

Jesse Tuhkanen

# **Tutkimusdatan hallinta SeAMKin tutkimushankkeissa avoimen TKI-toiminnan näkökulmasta**

Opinnäytetyö

Kevät 2019

SeAMK Tekniikka

Tietotekniikan koulutusohjelma

**SeAMK** 

SEINÄJOEN AMMATTIKORKEAKOULU  
SEINÄJOKI UNIVERSITY OF APPLIED SCIENCES

SEINÄJOEN AMMATTIKORKEAKOULU

## Opinnäytetyön tiivistelmä

Koulutusyksikkö: Seinäjoen Ammattikorkeakoulu

Tutkinto-ohjelma: Tietotekniikan koulutusohjelma

Suuntautumisvaihtoehto: Tietoverkkotekniikka

Tekijä: Jesse Tuhkanen

Työn nimi: Tutkimusdatan hallinta SeAMKin tutkimushankkeissa avoimen TKI-toiminnan näkökulmasta.

Ohjaaja: Pasi Mikkonen

Vuosi: 2019

Sivumäärä: 35

Liitteiden lukumäärä: 1

---

Opinnäytetyö tehtiin Seinäjoen ammattikorkeakoululle. Työn tarkoituksena oli perehtyä parhaisiin datanhallinnan käytänteisiin ja luoda tämän pohjalta ohjeistus SeAMKin henkilöstölle. Ohjeistusta hyödynnettäisiin organisaation TKI-toiminnassa. Tavoitteena oli löytää käytänteitä, jotka auttavat datan avointa jakamista tutkimusprosessin eri vaiheissa.

Työn teoriaosuudessa esitellään työssä käytettäviä datanhallinnan tekniikoita ja ohjeistukseen vaikuttaneita taustatietoja. Käytännön osuudessa esitellään näiden tietojen perusteella syntynyttä ohjeistusta ja työn kulkua. Työn lopputuloksena syntyi ohjeistus, jota voidaan käyttää TKI-hankkeissa datanhallinnan apuna.

Avainsanat: Datanhallinta, TKI-toiminta, Avoin jakaminen,

SEINÄJOKI UNIVERSITY OF APPLIED SCIENCES

## **Thesis abstract**

Faculty: School of Technology

Degree programme: Information Technology

Specialisation: Networking Technology

Author: Jesse Tuhkanen

Title of thesis: Research Data Management in SeAMK Research Projects from the Open RDI Point of View

Supervisor: Pasi Mikkonen

Year: 2019

Number of pages: 35

Number of appendices: 1

---

This thesis was made for Seinäjoki University of Applied Sciences. The purpose of this thesis was to look into the best data management conventions and to create a manual for the personnel of SeAMK based on the gathered knowledge. The instructions would then be utilized in the organization's RDI-projects. The aim was to find different practices that promote open sharing of data at different stages of research projects.

The theory part of the thesis studied different techniques of data management and other background details, which would affect the contents of the manual. The practical part of the thesis described the manual that was generated based on the information presented in the theory part and the workflow of the process. The result of the thesis was a manual that can be used as a data management guideline in RDI-projects in the future.

Keywords: data management, RDI, open sharing

## SISÄLTÖ

Opinnäytetyön tiivistelmä.....	2
Thesis abstract.....	3
SISÄLTÖ.....	4
Kuva- ja taulukkoluetelo .....	<b>Virhe. Kirjanmerkkiä ei ole määritetty.</b>
Käytetyt termit ja lyhenteet .....	7
1 JOHDANTO .....	8
1.1 Työn tausta .....	8
1.2 Työn tavoite .....	8
1.3 Työn rakenne .....	8
2 TKI-TOIMINTA .....	10
2.1 TKI-toiminta ammattikorkeakouluissa .....	10
2.2 Avoin TKI-toiminta Seinäjoen ammattikorkeakoulussa .....	10
3 MITÄ ON DATANHALLINTA .....	12
4 MITÄ AVOIMUUS ON JA MIKSI SITÄ OLISI TÄRKEÄ HARJOITTAA?.....	13
5 AVOIMEN TKI-HANKKEEN VALMISTELU .....	14
5.1 Aineistohallintasuunnitelma.....	14
5.2 DMPTuuli .....	14
6 AVOIMEN TKI-HANKKEEN TOTEUTUSVAIHEEN DATANHALLINNAN VAIHTOEHTOJA.....	16
6.1 Aineiston fyysinen hallinta.....	16
6.1.1 Tiedostojen hallinta ja nimeäminen.....	16
6.1.2 Versionhallinta .....	17
6.1.3 Aineiston dokumentaatio.....	18
6.2 Tallennusmediat.....	20
6.2.1 Oma tietokone.....	20
6.2.2 SSD-levyt.....	20
6.2.3 HDD-levyt.....	21
6.2.4 Verkkosemat .....	22

