide deduktiivseks avalikustamiseks. Seega muudame andmed ja dokumentatsiooni kättesaadavaks ainult selleks ette nähtud andmete jagamise lepingu alusel, millega kaasneb: (1) kohustus kasutada andmeid ainult teadusuuringute eesmärgil, mitte üksikute osalejate tuvastamiseks; 2) kohustus andmete turvamiseks sobiva tehnoloogia abil; ja (3) kohustus hävitada või tagastada andmed pärast analüüside lõppu.

**Näide 2:**

Mul on tundlikke andmed, kuna need sisaldavad isiku tuvastamist võimaldavad andmeid.

Juurdepääs uuringudokumentidele on piiratud esmase uurimisrühma liikmetega. Salvestatud andmetel eemaldatakse kogu identifitseeriv teave ja need sildistatakse uuesti uuringu koodinumbritega. Andmebaas, mis seob uuringukoodinumbrid nõusolekuvormide ja identifitseerimisandmetega, salvestatakse parooliga kaitstud arvutites turvalises lukustatud kontoris eraldi. Osalejate privaatsuse säilitamiseks koosnevad üksikandmete üksused ainult tulemusnäitajatest ilma demograafilise või identifitseeriva informatsioonita.

**Kuidas hallatakse autoriõiguste ja intellektuaalomandi küsimusi?**

Märkige, kellele kuulvad kogutavate ja loodavate andmete autoriõigused ja intellektuaalomandi õigused ning kasutamise ja taaskasutamise litsents(id). Mitme partneriga projektide puhul peaks intellektuaalomandi õigused olema lepingutes määratud. Mõelge võimalikele rahastaja-, institutsiooni- või grupipoliitikatele autoriõiguste või intellektuaalomandi suhtes. Mõelge ka kolmandate osapoolte andmete taaskasutamise lubadele ja andmete jagamiseks vajalikele piirangutele. Tooge välja kõik piirangud, mis on vajalikud andmete jagamisel, näiteks seoses patenteeritavate andmete kaitsega. Selgitaga kuidas andmeid korduvkasutuseks litsenseeritakse.

Lähtuge järgnevast:

* Kellele andmed kuuluvad?
* Mis litsents antakse andmete taaskasutuseks?
* Kas kolmandate osapoolte andmete taaskasutamisel on mingeid piiranguid?
* Kas andmete jagamist lükatakse edasi/piiratakse, näiteks kuni publikatsioonide avaldamiseni või tulemuste patenteerimiseni?

**Säilitamine ja varundamine**

**Kuidas andmeid uuringu jooksul salvestatakse ja varundatakse?**

Märkige, kui tihti ja kuhu andmed salvestatakse ja varundatakse. Kui palju varukoopiaid tehakse? Andmete hoidmine ainult sülearvutites, arvutite kõvaketastel ja välistel kõvaketastel on üsna riskantne. Eelistatav on kasutada ülikooli IT-meeskonna poolt pakutavat stabiilset ja hallatud salvestusruumi. Samamoodi on tavaliselt parem kasutada IT-teenuste poolt pakutavaid automaatseid varundusteenuseid kui tugineda manuaalsetele protsessidele. Kui otsustate kasutada kolmanda osapoole teenust, siis peaksite tagama, et see ei oleks vastuolus rahastajate, institutsioonide, osakondade poliitikaga, näiteks andmete juriidilise jurisdiktsiooniga või tundlike andmete kaitsega.

Lähtuge järgnevast:

* Kas on piisavalt salvestusruumi või tuleb arvestada lisanduvate tasudega lisateenuste eest?
* Kuidas andmeid varundatakse?
* Kes vastutab varundamise ja taastamise eest?
* Kuidas taastatakse andmed intsidendi korral?
* Kaaluge andmete turvalisust, peamiselt juhul, kui andmed on tundlikud, nt üksikasjalikud isikuandmed, poliitiliselt tundlik teave või ärisaladused. Pange tähele esmaseid riske ja kuidas neid juhtida.

Vaadake [UK Data Service Guidance'it](https://www.ukdataservice.ac.uk/manage-data/store) andmete salvestamise kohta.

**Näide 1:**

Kasutan võrgupõhist salvestusketast *XXX*, mida saavad kasutada kõik teadurid ja tudengid aktiivses kasutuses olevate andmete säilitamiseks. See on täielikult varundatud, paindlik ning paikneb mitmes asukohas. Sellele pääseb ligi läbi VPN-i ka väljastpoolt ülikooli.

**Näide 2:**

Andmeid säilitatakse kohapeal turvalises salasõnaga kaitstud andmeserveris. Üks komplekt kõvakettaid ja üks komplekt kassette säilitatakse *XXX* majas. Teine komplekt kõvakettaid ja kassette säilitatakse *XXX* majas.

**Näide 3:**

Andmeid (töötajate arvutites ja veebiserveris) hallatakse vastavalt kolledži IT osakonna tavapärastele praktikatele ning need on parooliga kaitstud. Kõik piiratud ligipääsuga, mitteavalikud andmed säilitatakse *XXX* (*Piiratud Juurdepääsuga Andmekeskuses*).

