HORISONT 2020 andmehaldusplaani koostamine

juhend

Koostaja:

Kai Kalvik

TalTechi raamatukogu erialainfotalitus

Tel 620 3554

kai.kalvik@taltech.ee

SISUKORD

SISSEJUHATUS

Horisont 2020 andmehaldusplaani punktid tuginevad juhendile **Guidelines on FAIR Data Management in Horizon 2020** (Version 3.0, 26 July 2016). Juurde on lisatud erinevatest allikatest pärit selgitused ja soovitused andmete haldamisega seonduva kohta.

Euroopa Komisjoni avatud juurdepääsu (AJ) põhimõtteid hakati rakendama 7. Raamprogrammis (RP7), kus soositi võimalikult lühikest embargoperioodi publitseerimisel ning hüvitati kirjastamiskulud autoritele, kes avaldasid oma artiklid koheselt AJ ajakirjas. RP7 OpenAIRE projektide käigus on toetatud AJ repositooriumite võrgustikku ning on loodud **OpenAIRE e-taristu ja portaal** teadustöö tulemuste ja teadusandmete sidumiseks.

AJ poliitikat rakendatakse ka raamprogrammis **Horisont 2020** (2014–2020), mis koondab ELi teadusuuringute rahastamisvahendeid ning on maailma suurim teadus- ja arendustegevust ning koostööd toetav programm.

Kui Horisont 2020 rahastusel valminud publikatsioonidele AJ nõue juba kehtib, siis uurimisandmete avalikustamise ja taaskasutamise edendamiseks rakendatakse paindlikku pilootprogrammi **Open Research Data Pilot** (ORD Pilot) ja üleminekuaega. Järk-järgult viiakse sisse nõuded, et **Horisont 2020 teadusprojektide taotlejad on kohustatud esitama andmehaldusplaani ja tegema projekti teadustöö andmed võimalikult avatuks.**

Esialgu vaid mõne teemavaldkonna puhul rakendatav pilootprojekt laieneb kõikidele Horisont 2020 teemadele 2017. aastal. Toetuse saaja võib põhjendatud vajaduse korral pilootprojektis osalemisest loobuda või avada andmed vaid osaliselt.

Hea andmehaldus ja AHP koostamine aitavad teadlasel teha tema andmed leitavaks, kättesaadavaks, koostalitlusvõimeliseks ja taaskasutatavaks. Horisont 2020-s väljendub see põhimõtte ja akronüümina **"FAIR data principles"** (**F**indable, **A**ccessible, **I**nteroperable and **R**eusable).

Põhimõte "as open as possible, as closed as necessary" – **"nii avatud kui võimalik, nii suletud kui vajalik"** suunab järgima teadustöö parimaid tavasid ning pidama kinni intellektuaalomandi, isikuandmete, ärisaladuste ja julgeoleku kaitsmise jm konfidentsiaalsuse nõuetest.

Andmehaldusplaan (**AHP**) (Data Management Plan, DMP) on uurimistöö juurde kuuluv kava, mille eesmärgiks on kirjeldada uurimistöö andmehalduse protsessi projekti jooksul ja pärast projekti lõppemist.

AHP on teadusandmete haldamise võtmeelement, milles kirjeldatakse:

- milliseid andmeid uurimistöö projekti käigus kogutakse ja kasutatakse;
- milliseid meetodeid ja standardeid rakendatakse;
- kas andmeid jagatakse/tehakse avalikult kättesaadavaks;
- kuidas andmeid säilitatakse ja arhiveeritakse.

Pilootprojekt hõlmab **uurimistöö alusandmeid** (vajalikud teaduspublikatsioonides saadud tulemuste kinnitamiseks) ja nendega seotud **metaandmeid** (kirjeldavad säilitatavaid andmeid) ning muid andmeid (ei ole otseselt seotud publikatsiooni või toorandmetega) ning nende metaandmeid.

Vaba juurdepääsuga andmed tuleb **säilitada repositooriumis**, võimaldada kolmandatele osapooltele tasuta juurdepääs ning võimalused andmekaeveks, andmete kasutamiseks, reprodutseerimiseks ja levitamiseks. Kui ei ole kavas pikaajalise säilitamisväärtusega andmeid repositooriumis arhiveerida, peavad AHP-s esitatud meetmed kindlustama andmehalduse tulemuslikkuse ka peale projekti lõppu.

Horisont 2020 AHP esialgne versioon esitatakse hiljemalt 6 kuu jooksul peale projekti algust. AHP vaadatakse läbi ja senine versioon uuendatakse uute andmete lisandumise, projekti muudatuste ja korraliste hindamiste puhul.

AHP koostamisel on soovitatav kasutada veebipõhist vahendit **DMPonline**, mida haldab Suurbritannias asuv Digital Curation Center (DCC). DMPonline pakub erinevate rahastajate nõuetele vastavaid AHP vorme, k.a Horisont 2020 jaoks koostatud vormi. Lisainfo New H2020 DMP guidelines, (12 Aug. 2016, DCC News).

DMPonline võimaldab oma kasutajakonto kaudu loodud plaane muuta, teistega jagada ja täiendada ning alla laadida.

Lisainfo:

- Teabeleht „Teaduse avaandmete pilootprogramm Horisont 2020 projektis“
- Teabeleht teadlasele „Kuidas saab OpenAIRE kasulik olla?“
- Teabeleht projekti koordinaatoritele „Avatud juurdepääs ja avaandmed programmis Horisont 2020“
- Info Euroopa Komisjoni veebilehel: Participant Portal H2020 Online Manual: Data Management
- Info OpenAIRE veebilehel: What is the Open Research Data Pilot

1. ÜLEVAADE ANDMETE KOHTA

Teadusandmete kogumise/genereerimise otstarve, selle seos projekti eesmärkidega.
Andmete päritolu. Kas taaskasutatakse juba olemasolevaid andmeid ja kuidas?
Andmete tüübid ja vormingud. Andmete oodatav maht. Kellele võiksid andmed olla kasulikud?

Andmeid on mitut liiki: digitaalsed ja mittedigitaalsed; numbrilised, kirjeldavad, visuaalsed või kombatavad andmed (nt mõõtmistulemused, algoritmid, laboripäevikud, DNA proovid, füüsilised kollektsioonid, taimeisendid, filmid jne). Teadusandmete hulka üldjuhul ei kuulu haldusandmed, õppematerjalid, teaduspublikatsioonid.

Terviku moodustavaid andmeelemente nimetatakse **andmehulgaks** või andmekogumiks (dataset).