6.2.5	Magneettinauha .....	23
6.2.6	CD- ja DVD-levyt.....	23
6.2.7	Pilvitallentaminen .....	23
6.2.8	USB-tikut.....	24
6.2.9	Vanhentuneet tallennusmediat.....	24
6.3	Varmuuskopiointi.....	24
6.3.1	Volume Shadow Copy Service.....	25
7	<b>AVOIMEN TKI-HANKKEEN TULOSTEN HALLINTA .....</b>	<b>26</b>
7.1	Pitkäaikaissäilytys ja -saatavuus .....	26
7.2	Konvertointi .....	28
8	<b>Ohjeistuksen tekeminen ja esittely .....</b>	<b>31</b>
9	<b>Yhteenveto ja pohdintaa .....</b>	<b>32</b>
	<b>LÄHTEET .....</b>	<b>33</b>
	<b>LIITTEET .....</b>	<b>35</b>

## **Kuva- ja taulukkoluetelo**

Kuva 1. Avoimen TKI-prosessin viisi päävaihetta. ....	11
Kuva 2. HDD-kovalevyjen vikaantumisasastetta esittävä kuvaaja .....	22
Taulukko 1. Säilytyskelpoisten tiedostomuotojen arviointitaulukko .....	29
Taulukko 2. Yhteenveto säilytyskelpoisten tiedostomuotojen arvioinnista .....	30

## Käytetyt termit ja lyhenteet

<b>Haka</b>	Käyttäjätunnistusjärjestelmä, jonka ansiosta käyttäjä pystyy kirjautumaan kotiorganisaation tunnuksilla moneen eri palveluun.
<b>HDD</b>	Hard disk drive. Mekaaninen tietokoneen massamuisti.
<b>SSD</b>	Solid-state drive. Flash-muistia käyttävä tietokoneen massamuisti, jossa ei ole liikkuvia osia.
<b>TKI-toiminta</b>	Tutkimus-, kehittämis- ja innovaatiotoiminta.

# 1 JOHDANTO

## 1.1 Työn tausta

Seinäjoen Ammattikorkeakoulu tarvitsee lisäohjeistusta siitä, miten tutkimusdataa kannattaa käsitellä, tallentaa ja säilyttää. Organisaation tutkimushankkeiden näkyvyyden ja vaikuttavuuden kannalta ohjeistuksen tekeminen ja avoin jakaminen olisi tärkeää.

## 1.2 Työn tavoite

Työn tavoitteena on luoda ohjeistus, mikä antaa perustietoa siitä, miten tutkimushankkeissa käsiteltä dataa kannattaa tallentaa, käsitellä ja säilyttää datan elinkaaren aikana, jotta mahdollistetaan datan jakaminen. Ohjeistuksen on tarkoitus palvella organisaation tutkimushenkilöstöä selkeyttäen ja linjaten käytäntöjä.

## 1.3 Työn rakenne

Luvussa 2 kerrotaan yleisesti TKI-toiminnasta, sekä TKI-toiminnasta Seinäjoen ammattikorkeakoulussa.

Luvussa 3 selvitetään, mitä datanhallinta tarkoittaa ja miksi sitä olisi tärkeää harjoittaa.

Luvussa 4 kerrotaan, mitä avoimuus tarkoittaa tutkimushankkeissa ja mihin sillä pyritään.

Luku 5 käsittelee datanhallinnan vaiheita TKI-hankkeen valmistelun aikana.

Luvussa 6 esitellään datanhallinnan tekniikoita TKI-hankkeen toteutusvaiheessa.

Luvussa 7 esitellään datanhallinnan kysymyksiä TKI-hankkeen tulosten hallinnan vaiheessa.



Luvussa 8 kerrotaan työn toteutuksesta ja esitellään työn tuloksena syntynyttä ohjeistusta.

## 2 TKI-TOIMINTA

### 2.1 TKI-toiminta ammattikorkeakouluissa

Ammattikorkeakoululaki (L 28.12.2018/1368) määrittää, että ammattikorkeakoulujen tehtäviin kuuluu soveltava tutkimus ja kehittämistyö sekä innovaatiotoiminta (TKI). Ammattikorkeakoulut ovat tärkeä osa suomalaista tutkimus- ja innovaatiojärjestelmää. TKI-toiminta tuottaa ammattikorkeakoulujen yhteistyökumppaneiden kanssa uutta tietoa ja osaamista kehittäen samalla niiden toimintaa. Tärkeimmät sidosryhmät ovat muut tutkimus- ja koulutusorganisaatiot, työ- ja elinkeinoelämä sekä julkishallinto. (Raivo 2017.)

### 2.2 Avoin TKI-toiminta Seinäjoen Ammattikorkeakoulussa

SeAMKin TKI-toiminta painottuu aluetta palvelevaan soveltavaan tutkimukseen, mikä pitää sisällään kehittämistyötä yritysten, opiskelijoiden ja ammattiosaajien kesken. SeAMKin TKI-toiminnan keskeisimmät kumppanit ovat Etelä-Pohjanmaan alueen yritykset ja organisaatiot. (Seinäjoen ammattikorkeakoulu [Viitattu 10.06.2018].) SeAMKin TKI-toiminnan painopisteet ovat ruokaratkaisut, älykkäät ja energiatehokkaat järjestelmät, osallistavat hyvinvointipalvelut sekä yrittäjyys ja kasvu (Seinäjoen ammattikorkeakoulu 2018).