**Varundus & versioonihaldus**

**Näide 1:**

Materjalide täielikud koopiad genereeritakse ja säilitatakse eraldi peamises- ja varuhoidlas nii uuringu juhi kui ka uuringu kaasjuhi jaoks (jooksvalt) ning koos kõikide ekspertide kogu liikmetega iga kuue kuu tagant. Projekti meeskond rakendab National Institute of Dental and Craniofacial Research’i versioonihalduse juhiseid, et korrastada andmete versioone ning tagada erinevate versioonide tuvastamine, korralik kontrollimine ja õigesti kasutamine.

**Näide 2:**

Kõik uurimisrühma poolt tehtavad muudatused intervjuude transkriptsioonide failides ja kodeerimisel järgivad [University of Leicester'i](https://www2.le.ac.uk/services/research-data/documents/UoL_VersionControlChart_d0-1.pdf) poolt soovitatud versioonihalduse standardeid.

**Näide 3:**

Andmete haldamisel kasutame tasuta andmeallikate hajushaldamise tööriista [Mercurial](https://www.mercurial-scm.org/%22%20%5Ct%20%22_blank), et andmed oleks lihtsasti tuvastatavad ning korrektselt kontrollitud ja kasutatud.

**Näide 4:**

Kõik andmed varundatakse igakuiselt käsitsi uurimisrühma kontoris asuva arvuti kõvakettale rühma liikme *XXX* poolt. Arvuti on kaitstud salasõnaga, mille kasutamise õigus antakse ainult rühma liikmetele. Sammvarundus toimub igal öösel ja täielik varundus toimub kord kuus. Vigade parandamise või uuendamise tõttu muudetud failid hoitakse alles arhiivisüsteemis. Kõik muudatused salvestatakse muudatuste ajalugu sisaldavasse dokumenti.

**Kuidas on lahendatud juurdepääsetavus ja turvalisus?**

Kui andmed on konfidentsiaalsed (näiteks isikuandmed, mis ei ole avalikes andmebaasides, konfidentsiaalsed andmed või ärisaladused), siis tuleb välja tuua sobivad turvameetmed ja märkida kasutatavad standardid, nt ISO27001.

Lähtuge järgnevast:

* Millised on riskid andmete turvalisusele ning kuidas neid riske hallatakse?
* Kuidas kontrollite ligipääsu andmetele, et neid turvaliselt hoiustada?
* Kuidas tagate, et uuringurühma liikmed saavad andmetele turvaliselt ligi?
* Kuidas tagate välitööde käigus loodud või kogutud andmete turvalise ülekandmise oma põhilisse turvatud süsteemi?

# Valik ja säilitamine

**Millised andmed on pikaajalise väärtusega, mida tuleks säilitada ja jagada?**

Mõelge, kuidas saaks andmeid taaskasutada, nt oma uurimistulemuste kinnitamiseks, uute uuringute läbiviimiseks või õpetamiseks. Otsustage, milliseid andmeid ja kui kaua säilitada. Otsus võib põhineda mis tahes kohustustel teatud andmeid säilitada, võimalikul taaskasutusväärtusel, sellel mida on majanduslikult otstarbekas säilitada ning andmete jagamiseks ja säilitamiseks tehtavate lisanduvate ettevalmistuste raskusest.

Lähtuge järgnevast:

* Milliseid andmeid tuleb lepingulistel, juriidilistel või regulatiivsetel eesmärkidel säilitada või hävitada?
* Kuidas otsustate, milliseid andmeid säilitada?
* Millised on andmete prognoositavad kasutusalad?
* Kui kaua andmeid hoitakse ja säilitatakse?

Vaata DCC juhendit: [How to appraise and select research data for curation.](http://www.dcc.ac.uk/resources/how-guides/appraise-select-data)

**Milline on andmekogumi pikaajaline säilitamiskava?**

Mõelge, kuidas pikaajalise väärtusega andmekogumeid säilitatakse ja kureeritakse pärast grandi kehtivusaega. Samuti tooge välja andmete ettevalmistamise ja dokumenteerimise plaanid jagamise ja arhiveerimise kohta. Kui te ei kavatse kasutada loodud hoidlat, peaks andmehalduskava näitama, et ressursid ja süsteemid andmete tõhusaks kureerimiseks on olemas ka pärast uurimistoetuse kehtivusaega.

Lähtuge järgnevast:

* Kus, nt millises hoidlas või arhiivis andmeid hoitakse?
* Kui palju maksab valitud hoidla või arhiivi kasutamine (kui üldse)?
* Kas olete arvestanud andmete jagamisele/säilitamisele kuluvate ettevalmistuse aja ja

Tallinna Tehnikaülikoolil on oma teadusandmete repositoorium.