Teadusandmed tuleb esitada võimalikult **töötlemata** kujul (raw/primary data), et võimaldada nende taaskasutamist ja uurimistulemuste kordamist.

Andmetele tuleb lisada ka kirjeldus andmete kogumise kohta ning **dokumentatsioon**, mis võimaldavad kogutud andmeid analüüsida ja mõtestada, nagu näiteks: metoodikakirjeldused, laboripäevikud, küsimustikud, koodiraamatud, tarkvara süntaks, andmebaasi struktuuri kirjeldus, andmesõnastikud, teave seadete ja kalibreerimise kohta jms teave.

Tavapärase laboritöö andmete dokumenteerimise võib jagada kolmele tasandile:

- projekti eesmärkidest, metoodikast ja vahenditest lähtuvalt;
- faili või andmebaasi tasemel (nt kuidas andmehulga failid on omavahel seotud), see on tavaliselt tekstifail nimega readme.txt;
- muutuja või üksuse tasemel (nt määrata tabelarvutuse faili nimetus nõnda, et see selgitaks muutuja tähendust).

Kõik arvutifailid võib jagada kahte kategooriasse - **binaarfailid** ja **tekstifailid**. Lihttekst on masinloetav ja seda toetavad praktiliselt kõik rakendused kõigil arvutiplatvormidel.

Andmete tekstifailina salvestamise eeliseks on see on see, et faili saab lugeda tekstiredaktor, nagu Windows Notepad ja see on inimloetav. Binaarfail ehk kahendvormingus salvestatud fail on loetav ainult arvutile. Binaarfaile juhib kohaldatav tarkvara, mis võib olla omanduslik.

1.1. Avatud vormingud

Avatud vormingud on laialt kasutusel ja omavad parimat võimalust olla loetavad ka tulevikus. Seevastu omandiõigusega kaitstud formaatide puhul, eriti nende puhul, mis on mittestandardsed ja nõuavad konkreetseid tarkvararakendusi või nende spetsiaalseid versioone, võivad perspektiivis tekkida kasutamise takistused. Kiired muutused tehnoloogiaturul põhjustavad failivormingute aegumist. Avatud formaat peab olema realiseeritav nii omandiõigusega kaitstud kui ka avatud lähtekoodiga tarkvaras vastavate tüüplitsentside alusel.

Avatud vormingud on näiteks ODF, TXT, CSV, RTF, HTML, XHTML, PDF, JPEG, PNG, SVG.

TXT – (Text File, tekstifail), mis sisaldab vaid tähti, kirjavahemärke, tühikuid ja reavahetusi, levinuim nimelaiend on .txt. Lihttekstifaile, mis kasutavad ainult ASCII koode, nimetatakse ka ASCII-failideks.

PDF – (Portable Document Format, porditav dokumendivorming) arvuti riist- ja tarkvarast sõltumatu failitüüp, mis säilitab dokumendi väljanägemise nii ekraanil kui väljatrükil.

RTF – (Rich Text Format, rikastatud tekstivorming) alates 1987. a Microsofti arendatav platvormist suhteliselt sõltumatu dokumendi failivorming, mida avab ja salvestab enamik teksitöötlusprogramme. RTF-is tekst erineb lihttekstist selle poolest, et saab kasutada erinevaid fonte ja suurusi, teksti värvi, rasvast ja kaldkirja ja erinevate raamjoontega tabeleid.

CSV – (Comma Separated Values, komaeraldusega väärtused) failivorming, kus iga rida tähendab ühte kirjet ja veerud eraldatakse üksteisest enamasti semikoolonitega või komadega.

XML – (Extensible Markup Language, laiendatav märgistuskeel) on struktureeritud reeglite kogu, mille alusel saab defineerida mistahes andmeid, mida soovitakse veebis vahetada. Saab kasutada mitmel eesmärgil.

JSON - (JavaScript Object Notation) on andmevahetusvorming, mis põhineb JavaScript programmeerimiskeele alamhulgal. JSON on tekstvormingus, sobiv inimlugemiseks ja -kirjutuseks. JSON-ist on saanud populaarne alternatiiv XML-ile.

Metaandmeid kirjeldatakse XML-RDF või JSON vormingus, mis võimaldab luua linktehnoloogia rakendusi.

1.2. Andmete teisendamine

Uurimistöö käigus tekib tõenäoliselt vajadus **andmete teisendamiseks** (nt ühest failiformaadist teise, uute näitajate arvutamine vanade andmete põhjal, teksti muutmine numbriliseks, krüpteerimine, kodeerimine). Põhjuseks võib olla uue tarkvara kasutuselevõtt, soov kindlustada andmefailide loetavus tulevikus, andmete jagamise vajadus jm. Enamik tarkvarapakette võimaldavad eksportida ja vahetada formaate nii, et saab luua tekstifaili impordiks teise programmi. Näiteks Microsoft Excel-is saab salvestada tabeli CSV-failina.

Kvalitatiivsed andmed saab muuta kvantitatiivseteks andmeteks kohaldades tekstilise kodeerimise ja kategoriseerimise tehnikaid. Küsimustikes on vastused tavaliselt kodeeritud numbritena tähemärgistringide asemel.

Andmete **visualiseerimise** eesmärgil võib näiteks lugeja ja nimetajaga suhtarvude asemel esitada andmed protsentidena, kuvades need tulp- või kook-diagrammina.

Andmete salvestamisel ja edastamisel on otstarbekas kasutada **andmete pakkimist**, mis tähendab andmetes korduvate märkide ja mustrite ühekordset kirjeldamist ehk informatsiooni kodeerimist, mille tulemusel võtab andmete salvestamine mälus vähem ruumi. Kasutades tihendusalgoritmile vastavat hõrendusalgoritmi, saab tihendatud andmed algkujule viia. Failide kokku-ja lahtipakkimine on üsna aeganõudev protsess.

Standardne **andmetihendusvorming** failide arhiveerimiseks on ZIP-vorming, mida kasutatakse Windows, Mac, Linux ja Unix platvormidel. Zip on kadudeta vormingu tüüpi, mis tähendab, et fail peaks peale lahtipakkimist olema identne algsega. Enamkasutatavad programmid on Zip, GNU Zip (.gzip, .tar.gz) ja Stuffit. On ka kadudega andmepakkimise algoritme, seotud mõne multimeedia vorminguga.

Andmete edastamise terviklikkust saab kontrollida **kontrollsumma** (checksum) – edastusühiku bittide koguarv – arvutuse abil. Kontrollkoode rakendatakse ka repositooriumite juures andmete ülekandmisel.