SeAMK on sitoutunut noudattamaan avoimen tieteen ja tutkimuksen periaatteita (Seinäjoen ammattikorkeakoulu [Viitattu 11.06.2018]). Tavoitteena on, että hankkeiden menetelmät, aineistot, tulokset ja tuotokset olisivat kaikkien halukkaiden käytävissä ja TKI-toiminnan laatua, luotettavuutta, näkyvyyttä ja laatua parannettaisiin (Seinäjoen ammattikorkeakoulu 2018).

Avoimen TKI-toiminnan prosessissa on viisi päävaihetta:

1. Avoimen TKI-hankkeen valmistelu
2. Avoimen TKI-hankkeen toteutus
3. Avoimen TKI-hankkeen tulosten hallinta
4. Avoimen TKI-hankkeen tulosten avaaminen ja julkaiseminen

5. Avoimesta TKI:sta liiketoimintaa. (Päällysaho & Latvanen 8.3.2017.)

Prosessi muodostaa kehämäisen jatkumon, missä vaiheet eivät välttämättä ole aina selkeitä tai siirtymät suoraviivaisia. Hyvin toteutettu tutkimusaineiston hallinnointi näiden prosessien eri vaiheissa edistää avoimuuden toteutumista (Päällysaho & Latvanen 8.3.2017.)



Kuva 1. Avoimen TKI-prosessin viisi päävaihetta. (Päällysaho & Latvanen 8.3.2017)

### 3 MITÄ ON DATANHALLINTA

Datanhallinta on niiden toimintatapojen harjoittamista, mitkä helpottavat tutkimuksessa syntyvän datan löytämistä ja ymmärtämistä. Se helpottaa datan käytettävyyttä projektin aikana ja sen jälkeen. (Briney 2015, 1.2.2.)

Oikeilla toimintavaoilla varmistetaan, ettei tutkimusdata vaarannu missään tutkimusprojektin vaiheessa. (Briney 2015, 1.1). Ilman kunnollista datanhallintaa on olemassa suuri riski, että data vaurioituu tai sitä ei pystytä enää lukemaan. (Briney 2015, 1.2.2). Hyvän hallintatavan ansiosta tietoja ei käsittelyprosessin aikana asiatomasti muuteta, tiedot eivät vahingoitu ja tietojen virheettömyyteen ja laatuun voidaan luottaa. (Avoin tiede ja tutkimus [Viitattu 17.7.2017]).

Datanhallinnan osa-alueita ovat aineistonhallintasuunnitelman tekeminen, aineiston dokumentointi ja organisointi, analysointi, sensitiivisen datan suojaaminen, asianmukaisen tallennustilan käyttäminen, datan varmuuskopioiminen ja datan jakaminen. Datanhallintaa pitää sen laajuuden vuoksi harjoittaa ennen tutkimusprojektin alkua, sen aikana ja sen jälkeen. (Briney 2015, 1.2.2.)

Laadukkaan tutkimusaineiston ominaisuuksia ovat käytettävyys, saatavuus, eheys, virheettömyys sekä avoimuus ja luottamuksellisuus. Tutkimustulosten ja -aineistojen on oltava löydettävissä ja käytettävissä myös pidemmällä aikavälillä. (Avoin tiede ja tutkimus [Viitattu 17.7.2017].)

## 4 MITÄ AVOIMUUS ON JA MIKSI SITÄ OLISI TÄRKEÄ HARJOITTA?

Avoimuuden edellytyksenä on laajamittainen pääsy tutkimuksen tuottamiin ja tarvitsemiin julkaisuihin, tietoaineistoihin ja menetelmiin (Avoin tiede ja tutkimus [Viitattu 17.7.2017]). Datanhallinnan harjoittamisen myötä datan jakaminen helpottuu ja datan uudelleenkäyttö mahdollistetaan (Briney 2015, 1.1).

Jakamisen ansiosta tutkimusten toistettavuusongelmat ja resurssien tarve vähenevät ja täten tieteen kehittäminen nopeutuu (Briney 2015, 1.1). Avoimesti saatavilla olevat tulokset ja tietoaineistot mahdollistavat tieteellisten havaintojen kyseenalaistamisen ja verifioimisen, jolloin tieteen luotettavuus kasvaa. Se myös avaa uusia tutkimustulosten analysointimahdollisuuksia. (Briney 2015, 1.1.)

Datanhallinnan ja avoimuuden tärkeyden korostumisen syitä ovat tutkimusdatan digitalisoituminen, minkä ansiosta datan jakaminen on helpompaa ja dataa voidaan jakaa globaalilla mittakaavalla nopeasti. Rahoittajat haluavat myös varmistua, että resursseja käytetään tehokkaasti. (Briney 2015, 1.1.) Tavoitteena on hyvän laadun ja hyvien käytäntöjen edistäminen ja samalla huonolaatuisten tutkimusten ja väärönnösten välttäminen. Avoimen tieteen ja tutkimuksen tarkoituksena on kehittää tutkimuksen kestävyttä, käytettävyyttä, saatavuutta ja luottamusta. (Avoin tiede ja tutkimus [Viitattu 17.7.2017].)