**Andmete jagamine**

**Kuidas andmeid jagatakse?**

Kaaluge kus, kuidas ja kellele pikaajaliseks säilitamiseks mõeldud andmed kättesaadavaks tehakse. Andmete jagamise meetodite valik tuleneb erinevatest teguritest, nagu näiteks andmete tüüp, suurus, keerukus ja tundlikkus. Kui võimalik, mainige varasemaid näiteid andmete tõhusast jagamisest. Kuidas minimeerida piitanguid.

Andmete anonümiseerimine, nõusoleku saamine avaldamiseks uuringus osalejatelt. Vajadusel kohaldage embargo perioodi, kuid mitte pikaajalist ainukasutust. Kui andmetel on korduvkasutamise potentsiaali, peaksite kasutama seda hõlbustavaid standardeid ja vorminguid ning tagama, et andmete leidmiseks vajalikud asjakohased metaandmed oleksid veebis kättesaadavad. Rakendada tuleks püsivaid tunnuseid, et andmeid oleks võimalik usaldusväärselt ja tõhusalt leida, tsiteerida ja taaskasutada.

Lähtuge järgnevast:

* Kuidas saavad potentsiaalsed kasutajad andmete kohta teada?
* Kellega ja millistel tingimustel andmeid jagatakse?
* Kas jagate andmeid hoidla kaudu, käsitlete taotlusi otse või kasutate mõnda muud moodust?
* Millal andmed kättesaadavaks tehakse?
* Kas hangite andmetele püsiva identifikaatori?

**Kas andmete jagamisel on vaja piiranguid?**

Esitage kõik eeldatavad raskused tunnustatud pikaajalise väärtusega andmete jagamisel, koos raskuste põhjuste ja võimalike meetmetega takistuste ületamiseks. Piirangud võivad olla tingitud näiteks konfidentsiaalsusest, nõusolekulepingute puudumisest või intellektuaalomandi õigustest. Mõelge, kas mitteavaldamise leping pakuks konfidentsiaalsete andmete puhul piisavat kaitset.

Lähtuge järgnevast:

* Mida teete piirangute ületamiseks või minimeerimiseks?
* Kui kaua vajate andmete ainukasutust ja miks?
* Kas on vaja andmete jagamise lepingut (või samaväärset)?

**Näide:**

Selle teadustöö andmekogumid, mis on väljaande aluseks, antakse hoiule *XXX* institutsionaalsesse andmehoidlasse ning need tehakse avalikkusele kättesaadavaks väljaande avaldamise hetkel. Andmeid säilitatakse hoidlas vastavalt rahastaja ja ülikooli andmepoliitikale. Hoidlasse XXX salvestatud failidele antakse digitaalse objekti identifikaator (DOI) ja sellega seotud metaandmed. Hoidlas asuvatele andmekogumitele väljastatud DOI saab lisada publikatsioonides viitena andmetele, mis võimaldab tuvastada ja pääseda ligi väljaande aluseks olevatele andmekogumitele. Hoidlas XXX hoitavate andmekogumite metaandmed on avalikult otsitavad ja leitavad ning näitavad, kuidas ja millistel tingimustel andmekogumile juurde pääseb.

**Vastutusala ja ressursid**

**Kes vastutab andmehalduse eest?**

Pange kirja rollid ja vastutusalad kõikide seotud tegevuste kohta, nt andmete kogumine, metaandmete loomine, andmete kvaliteet, salvestamine, varundamine, arhiveerimine ja jagamine. Mõelge, kes vastutab vajalike eeskirjade ja nõuete järgimise eest. Võimalusel nimetage konkreetsed isikud. Kui andmete omamine ja vastutus on osa partnerite konsortsiumileppest, siis samuti selgitage vastutust.

Lähtuge järgnevast:

* Kes vastutab andmehaldusplaani koostamise, kasutamise, üle vaatamise ja muutmise eest?
* Kes vastutab iga andmehaldustegevuse eest?
* Kuidas jagatakse vastutus partnerite vahel koostööprojektides?
* Kas andmete omamine ja vastutus andmehalduse osas on osa partnerite vahel sõlmitud konsortsiumikokkuleppest või lepingust?

Vaadake Suurbritannia [andmeteenistuse juhiseid](https://www.ukdataservice.ac.uk/manage-data/plan/roles-and-responsibilities) andmete haldamise rollide ja vastutuse kohta.

<https://dataoneorg.github.io/Education/bestpractices/define-roles-and.html>

**Milliseid vahendeid on vaja plaani teostamiseks?**

Kaaluge hoolikalt ja põhjendage plaani elluviimiseks vajalikke vahendeid, sealhulgas säilitamiskulud, tarkvara, riistvara, tehniline ekspertiis, personali tööaeg, andmete ettevalmistamise hoiustamiskulud ja võimalikud repositooriumi kulud. Pange kirja ja põhjendage, kui mõni ressurss on ette nähtud vaid selle konkreetse projekti jaoks.

Lähtuge järgnevast:

* Kas on vajalik kaasata täiendavad eksperte või on vajalik olemasoleva personali koolitus?
* Kas vajate täiendavat või erilist riistvara või tarkvara, mida teie asutus ei paku?
* Kas andmerepositooriumid rakendavad tasusid?