Andmebaasi andmete korrastamiseks, liigse kordumise ja sellest tingitud vastuolude vältimiseks tuleks kasutada **andmete normaliseerimist** ehk andmete viimist andmemudeli normaalkujule (statistiline normaliseerimine ja andmebaasi normaliseerimine).

1.3. Valik andmete töötlemise, analüüsimise ja visualiseerimise tööriistu

OpenRefine (varem Google Refine), http://openrefine.org/index.html
Töölauarakendus võimaldab andmete puhastamist, muutmist ühest vormingust teise, väliste andmete lisamist andmestikule.

NodeGoat, http://nodegoat.net/about
Andmetöötluse tööriist, mis võimaldab luua andmemudelil põhinevaid andmekogumeid ning töödelda, analüüsida ja visualiseerida andmestiku objekte relatsiooniliselt, kronoloogiliselt ja ruumiliselt.

RAW Graphs, https://rawgraphs.io/
Andmete visualiseerimise tööriist (arvututabeli andmete visualiseerimiseks suur valik diagramme).

Voyager, http://vega.github.io/voyager
Visuaalse analüüsi tööriist, vt ka https://vega.github.io/vega/about/projects/

Palladio, http://hdlab.stanford.edu/palladio/
Tasuta veebipõhine andmete visualiseerimise platvorm (andmete interaktiivne esitamine nt ajateljel, võrgustikuna).

Carto, https://carto.com/
Asukohaandmete visualiseerimine ja kaartide kujundamine; andmete geokodeerimine.

RegexOne, https://regexone.com/
Regulaaravaldis määratleb mustri, mida saab kasutada otsinguks ja asendamiseks tekstis (kood, logifailid, arvutustabelid, dokumendid).

2. FAIR DATA

FAIR Data põhineb neljal printsiibil, mille alusel andmed peavad olema leitavad, kättesaadavad, koostalitlusvõimelised ja taaskasutatavad.

2.1. Andmete leitavaks tegemine, metaandmetega varustamine

Kirjeldage üldjoontes failidele nime andmise põhimõtet, versioonide käsitlemist, märksõnade määramist, metaandmete loomist.

Kas andmete leitavaks tegemisel on standardlahendusena kasutusel unikaalsed püsiidentifikaatorid, nagu nt DOI (Digital Object Identifiers)?

Kas on kasutusel metaandmete standard? Kui valdkonnapõhiseid metaandmete standardeid ei ole, siis kirjeldada, millist tüüpi metaandmed luuakse ja kuidas.

Failihalduse strateegia kujundamisel tuleks silmas pidada järgmisi elemente: versiooni number; loomise kuupäev; looja nimi; sisukirjeldus; grupi või üksuse nimi, mis on andmetega seotud; avaldamise kuupäev; projekti number. **Faili ja kausta nimemeetod** on võti hästiorganiseeritud kataloogi ja draivi struktuuri pidamiseks. Faili nimi on peamine andmefaili indentifitseerija. Soovitav on rakendada kogu grupile kehtivad juhised failide nimetamisel. Kuna teadusprojekti käigus luuakse arvukalt faile erinevates formaatides ja versioonides, aitab süsteemne ja järjepidev failihaldus andmeid tõhusalt tuvastada ja kasutada, hoides ära andmete “üle pea kasvamise”.

Mõned üldised soovitused:

- Umbes 25 tähemärgist piisab vajaliku info kirjeldamiseks; hoiduda tuleks erisümbolitest faili nimes, sest neid kasutatakse sageli opsüsteemides erikäskude jaoks; kasutada allkriipse tühikute asemel; kuupäevad tuleb vormindada järjepidevalt ühtemoodi;
- kui töö on seotud rohkem kui ühe arvutiga, peavad failid olema sünkroniseeritud;
- suure hulga failide ümbernimetamisel saab enamasti kasutada opsüsteemi sisseehitatud vahendeid (nt digikaamerate automaatselt määratud pildifailide nimed vajavad muutmist).

Üldlevinud viis andmefaili **versioonide** kirjeldamiseks on kasutada järgarve 1, 2 ja 3 suurte versioonimuudatuste ja kümnendkohti väikeste muudatuste jaoks (nt versioon 1.1.) Alati

tuleb jälgitava protsessi muudatused andmetena salvestada, et andmete põhjal oleks võimalik arengut ja muudatusi jälgida.

Teadusandmed peavad olema varustatud standardsete **metaandmetega** ja **unikaalsete digitaalsete identifikaatoritega**.

Metaandmed on struktureeritud informatsioon andmeüksuse kohta, mis võimaldab andmeid kirjeldada ja arhiveerida, andmeid leida ja identifitseerida ning andmeüksusele viidata. Metaandmed lisatakse andmetele nende üleslaadimisel repositooriumisse seal oleva vormi kaudu. Metaandmetel on üha suurenev tähtsus koostöövõimeliste süsteemide ehitamisel.

Märksõnad on publikatsiooni või andmekogu teemat väljendavad ja täpsustavad sõnad või mõisted, mis on saadud sisuanalüüsi tulemusel. **Võtmesõnad** on dokumendi sisu peegeldavad otsingusõnad, mis on saadud tekstianalüüsi tulemusel. Kui märksõna on pärit **kontrollitud sõnastikust** (tesaurusest või ontoloogiast), siis tuleb sellele viidata.

Objekti digitaalidentifikaator (Digital Object Identifier, DOI) on püsiidentifikaator, mida kasutatakse objekti (mis ise võib olla füüsiline või digitaalne) üheselt identifitseerimiseks digitaalses keskkonnas. Identifikaatorid võimaldavad publikatsioone ja andmehulkasid otsida, tuvastada ja linkimise teel seostada. **DOI** on unikaalne (täht)numbriline märgijada, mis koosneb prefiksist ja sufiksist, mida eraldab kaldkriips. Esimeseks teadusandmetele antud DOI-ks on 2004. aastal kliimamudeli katsetusega seotud andmekogu DOI: 10.1594/WDCC/EH4_OPYC_SRES_A2

DataCite on ülemaailmne mittetulundusühing, mis koostöös oma liikmetega annab andmesäilituskeskustele õiguse omistada teadusandmetele DOI-sid. Andmesäilituskeskus (repositoorium) väljastab andmekogule DOI nime ning DataCite hoiustab registreeritud andmekoguga seotud metaandmed, andmekogu veebilehe URL-i ja andmetele määratud DOI-d.