## 5 AVOIMEN TKI-HANKKEEN VALMISTELU

### 5.1 Aineistohallintasuunnitelma

Aineistohallintasuunnitelma on osa tutkimussuunnitelmaa, jonka avulla varmistetaan, että tutkimusaineistojen suhteen noudatetaan hyvää tieteellistä käytäntöä. (Tietoarkisto 2016a).

Rahoittajat vaativat nykyään yhä useammin aineistohallintasuunnitelmaa, sillä he ymmärtävät, että data on haurasta ja ilman sitä aineisto on vaarassa hävitä. Aineistohallinnan vaatiminen on siis rahoittajien tapa varmistua siitä, ettei resursseja mene hukkaan. Aineistohallintasuunnitelma kannattaa tehdä riippumatta siitä vaatiiko rahoittaja sitä, sillä paras tapa parantaa datanhallintaa on suunnitella toimintatavat etukäteen. Tästä johtuen aineistohallintasuunnitelma kannattaa tehdä ennen tutkimusprojektin alkua. Aineistohallintasuunnitelma vastaa yleensä seuraaviin kysymyksiin:

- Minkälaista aineistoa tullaan keräämään
- Kuinka aineisto dokumentoidaan ja organisoidaan
- Miten aineisto säilytetään ja miten se turvataan
- Miten aineistoa hallinnoidaan projektin päätyttyä
- Miten uudelleenkäyttö mahdollistetaan. (Briney 2015, 3.2 - 3.2.2.)

### 5.2 DMPTuuli

DMPTuuli on Tuuli-projektin tuottama työkalu, joka auttaa tutkijaa aineistohallintasuunnitelman kirjoittamisessa. Työkalu tarjoaa ohjeistuksia ja malleja, joiden avulla tutkijat voivat luoda tehokkaan aineistohallintasuunnitelman, jossa otetaan huomioon koko projektin elinkaari. Työkalu sisältää malleja, jotka esittävät eri rahoittajien ja instituutioiden vaatimuksia aineistohallintaan. (DMPTuuli [Viitattu 15.5.2017].) Työkalua voi käyttää apuna tutkimussuunnitelman tekemiseen, rahoituksen hakemiseen ja aineistohallintaan. DMPTuulin on luotu Tuuli-projekti toimesta avoimen

lähdekoodiin perustuvan DMPonline-työkalun pohjalta. (Avoin tiede ja tutkimus [Vii-  
tattu 15.5.2017].)

## **6 AVOIMEN TKI-HANKKEEN TOTEUTUSVAIHEEN DATANHALLINNAN VAIHTOEHTOJA**

### **6.1 Aineiston fyysinen hallinta**

Vastuu aineiston säilytyksestä ja käyttökuntoisuuden varmistamisesta on yleensä tutkimushankkeella sen aktiivivaiheen ajan. Tämän vuoksi on tärkeää, että Tutkijoilla ja tutkimusryhmillä on perustietoa siitä miten fyysinen säilytys kannatta toteuttaa. Näin aineisto säilyy käyttökelpoisena. (Tietoarkisto 2016b). Ilman hyvää ja loogista tiedostojen hallintamenetelmää tiedostot päätyvät herkästi epäjärjestykseen. Hallintamenetelmien pitäisi olla sellaisia, että ne integroituvat sulavasti päivittäiseen tutkimustyöhön ja niiden ylläpitäminen olisi vaivatonta rutiinia. Organisoinnin tärkeys korostuu, kun aineistoa työstetään tutkimusryhmässä. Aineiston järjestyksessä pysyminen varmistetaan siten, että tutkimusryhmän jäsenet sopivat yhteiset hallintatavat. Hallintatavat pitää sisällään tiedostojen organisoinnin, nimeämisen ja versiohallinnan sekä dokumentaation. (Briney 2015, 5.1.1.)

#### **6.1.1 Tiedostojen hallinta ja nimeäminen**

Tutkimusaineistolle luodaan oma hakemisto, johon tallennetaan kaikki aineistoon liittyvät tiedostot, kuten datatiedostot ja kuvailutiedot. Tiedostojen perustiedot tallennetaan kuvailutietojen yhteyteen. (Tietoarkisto 2016b.) Tiedostojen hallinnan kannalta on tärkeää rakentaa looginen kansiorakenne. Tiedostot jaotellaan kansioihin siten, että tietyn aihepiirin tiedostot ovat samassa paikassa. Kansioita luodaan aluksi pieni määrä kattamaan laajempia aihekokonaisuuksia ja näiden sisälle luodaan alikansioita tarkentamaan aihepiiriä. (University of Cambridge [Viitattu 12.7.2017].) Tiedostoja voi jaotella kansioihin esimerkiksi projektin, tutkijan, päivämäärän tai datatyypin mukaan. Oikean ryhmittelytavan valinnassa on tärkeää miettiä, minkä tiedon perusteella tiedostoja tulee tulevaisuudessa etsittyä. (Briney 2015, 5.1.1.)