DataCite Eesti konsortsiumi kaudu DOI-süsteemiga liitunud andmehoidlad ja teadustaristud.

DataCite-i integreeritud **otsingumootor** **http://search.datacite.org** võimaldab DOI-dega varustatud teadusandmeid metaandmete põhjal otsida ja selekteerida. Viitamiseks (Cite)

saab valida sobivas viitamisstiilis ja failiformaadis vormistatud viitekirje (nt APA, Harvard, MLA, Vancouver, Chicago, IEEE, BibTeX, RIS).

Metaandmete standardeid on mitmeid, ka osa repositooriume on kehtestanud oma nõuded. DataCite (DataCite Metadata Schema) peab DOI loomisel kohustuslikuks järgmisi metaandmete elemente:

- identifikaator (DOI, mille väljastab andmesäilituskeskus)
- andmete looja nimi
- pealkiri/nimetus,
- kirjastaja/väljaandja
- andmete avaldamise aasta,
- ressursi, teabe tüüp (andmekogu, tekst, pildi- või helifail jm).

Dublin Core Metadata Element Set eristab 15 elementi, vt http://dublincore.org/metadata-basics/

Data Documentation Initiative (DDI) näeb ette veel täiendavaid metaandmete elemente andmete kogumise, muutujate-tasemel kirjelduste ja metoodika kohta, vt http://www.ddialliance.org

Metaandmete standardid on leitavad veel veebilehtedel:

UK Digital Curation Centre: List of Metadata Standards: http://www.dcc.ac.uk/drupal/resources/metadata-standards

The Research Data Alliance Metadata Standards Directory, http://rd-alliance.github.io/metadata-directory

2.2. Andmetele vaba juurdepääsu loomine

Millised andmed tehakse avalikult kättesaadavaks? Kui (osad) andmed jäetakse avalikustamata, siis esitada põhjendus.

Kuidas tehakse andmed kättesaadavaks? Milliseid meetodid või tarkvaravahendid on vajalikud juurdepääsuks andmetele? Kas dokumentatsioon tarkvara kohta on juurde lisatud? Kas on võimalik hõlmata ka avatud lähtekoodiga tarkvara?

Kus andmed, nendega seotud metaandmed, dokumentatsioon ja kood hoiustatakse? Kuidas luuakse juurdepääs juhul kui esineb piiranguid?

Repositoorium (repository) on publikatsioonide ja andmete pikaajalise stabiilse säilitamise keskkond. Repositooriumid omavad võtmerolli andmete leidmisel ja taaskasutamisel. Eristatakse mitut liiki repositooriume: institutsionaalsed, asukohapõhised, valdkonnapõhised, interdistsiplinaarsed.

Teadlased peaksid hindama repositooriumi teenuseid lähtudes sobivusest oma andmetega. Repositooriumi kasutajatele peavad olema esitletud andmete identifitseerimise, juurdepääsu ja kasutamise põhimõtted, k.a repositooriumi säilitamise, andmeturve jm poliitikad.

Repositooriumi valikul peaks aegsasti hankima infot aktsepteeritud failitüüpide, nõutavate metaandmete ja dokumentatsiooni ning ka arhiveerimisega seotud kulude osas. Samuti oleks hea, et andmete kasutaja saaks leitud andmete juures ka viitekirja luua. Repositooriumi valikul tuleks eelistada **OpenAIRE portaaliga** liidestatud andmehoidlaid.

Usaldusväärsed andmehoidlad on varustatud dokumenteeritud poliitikate ja tegevuskordadega, järgides standardeid ja parimaid praktikaid. Repositooriumid saavad oma usaldusväärsust ja läbipaisvust kinnitada auditi ja sertifitseerimise kaudu. Kolm tuntud auditeerimissüsteemi on: DRAMBORA, The Data Seal of Approval ja ISO 16363 Space data and information transfer systems — Audit and certification of trustworthy digital repositories.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Teadusandmete repositooriumide register (Registry of Research Data Repositories) hõlmab ligi 2000 andmehoidlat üle maailma, sisaldades institutsionaalseid, valdkonnapõhiseid ja multidistsiplinaarseid repositooriume, vt **http://www.re3data.org/search**

OpenAIRE

OpenAIRE (Open Access Infrastructure for Research in Europe), https://www.openaire.eu on avatud teaduse edendamiseks loodud e-taristu, mille portaali kaudu püütakse teha kättesaadavaks võimalikult suur osa Euroopas rahastatud teadusprojektidest, publikatsioonidest ja teadusandmetest, k.a Horizon 2020 pilootprojekti **Open Research Data Pilot** projektide teadusandmed, vt lähemalt https://www.openaire.eu/h2020-oa-data-pilot.

OpenAIRE lingib nendele repositooriumitele, mis on OpenAIRE-ga liidestatud ja tehniliselt koostalitlusvõimelised (OpenAIRE compliant) ning avatud juurdepääsuga.

ZENODO, https://zenodo.org/ on OpenAIRE ja CERNi koostöös valminud tsentraliseeritud, multidistsiplinaarne teaduspublikatsioonide ja -andmete repositoorium, mis tasuta, turvaliselt ja võimalikult avatuna säilitab teadustöö tulemused ja varustab need DOI püsiidentifikaatoritega.

Andmete esitamisel tuleb arvestada, kas valitud formaadid võimaldavad andmete jagamist ja pikaajalist juurdepääsu andmetele?

Avatud formaadid on laialt kasutusel ja omavad parimat võimalust olla loetavad ka tulevikus. Seevastu omandiõigusega kaitstud formaatide puhul, eriti nende puhul, mis on mittestandardsed ja nõuavad konkreetseid tarkvararakendusi või nende spetsiaalseid versioone, võivad perspektiivis tekkida kasutamise takistused. Kiired muutused tehnoloogiaturul põhjustavad failivormingute aegumist. Avatud formaat peab olema realiseeritav nii omandiõigusega kaitstud kui ka avatud lähtekoodiga tarkvaras vastavate tüüplitsentside alusel.

Kui andmekaitse ja privaatsuse nõuded takistavad andmete jagamist, siis võib jagada andmeid anonüümistatud kujul, mis eeldab andmekasutuse kokkuleppe sõlmimist. Andmetele saab määrata ka embargo ehk juurdepääsupiirangu aja. Vaata lähemalt piirangute kohta ka p. 2.4.

2.3. Andmete koostalitlusvõime

Andmete koostalitlusvõime hindamine. Täpsustada, milliseid andmete ja metaandmete sõnastikke, standardeid või metodoloogiaid on kavas järgida, et hõlbustada koostalitlusvõimet.