Johdonmukaisen nimeämistavan käyttäminen täydentää aineistonhallintaa lisäämällä järjestystä kansioiden sisällä. Tällä varmistetaan se, että tiedostot on helposti



löydettävissä myös myöhemmässä vaiheessa projektia. (Briney 2015, 5.2.1.) Nimeämiskäytäntö kannattaa päättää tutkimusprojektin alussa. Kun nimeämiskäytäntö on päätetty, on siitä tärkeää pitää kiinni, jotta kaikki tiedostot on nimetty johdonmukaisesti koko tutkimusprojektin ajan. Nimien tulee olla tarkoituksellisia koko työryhmälle ja niiden tulisi helpottaa tiedostojen löytämistä. Työryhmän olisi hyvä sopia myös sanaston, välimerkkien, päivämäärien ja numeroiden käytöstä. (University of Cambridge [Viitattu 12.7.2017].) Tärkeintä on miettiä mistä nimessä olevista tiedoista on eniten hyötyä itselle. Tällaisessa nimessä pitäisi olla nopeasti ja yksilöivästi tunnistettavissa mitä kyseinen tiedosto sisältää. Peruseriaatteena nimeämisessä on se, että nimet ovat kuvaavia, johdonmukaisia ja mielellään lyhyitä. Erikoismerkkien ja välien käyttöä tulisi välttää. Muuten nimeäminen on hyvin joustavaa. Nimessä voi esiintyä esimerkiksi:

- tutkimuksen tyyppi tai numero
- tutkijan nimi tai nimikirjaimet
- näytteen tyyppi tai numero
- analyysin tyyppi
- päivämäärä
- sijainnin nimi
- versionumero. (Briney 2015, 5.2.1.)

Paras nimeämistapa saattaa vaihdella aineistokohtaisesti, joten eri tapojen käyttäminen on sallittua. (Briney 2015, 5.2.1).

### **6.1.2 Versionhallinta**

Yleensä tutkimusprojekteissa yksittäistä tiedostoa työstää useampi henkilö, jos versionhallinta ei ole, voi syntyä sekaannusta siitä, mikä tiedostoista on viimeisin versio. Versiohallinnan yksi tärkeistä työkaluista on versionumerointi. (University of Cambridge [Viitattu 12.7.2017].) Versionumeroinnin ansiosta usean henkilön työstäessä samaa tiedostoa, ja mahdollisesti eri sijainneissa, tiedetään aina, mistä versiosta on kyse, ja mikä on viimeisin versio. Tämän myötä tiedostojen jakaminen helpottuu, kun tiedostoja voidaan huoletta jakaa tutkimushenkilöiden välillä mene-

mättä sekaisin siitä, mistä versiosta on kysymys. Lisäksi palaaminen tiedoston aikaisempaan tilanteeseen helpottuu. (Briney 2015, 5.2.2.) Versionumeroinnissa jokainen merkittävä muutos tiedoston nimeen voidaan merkitä numeroin, esim. v01 tai v02. Vastaavasti tällä versionumeroinnilla pienet muutokset voitaisiin merkitä nimeen desimaaliluvuin, esim. v01\_01 tarkoittaen, että tiedoston ensimmäiseen versioon on tehty pieni muutos. (University of Cambridge [Viitattu 12.7.2017].) Koska versionumerointi ei itsessään kerro mitä muutoksia dokumenttiin on tehty, kannattaa varsinaisen tiedoston yhteyteen luoda erillinen tiedosto, johon luetellaan, mitä muutoksia missäkin versiossa on tehty. Tähän tarkoitukseen sopii myös kuvailutiedosto. (Briney 2015, 5.2.2.)

### **6.1.3 Aineiston dokumentaatio**

Aineiston dokumentointi on erittäin tärkeää. Se mahdollistaa sen, että muut ymmärtävät aineiston sisällön ja pystyvät sitä myös työstämään. Tämä on tärkeää, jos dataa aiotaan jakaa. Ilman dokumentointia data on pahimmassa tapauksessa täysin merkityksetön lukijalle. Kunnolla tehty dokumentaatio myös osoittaa, että tutkimus on suoritettu hyvien käytänteiden mukaisesti. (Briney 2015, 4.3.) Näiden syiden takia myös hallintamenetelmät (organisointi, nimeämiskäytännöt, versiohallinta) on syytä dokumentoida. Hallintamenetelmät kuvaillaan selkeästi ja ytimekkäästi ja kerrotaan, miten ne toimivat. Hyvä käytäntö hallintamenetelmien kuvailuun on erillinen ylätasoinen kuvailutiedosto. Mikäli käytännöt koskevat alemman tason tiedostoja, avataan hallintatavat niiden yhteyteen tallennetulla kuvailutiedostolla. (Briney 2015, 5.3.2.)

Kuvailutiedosto on helppo ja tehokas tapa dokumentoida tutkimusdataa. Se mahdollistaa datan tarkoituksen ja sisällön ymmärtämisen heti kun aineisto otetaan käsittelyyn. Kuvailutiedostoon kuvaillaan tiedostojen sisältö ja järjestys pääpiirteittäin. Sellainen kannattaa luoda aina tiedostojen yhteyteen, mikäli datan ymmärtämiseen tarvitaan taustatietoja tai se ei ole yksiselitteinen. Kuvailutiedostoja voi luoda monitasoisesti. Projektin ylätasoinen kansiossa voi esimerkiksi lukea yleiset tiedot projektista, kuten:

- Projektin nimi

- Projektin yhteenveto
- Rahoitustiedot
- Yhteys henkilön yhteystiedot
- Oma nimi ja titteli
- Muiden projektissa työskentelevien henkilöiden tiedot
- Datan ja sitä tukevien tietojen sijainnit
- Hallinta ja nimeämiskäytännöt. (Briney 2015, 4.3.1.)