Kas standardsõnavara on kasutusel andmekogumi kõikide andmetüüpide jaoks, et võimaldada interdistsiplinaarset koostalitlusvõimet? Kui ei, kas siis leidub vasteid enamkasutatavatest ontoloogiatest?

Hoiustades teadustöö tulemused repositooriumis, mis ühildub metaandmete standardiga Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH), on tulemused

leitavad ka suurte otsingumootorite ja portaalide, k.a OpenAIRE portaali kaudu. Sellise andmevahetuse eelduseks on, et osapooled rakendavad metaandmete avaldamise ja andmekorje juures nimetatud standardprotokolli OAI-PMH.

Kontrollitud märksõnastik (tesaurus, ontoloogia, taksonoomia) on teemakategooriatesse liigitatud märksõnade korrastatud loetelu, mis sisaldab termineid, äraviitetermineid ja sünonüüme ning hierarhiasuhteid terminite vahel.

Semantilise veebi standarditega vastavuses olevad märksõnastikud on osaks infosüsteemide ja teadmusvõrgustike loomisel. Standard Simple Knowledge Organization System (SKOS) on suunatud just hierarhilise teabe esitamiseks, mida on rakendatud näiteks USA Kongressi Raamatukogu märksõnastiku (Library of Congress Subject Headings) ja Euroopa Liidu mitmekeelse tesauruse Eurovoc puhul. Hea näide linkandmete tehnoloogia kasutamise kohta on ka vabalt allalaetav saksa- ja ingliskeelne Standard-Thesaurus Wirtschaft / STW Thesaurus for Economics, kus deskriptoritele on määratud püsiidentifikaator ning võimalik on linkida teistele märksõnastikele.

Lisainfo tesauruste kohta asub veebi tehnilisi standardeid haldava konsortsiumi W3C (World Wide Web Consortium) Find & Share SKOS Data lehel.

Data Documentation Initiative (DDI) **Alliance** standardid (RDF Vocabularies, Controlled Vocabularies) käsitlevad andmete kirjeldamist, dokumenteerimist ja haldamist, seostatuna andmete elutsükli etappidega. DDI on eelkõige statistiliste ja sotsioloogiliste andmete kirjeldamise standard. Vt lisaks http://www.ddialliance.org/

2.4. Andmete taaskasutamine, litsentsid

Täpsustada, kuidas andmed on litsentsitud, et lubada võimalikult laialdast taaskasutamist.

Millal on andmed kättesaadavad taaskasutamiseks? Vajadusel märkida, miks ja mis perioodiks on ajaline juurdepääsupiirang (embargo) vajalik.

Kas loodud andmed ja/või projektis kasutatud andmed on kasutatavad kolmandate isikute poolt, eriti pärast projekti lõppu? Selgitada põhjus, kui osade andmete taaskasutamine on piiratud.

Määratleda aeg milleni andmed jäävad korduvkasutatavaks.

Kirjeldage andmete kvaliteedi tagamise protsesse.

Kohe teadusprojekti alguses tuleb lahendada kogutavate andmete omandiõiguse ja intellektuaalomandi kaitse küsimused, mis võivad institutsionaalsete ja rahvusvaheliste koostööprojektide puhul olla keerukad (kes, kuidas ja millal andmeid kasutavad ning jagavad). Sel juhul tasub kaaluda Intellektuaalomandi alaste õiguste määratlemist konsortsiumi lepingus.

Intellektuaalomandi alla kuuluva autoriõiguse sisu moodustavad isiklikud ja varalised õigused. Kui isiklikud ehk mittevaralised õigused (õigus autorsusele, teose puutumatusele) kuulub autorile, siis varalised õigused (õigus salvestada, avalikustada, kopeerida jm) kuuluvad üldjuhul ülikoolile. Kui teadlane töötab ülikoolis, siis võivad andmed kuuluda ülikoolile. Institutsionaalne andmehalduse ja varaliste õiguste poliitika võib mõjutada andmete kuuluvust ja jagamist.

Andmed ja faktid iseenesest ei ole autoriõiguse kaitse all. Kui andmekogu loomise protsess on olnud loominguline, siis võib see olla autoriõigusega kaitstav. Rakendades standardlitsentse või selgitades piiranguid kasutustingimustes, edendavad teadlased andmete teadlikku taaskasutamist.

Andmete kasutamise tingimuste määramisel on soovitatav kasutada **avatud sisulitsentse** (open licenses), mille seast levinuim on **Creative Commons** (CC).

2.4.1. Creative commonsi litsentsid

Creative Commons on USA-s asuv mittetulundusühing, mis juba aastaid arendab õiguslikke vahendid loometööde jagamiseks Interneti teel. CC litsentsid võimaldavad teose autoril või autoriõiguste omajal (litsentsiandjal) osast oma õigusest loobuda, et avaldatud teose kasutamine oleks lihtsam (litsentsisaajaks on üldsus).

CC litsentsid on lihtsalt mõistetavad ja masinloetava koodi tõttu koostalitlusvõimelised eri süsteemidega. CC platvormil saab valida sobiva litsentsi: https://creativecommons.org/choose/ .

Igal litsentsil on oma kindel graafiline kujund ja lühend. CC peamised 6 litsentsi lühenditena on:

BY - autorile viitamine

BY-SA - autorile viitamine + jagamine samadel tingimustel

BY-ND - autorile viitamine + tuletatud teoste keeld

BY-NC - autorile viitamine + mitteäriline eesmärk

BY-NC-SA - autorile viitamine + mitteäriline eesmärk + jagamine samadel tingimustel

BY-NC-ND - autorile viitamine + mitteäriline eesmärk + tuletatud teoste keeld

DataCite soovitab teadlastel andmekogumitele määrata Creative Commons Zero (CC0) ehk "no copyright reserved" litsentsi, et võimaldada andmete maksimaalset taaskasutamist, vt https://wiki.creativecommons.org/wiki/CC0_use_for_data. Eelistada võiks ka CC BY, CC BY SA litsentsi.

CC litsentsidest on eestindatud Creative Commonsi litsentsi 3.0 versioon, vt http://hitsa.ee/teenused/autorioigused.

2.4.2. Juurdepääsutingimuste ja piirangute määramine

Kui andmetele ei ole võimalik anda avatud juurdepääsu, on võimalik valida kolme enamlevinud **juurdepääsupiirangu** kehtestamise vahel.

Embargo: juurdepääs andmetele on piiratud teatava ajavahemiku (enamasti 6-12 kuud) jooksul. Piirang tuleneb sageli vajadusest saada piisavalt aega andmetel põhineva publikatsiooni avaldamiseks. Embargo pikkust võib mõjutada rahastaja nõue andmete avaldamiseks ettenähtud ajal või teatud aja jooksul peale projekti lõppemist.

Juurdepääsu võimaldamine ainult autenditud kasutajatele: andmetele juurdepääsuks on nõutud liikmeks olemine teatud uurimisgrupis või asutuses.

Andmete kasutamise lepingud: kokkulepe andmete tootja ja andmete kasutaja vahel, mis võib selgelt esitada reeglid andmete korduvkasutamiseks, hoidmiseks, uuesti levitamiseks ja kustutamiseks.

Taaskasutades **kolmanda osapoole** andmeid, eriti varaliste õigustega andmeid, peab vajadusel küsima õiguste osas andmete tootjalt. Kui andmeid kogunud teadlane on andmete esmane kasutaja, siis järgmised kasutajad võidakse koondada ka mõiste “teisene kasutaja” alla, kes võivad andmeid kasutada avaldatud tulemuste kontrollimiseks, järgmistes uuringutes või hoopis õppetöös.

Andmete omandiõigusega seotud küsimuste korral tuleb kindlaks teha andmete omanik ja tingimused andmete kasutamiseks.

2.4.3. Andmete kvaliteedi tagamine

Riigi Infosüsteemi Ameti (RIA) **andmekvaliteedi tagamise juhend** esitab andmekvaliteedi tunnuste komplekti, mis koosneb 9-st tunnusest (nt ajakohasus, täpsus, ühilduvus, veakäsitlus), mille osas tuuakse välja andmekvaliteedi nõuded, tagamise meetmed ja kontrollküsimused. Kuigi RIA mahukas juhend käsitleb eelkõige riigi infosüsteemi andmekogude haldamist, pakuvad andmekvaliteedi nõuete kontrollküsimused jm RIA materjalid vajalikku teavet kõigile andmekogude koostajatele ja omanikele.

Andmetega seonduvad protsessid tuleb projekti käigus **dokumenteerida** (nt andmete standardne/automatiseeritud kogumine, seadmete kalibreerimine, korduvate proovide ja mõõtmiste tegemine, andmete eksperthindamine).

Laborites kogutavate andmete kvaliteedi ja dokumenteerimise kohta pakub pidepunkte määrus "Hea laboritava nõuded ja kord", kehtestatud kemikaaliseaduse § 5 alusel, vastu võetud 17.12.2015.

Andmed võivad oma digitaalse vormi ja heterogeensuse tõttu põhjustada mitmeid väljakutseid. On mitmed riskid, mis võivad viia andmete kõlbmatuks muutumiseni, nt tehnoloogilise võimekuse puudumine andmete ülekandmiseks füüsiliselt andmekandjalt ning riist – ja tarkvara iganenuks või kasutamiskõlbmatuks muutumine aja jooksul.

3. VAHENDITE JAOTUS, KULUDE HINDAMINE

Kulude hindamine andmete leitavaks, kättesaadavaks, koostalitlusvõimeliseks ja taaskasutatavaks tegemiseks. Kirjeldus kulude katmise kohta.

Projekti andmehalduse eest vastutavate isikute määramine.

Pikaajalise säilitamise kulude ja võimaliku kasu kirjeldamine.

Hinnake andmehalduse tegevuste kulusid aja, tööjõu, seadmete, tarkvara, riistava, IT taristu kasutamise osas. Lisakulusid võivad põhjustada mitmed töölõigud, järgnevalt vaid mõned näited:

- kvalitatiivsete andmete kogumine ja transkribeerimine (intervjuude salvestamine, kirjalikule kujule viimine, tõlkimine);
- dokumentide, fotode jm materjalide digiteerimine, vajadusel ka optilise tekstituvastustarkvara (Optical Character Recognition) ning manuaalse andmesisestuse ja kontrollimeetodi rakendamine;
- andmete ülekandmine ja kaugjuurdepääs turvalise FTP serveri kaudu;
- anonüümistamine, mille peale kuluv aeg sõltub tundlike andmete mahust ja keerukusest, kus õige planeerimine ja meetodite valik vähendab kulusid;
- andmete jagamine ja pikaajaline hoiustamine, kus kulu sõltub andmekeskuse/repositooriumi valikust (nt nõudmised andmete formaadi ja dokumentatsiooni kohta; kas tegemist on DataCite Eesti konsortsiumi kaudu liitunud andmehoidlatega või mitte).

Põhjaliku ülevaate võimalike kulude kohta annab UK Data Service'i tabel Data management costing tool.

Keerulisemate juhtumite puhul, mis on tingitud kas andmete tüübist, mahust või asjaolust, et projekti tegevused on geograafiliselt laiali erinevate asutuste vahel, võib andmete haldamine nõuda ekspertiisi või spetsiaalset varustust. Tegevuste õigeaegne planeerimine ning konsulteerimine ülikooli tugistruktuuridega (IT teenistus, teadusosakond, raamatukogu) aitab ressursse kokku hoida.

Formaalsete protseduuride ja **vastutusalade määramine** aitab tagada andmete autentsuse ja terviklikkuse säilimise. Planeerige vastutusalad kõigi tegevuste kohta (andmete

sisestamine, metaandmete loomine, andmete kvaliteedi tagamine, andmete hoidmine ja varundamine, andmete arhiveerimine ja jagamine) ning lisage ka isikute nimed.

Andmete pikaajaline säilitamine võimaldab vältida sarnaste kordusuuringute tegemist. Arhiveeritud andmeid saab taaskasutada uurimistulemuste kinnitamiseks, uute uuringute läbiviimiseks ning õpetamisel.

4. ANDMETE VARUNDAMINE, TURVALISUS

Plaan süstemaatiliseks varukoopiate tegemiseks andmefailidest, varundamise regulaarsuse ja ulatuse määramine (muutvarunduse puhul varundatakse ainult muutunud failid, täisvarunduse korral kõik failid), varukoopiate säilitamise tähtaja määramine. **Turvameetmete ja standardite** kirjeldamine tundlike andmete kaitseks. **Volitamine**, mis tagab projektis osalejatel juurdepääsu ja sooritamisõiguse ainult kindlaks määratud andmete osas. Uuringuandmete **säilimise tagamiseks** tuleks varukoopiaid hoiustada kolmes regulaarselt hooldatud kohas. Varukoopiad tuleks hoida lähteandmetest eraldi.