Alakansion kuvailutiedosto kertoo tiedostojen järjestystä ja sisältöä kuvaavia asioita. (Briney 2015, 4.3.1). Yksittäisestä tiedostosta kuvaillaan:

- tiedoston nimi
- missä tiedosto sijaitsee (tiedostopolku)
- tiedoston koko
- tiedostoformaatti
- millä ohjelmalla tiedosto on tuotettu
- milloin tiedosto on tuotettu
- kuka tiedoston on tuottanut
- tiedoston versio
- tiedoston käyttöoikeudet (Tietoarkisto 2015).

Tietohakemisto on hyvä tapa avata paljon muuttujia sisältäviä tietoaineistoa. Tämä on hyödyllinen siksi, että kontekstuaalinen tieto ei yleensä mahdu tiedostoon itsessään. Tavallisesti tietohakemisto sisältää muuttujan seuraavat tiedot:

- muuttujan nimi
- muuttujan määritelmä
- kuinka muuttuja mitattiin
- datayksiköt
- minimi- ja maksimiarvot
- koodatut arvot ja niiden tarkoitus
- puuttuvien arvojen määrittely
- mittauksen tarkkuus
- tunnetut ongelmat
- toisten muuttujien suhde

- muut muistiinpanot datasta. (Briney 2015, 4.3.3.)

Kyseiset tiedot voi kirjata kuvailutiedostoon tai erilliseen tiedostoon. (Briney 2015, 4.3.3).

## **6.2 Tallennusmediat**

Jo ennen tutkimushankkeen alkua oikean tallennusmedian valinnassa tulee miettiä, paljonko tilaa tarvitaan aineiston tallentamiseen. Tämän jälkeen on hyvä miettiä, mitä vaihtoehtoja on jo valmiiksi tarjolla, mikä on itselle riittävän tietoturvan taso yksityisyyden ja datan säilyvyyden kannalta. (Briney 2015, 8.1.3.)

### **6.2.1 Oma tietokone**

Tietokoneen muisti on nopea ja helppo paikka aineiston säilyttämiseen, koska sitä käytetään valmiiksi datan tarkastelussa ja käsittelyssä. Tietokone voi vikaantua, ja kannettavat tietokoneet ovat myös alttiita varkauksille, joten aineiston mahdollinen menetys pitää ottaa huomioon. Tästä huolimatta tietokone on hyvä paikka tallentaa aineisto, mikäli aineiston tallentamiseen käytetään tietokoneen ohella muitakin vaihtoehtoja. (Briney 2015, 8.1.2.)