Autentsete andmete säilimiseks projekti käigus peaks andmed koondama ühte peafaili (master fail) ja määrama isiku, kellel on õigus faili/andmebaasi nii lugeda kui sinna kirjutada, fikseerides kõik peafaili tehtavad muudatused ning hoides ära tahtmatud muudatused ja kaod andmetes. **Turvaliseks pikaajaliseks hoiustamiseks** on soovitatav kasutada oma asutuse võrgu kaudu käideldatavat võrguketast, varundusserverit ja eelistada tuleks automaatset varundamist. Mistahes tsentraliseeritud varundamisteenuse kasutamise puhul tuleb arvestada säilitamise poliitikaga. Arvutid ning välised mäluseadmed ei sobi andmete, eriti originaalandmete pikaajaliseks säilitamiseks. Lootmine oma arvuti või väliste kõvaketaste peale võib viia andmete riknemise või hävimiseni.

Pilveteenuse (turvalisele internetikettale varundamisel) ehk kaugvarunduse kasutamisel on mitmeid eeliseid. Puudustena võib nimetada järgmist: mõned teenusepakkujad võivad andmeid salvestada väljaspool Euroopa majandusruumi, mis võib olla vastuolus Euroopa andmekaitse direktiiviga ja asutuse või teadusprojekti rahastaja poliitikaga; salvestatud andmed ei pruugi olla täiesti privaatsed kui nad on krüpteerimata; säilib võimalus, et teenusepakkuja loobub äritegevusest.

Tundlik informatsioon on teave, mille avalikustamisel on suur tõenäosus tekitada psühholoogilist, sotsiaalset, emotsionaalset või füüsilist kahju. Kaitsmist vajavad järgmist tüüpi andmed: isiku tuvastamist võimaldavad andmed, kaitstud terviseandmed. Väga tundlikud andmed paberkandjal peaksid olema hoiustatud seifis. Tundlikud andmed tuleks salvestada arvutisse, mis ei ole ühendatud ühegi võrguga. Kui see ei ole võimalik, siis tuleb andmed krüpteerida või arvuti krüpteerida. Samas võib ka selline arvuti sattuda varguse või

andmete kuritahtliku muutmise ohvriks. Tundlikke andmete hoidmisel USB mälupulgal tuleks kasutada krüpteerimise tarkvaraga varustatud seadmeid

Tundliku sisuga andmete eemaldamiseks on kolm peamist võimalust: andmete kustutamine spetsiaalse tarkvara abil (ei pruugi olla parim lahendus pooljuhtketaste ja mälupulkade puhul); demagneetimine, st andmete muutmine loetamatuiks andmekandjale tugeva magnetvälja rakendamise teel (see võib kõvakettad muuta kasutuskõlbmatuks); füüsiline hävitamine lõikumise, purustamise jms teel.

Kui identifitseeritavate andmete kasutamist enam ei vajata, tuleb need hävitada ning tulevased teadusuuringud tuleb teha mitteidentifitseeritavate või anonüümistatud andmete põhjal. See kehtib nii paberkandjal kui ka elektroonilisel kujul kirjete kohta.

Usaldusväärse krüpteerimismeetodi ja **tarkvara** valimine koos vajalike paroolidega on väga oluline. Krüpteerimisel teisendatakse andmed sellisele kujule, mida on võimalik lugeda ainult vastava krüptovõtme või turvalise salasõna olemasolul. See kaitseb andmeid ka arvuti varastamise korral. Krüptograafiliste võtmete, eriti peaparoolide koostamisel on otsustava tähtsusega, et ei koostataks nõrka krüptograafilist võtit. Andmeid ei saa taastada, kui krüpteerimise võti on kadunud. Võtmed võivad kaotsi minna või ununeda, ei pruugi olla enam kättesaadavad (kui võtme omanik on lahkunud), võivad hävineda või kustutatakse kogemata. Kui võtmed ei ole enam kättesaadavad, siis ei saa nendega kaitstud andmeid dekrüpteerida ega nende autentsust kontrollida.

4.1. Krüpteerimise tarkvara

BitLocker Drive Encryption (BDE) on Microsofti kõvaketaste krüpteerimise programm, mis on lisatud Windows Vista (Ultimate, Enterprise), Windows 8.1 (Pro, Enterprise), Windows 10 versioonidele. BitLocker pakub parimat kaitset arvutitele alates TMP (Trusted Platform Module) versioonist 1.2. Kui arvuti ei ole varustatud riistvara komponendiga TPM, siis tuleb salvestada BitLocker-i käivitusvõti irdseadmesse, nt USB mälupulgale.

FileVault on Mac OS-i krüpteerimise meetod, rakendatav alates Mac OS X 10.3 versioonist, FileVault 2 terve kettaseadme meetod alates OS X Lion versioonist. FileVault 2-ga on võimalik krüpteerida ka oma väliseid andmekandjaid (mälupulk, väline kõvaketas), vajalik on andmekandja uuesti formaatimine.

VeraCrypt on vabavaraline tarkvara, mis võimaldab luua krüpteeritud kaustu ja terveid andmekandjaid (USB mälupulk, kõvaketas) ning pakub võimalust andmete krüpteerimiseks nende kasutamise ajal. VeraCrypti eeliseks on avatud lähtekood ning et see on toetatud kõigil enamlevinud platvormidel: Windows, Linux, Mac OS X. Tänu sellele on VeraCryptiga krüpteeritud faile erinevate arvutite ja platvormide vahel lihtne jagada.

DigiDoc3 krüpto (Windows, Mac OS X, Linux) on ID-tarkvara osa, mis sobib failide lühiajaliseks krüpteerimiseks ja nende turvaliseks edastamiseks üle sidekanali. Krüpteerimisel luuakse turvaümbriku fail laiendiga .cdoc ning määratakse inimesed kellega tahetakse dokumenti jagada (adressaadiks võiks määrata ka iseenda). Salastatud info läheb kaotsi, kui adressaadi ID-kaarti ei saa kasutada dekrüpteerimiseks (nt on ID-kaart vahetunud, kadunud, aegunud või on kiip vigastatud).

5. EETILISED ASPEKTID

Kas on eetilisi või juriidilisi küsimusi, mis võivad mõjutada andmete jagamist?

Kas isikuandmetega seotud küsimustikesse on lisatud informeeritud nõusoleku andmine andmete jagamiseks ja pikaajaliseks säilitamiseks?