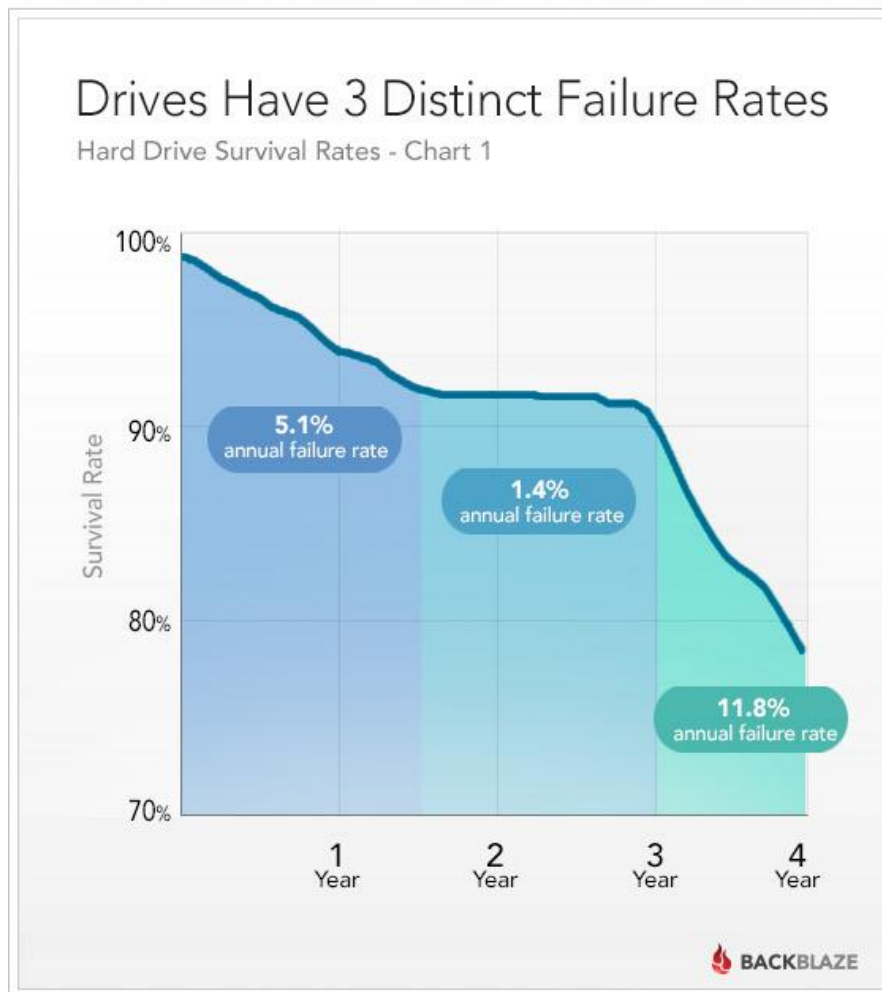
### **6.2.2 SSD-levyt**

SSD-levyissä ei ole mekaanisia liikkuvia osia, kuten HDD-levyissä, joten ne ovat nopeampia, luotettavampia ja niiden iskunkestävyys on parempi. Rajoittava tekijä SSD-levyjen eliniän ja luotettavuuden suhteen on se, että levyn muistisoluihin voi kirjoittaa vain rajallisen monta kertaa. Kuitenkin normaalissa käytössä tämä on harvoin ongelma, sillä SSD-levyn elinaika on pitempi kuin muun tietokoneen. Tietokone vaihdetaan todennäköisesti ennen kuin SSD-levy vikaantuu. (Bauer 2018.)

### 6.2.3 HDD-levyt

Ulkoisten kiintolevyjen muistikapasiteettien kasvu ja hintojen lasku on tehnyt siitä suositun vaihtoehdon. Ulkoinen kiintolevy on hyvä vaihtoehto, sillä varmuuskopioiden tekeminen saadaan toteutettua automatisoidusti. Kiintolevyt kuitenkin vikaantuvat ennen pitkää. Elinkaarena on hyvä pitää viittä vuotta. Kiintolevyjäkin käytettäessä on hyvä pitää varmuuskopioita muissakin paikoissa. (Briney 2015, 8.1.2.)

Yhtiö nimeltä Backblaze seurasi HDD-levyjensä vikaantumisasetta, ja yritys loi vuonna 2013 syntyneen datan pohjalta kuvaajan, mikä kuvaa perinteisien HDD-levyjen vikaantumisasetta niiden ikääntyessä (Kuva 2). Levyt olivat tämän ajan jatkuvassa datan säilytys-, tallennus- ja varmuuskopiointikäytössä. Kuvaajasta nähdään, että ensimmäisen neljän vuoden aikana on nähtävissä kolme selkeää vikaantumisasetta. Ensimmäisen puolentoista vuoden aikana vikaantumisen tahti on 5,1 % vuodessa. Seuraavan puolentoista vuoden aikana vikaantumisen tahti on tasaisempi 1,4 % vuodessa. Tämän jälkeen tahti kiihtyy huomattavasti, jolloin vikaantumisen tahti on 11,8 % vuodessa. Neljän vuoden jälkeen 78 % levyistä oli vielä toimintakunnossa. (Beach 2013.)



Kuva 2. HDD-kovalevyjen vikaantumisasastetta esittävä kuvaaja (Beach 2013.)

#### 6.2.4 Verkkoasemat

Organisaation verkkoasemat ovat yleensä kustannustehokkaita. Verkkoasemien hyvä puoli on se, ettei niistä olla itse vastuussa, niiden käyttökuntauisuudesta vastaa joku muu. Verkkoasemien koot ovat rajallisia, niihin pääsy on yleensä rajoitettua sijainnin puolesta. Jaettu verkkoasema on hyvä, jos sitä ylläpitää pätevä henkilö. (Briney 2015, 8.1.2.)

### **6.2.5 Magneettinauha**

Magneettinauha on nykyään harvinainen tapa datan tallennuksessa, mutta datan säilyvyyden kannalta varmimpia vaihtoehtoja pitkällä aikavälillä. Magneettinauhalla on mahdollista toteuttaa automaattinen varmuuskopiointi. (Briney 2015, 8.1.2.)

### **6.2.6 CD- ja DVD-levyt**

CD- ja DVD-levyt säilyttävät tiedot hyvin, mutta ne ovat erittäin työläitä käytettävyyden ja varsinkin tiedon tallentamisen kannalta. Kun tallennusmedian käyttö on työlästä, on epätodennäköisempää, että varmuuskopioita tallennetaan säännöllisesti. (Briney 2015, 8.1.2.)

### **6.2.7 Pilvitallentaminen**

Pilvitallentamisen hyötyjä on, että se mahdollistaa laajemman pääsyn dataan fyysisen sijainnin ja laitteiden puolesta. Tiedostot on mahdollista synkronoida tietokoneen muistiin. Jotkin pilvipalvelut varmuuskopioivat tallennettuja tiedostoja, jolloin tiedostojen versiohistorian kautta pystytään palauttamaan tietoja. Pilvitallentaminen on parhaimmillaan turvallinen ja luotettava tapa tallentaa aineistoa, mutta tämä riippuu pilvipalvelun tarjoajasta ja sen liiketoimintamallista. Mikäli yhtiö menee konkurssiin tai sen ostaa jokin toinen yhtiö, ei ole takeita siitä, että data on enää saatavissa. Pilvipalvelun ohella täytyy siis käyttää muitakin tallennusvaihtoehtoja. Pilvitallentamista käytettäessä pitää olla erityisen tarkkana yhtiön käyttöehtojen kanssa. Käyttöehdot pitääkin lukea aina ennen palvelun käyttämistä. Tietyissä palveluiden käyttöehdoissa lukee, että kun data ladataan palveluun, annetaan samalla yhtiölle laajat käyttöoikeudet tähän dataan. Henkilötietoja ei saa koskaan ladata pilveen. (Briney 2015, 8.1.2.)