Kui uuring hõlmab inimeste osalemist, siis peab andmehaldusplaanis olema kirjeldatud nende **konfidentsiaalsuse kaitse**, st strateegiad tundlike andmete käsitlemiseks, hoidmiseks, juurdepääsu piiramiseks ja jagamiseks. Juba andmehalduse planeerimise faasis tuleb läbi mõelda kuidas muuta andmed mitteidentifitseeritavaks ning kas ja kuidas andmeid jagada uurimisprojekti lõppedes.

Uuringu korraldajad peaksid tutvuma **asjakohaste seadustega** enne konfidentsiaalsete andmete kogumist. Eri jurisdiktsioonides on erinevad seadused, mis reguleerivad kuidas konfidentsiaalset informatsiooni hallata ja privaatsust kaitsta.

Uuringus osalevalt inimeselt on andmeid koguma hakates vaja saada informeeritud ehk **teadlik nõusolek** selles, mida tehakse kogutavate andmetega, kes saavad juurdepääsu andmetele ja kuidas andmeid jagatakse.

Levinud meede konfidentsiaalset informatsiooni sisaldavate andmete jagamiseks on nende andmete eelnev **anonüümistamine** ehk isikutuvastusteabe kõrvaldamine või muutmine.

Isiku tuvastamist võimaldavad tunnused võivad olla **otsesed, kaudsed** või **geograafilised**. Otsesed tunnused on nimi, isikukood, telefoninumber, meiliaadress jms. Neid on lihtsam määratleda ja neist hoiduda. Kaudsete tunnuste alla kuuluvad sellised tunnused ja muutujad nagu näiteks rass, palk, sugu, haridus, postiindeks, mis üheskoos esinedes võivad tuua erisused nähtavaks ja identifitseeritavaks. Konfidentsiaalsuse kaitsmine nõuab mitte ainult otseste identifikaatorite, vaid ka kaudsete tunnuste ja näitajate kaalutletud kasutamist ja käsitlemist.

On mitmeid võimalusi andmete anonüümistamiseks, et vähendada võimaliku isiku tuvastamist: eemaldada probleemsed muutujad ja markeerida nende eemaldamine andmekogumist (samas tuleb silmas pidada, et muutujate eemaldamine võib negatiivselt mõjutada andmekogumi kasutamist järgnevates analüüsides);

Rakendada statistilisi meetodeid, näiteks:

- muutuja muutmist (või eemaldamist) ülemise või alumise piiri vahemikus, mis võiksid välja selgitada võõrväärtusi või indiviide (top-coding, bottom-coding);
- kahe või enama muutuja andmete ühendamine üheks koondmuutujaks kokku;
- valimi võtmine, st avaldate juhusliku valimina suurema originaalandmete komplekti, mis pakub uuringu tulemustega võrreldavaid tulemusi (sampling);
- andmete saalimine, st muutujate väärtuste paarikaupa omavahel vahetamine (swapping).

Anonüümistamise kohta täpsema ülevaate saamiseks tuleb uurida **statistilise andmetöötluse meetodeid**. Kasutada saab ka OpenAIRE andmete anonüümistamise tööriista **Amnesia**, mis võimaldab andmetest identifitseerivat teavet eemaldada või teisendada.

KASUTATUD ALLIKAD

Autoriõigused – Creative Commons, http://www.hitsa.ee/teenused/autorioigused

Cybernetica AS. Andmekaitse ja infoturbe leksikon, http://akit.cyber.ee/term/

Datacite Eesti, http://datacite.ut.ee/

DataCite. Registry of Research Data Repositories, http://www.re3data.org/

DataCite Metadata Working Group (2016) DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0. DataCite e.V, http://doi.org/10.5438/0012

Digital Curation Centre (2004) List of Metadata standards, http://www.dcc.ac.uk/resources/metadata-standards/list

Digital Curation Centre (DCC) DMPonline, https://dmponline.dcc.ac.uk/

Dublin Core Metadata Initiative: Metadata Basics, http://dublincore.org/metadata-basics/

Eesti avaliku teabe masinloetava avalikustamise roheline raamat (2014), https://opendata.riik.ee/sites/default/files/manuals/avaliku-teabe-masinloetava-avalikustamise-roheline-raamat-20141125_0.odt

European Commission Participant Portal H2020 Online Manual, http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Horisont 2020: EL teadusuuringute ja innovatsiooni raamprogramm, http://www.horisont2020.ee

International DOI Foundation (2016) Digital object identifier system handbook, https://www.doi.org/hb.html

Lich, B. BitLocker, https://technet.microsoft.com/en-us/itpro/windows/keep-secure/bitlocker-overview

Microsoft (2016) VeraCrypt, https://veracrypt.codeplex.com/

Mäkelä, E. (2017) Using cultural heritage data in research. [Workshop]

OECD Statistics Directorate The OECD glossary of statistical terms, https://stats.oecd.org/glossary/index.htm

Open archives initiative protocol for Metadata harvesting, http://www.openarchives.org/pmh/

Research data domain - CASRAI dictionary (2015), http://dictionary.casrai.org/Category:Research_Data_Domain

Riigi Infosüsteemi Amet (2016) Infoturbe teadlikkuse tõstmise koolituse jaosmaterjal, https://www.ria.ee/public/toetusskeem/koolitusmaterjalid/teadl_Jaotusmaterjalid2016.pdf

Riigi Infosüsteemi Amet. Andmekvaliteedi tagamine, https://www.ria.ee/ee/andmekvaliteedi-tagamine.html

Sihtasutus Eesti Teadusagentuur. Avatud teadus, http://www.etag.ee/tegevused/teemad/avatud-teadus/

SK ID Solutions AS Mis on DigiDoc3 krüpto ning kuidas seda avada Windowsis, Mac OS x-is ja Linuxis, http://www.id.ee/index.php?id=36034

Tartu Ülikooli Raamatukogu. Avatud teadus (Open Science), https://utlib.ut.ee/avatud-teadus-open-science

The University of North Carolina at Chapel Hill and The University of Edinburgh (2016) Research data management and sharing, https://www.coursera.org/learn/data-management

UK Data Service (2016) Costing data management, https://www.ukdataservice.ac.uk/manage-data/plan/costing

Use FileVault to encrypt the startup disk on your Mac, https://support.apple.com/en-us/HT204837

Vallaste, H. (2000) E-teatmik: IT ja sidetehnika seletav sõnaraamat, http://vallaste.ee/index.htm

Whyte, A. (2015) 'Where to keep research data: DCC checklist for evaluating data repositories' v.1.1 Edinburgh: Digital Curation Centre, http://www.dcc.ac.uk/resources/how-guides