### 6.2.8 USB-tikut

USB-muistitikkuja ei kannata käyttää aineiston säilyttämiseen ollenkaan. USB-tikut hukkuvat helposti, ne ovat erittäin huono vaihtoehto tietoturvan kannalta. USB-muistitikku on suositeltavaa käyttää vain datan siirtämiseen. (Briney 2015, 8.1.2.)

### 6.2.9 Vanhentuneet tallennusmediat

Mikäli tutkimusdataa on vanhentuneella tallennusmedialla esimerkiksi disketillä tai vanhentuneella tietokoneella, kannattaa data kopioida uudemmille tallennusmedioille tai laitteille mahdollisimman nopeasti. Vanhentuneen tallennusmedian avaaminen voi vaikeutua ajan kuluessa huomattavasti ja lopulta se ei välttämättä enää onnistu. (Briney 2015, 8.1.3.)

## 6.3 Varmuuskopiointi

Kun datan varmuuskopio- ja säilytysmenetelmät ovat kunnossa, pystytään suojaamaan datan menetykseltä. Varmuuskopiokäytännön on suojeltava aineistoa eri vahingoilta, kuten vahingossa aineistoon tehdyiltä muutoksilta tai poistamiselta, tallennusmedian tai ohjelmiston aiheuttamilta vikaantumisilta, viruksilta, hakkereilta tai luonnonkatastrofeilta. Varmuuskopiointiin on olemassa paljon vaihtoehtoja, ja aineiston kopioiminen eri tallennusmedioille on yleensä helppoa. Aineiston koko voi joissain tapauksissa aiheuttaa rajoituksia, jolloin on mietittävä eritysratkaisuja. Oman organisaation varmuuskopiokäytäntöjä tulee noudattaa. Varmuuskopioita tulee tehdä säännöllisesti, ja ainakin yksi varmuuskopioista pitää hajauttaa fyysisesti muualle. Tallennusmedian tulee soveltua varmuuskopiointiin ja vanhoja varmuuskopioita ei kannata ylikirjoittaa uusilla. (Tietoarkisto 2016b.)

Sopiva varmuuskopioiden määrä on kolme. Tämä on hyvä tasapaino riittävän suojan ja hallinnan helppouden välillä. Varmuuskopiointi kannattaa toteuttaa siten, että käytetään vähintään kahta eri tallennusmediaa, ja yhtä varmuuskopiota pidetään fyysisesti eri sijainnissa. Kaksi eri tallennusmediaa kompensoivat tietyn tallennusmedian sisältämät riskit, fyysisesti eri paikassa sijaitseva varmuuskopio suojelee



paikallisesti dataa hävittäviä tapahtumia vastaan. Varmuuskopioita tulisi testata aika-ajoin ja opetella palauttamaan tietoja varmuuskopioista. (Briney 2015, 8.1.1.)

### **6.3.1 Volume Shadow Copy Service**

Shadow Copy on palvelu, joka pystyy luomaan esimerkiksi tiedostopalvelimen tiedostoista kopioita tiettyinä ajankohtina. Luotujen kopioiden ansiosta tiedostoista ja kansioista voidaan avata aiempia versioita niistä ajankohdista, jolloin palvelu on kyseisestä sijainnista kopion tehnyt. Tämä on käyttäjälle hyödyllistä, sillä hän voi palauttaa tiedoston, mikäli tiedosto on ylikirjoitettu epähuomiossa. Lisäksi tiedostoon tehtyjä muutoksia voidaan tarkastella tiedoston eri versioita vertailemalla. (Microsoft [Viitattu 9.11.2017].)

Palvelua käytettäessä pitää huomioida, että säännöllisten varmuuskopioiden luomista ei pidä korvata tällä palvelulla. Lisäksi palvelimen tilan täytyttyä vanhin kopio poistetaan, jotta uusia kopioita saadaan luotua. Kun kopio on poistettu, ei kyseisestä ajankohdasta pystytä enää tiedostoja palauttamaan. (Microsoft [Viitattu 9.11.2017].)

## 7 AVOIMEN TKI-HANKKEEN TULOSTEN HALLINTA

### 7.1 Pitkäaikaissäilytys ja -saatavuus

Pitkäaikaissäilytyksen tarkoituksena on säilyttää digitaalisia aineistoja, sekä varmistaa niiden ymmärrettävyys ja luettavuus. Tällöin mahdollistetaan aineistojen hyödynnettävyys muihin tarkoituksiin tulevaisuudessakin. (Kansallinen digitaalinen kirjasto [Viitattu 6.7.2017].) Informaatiota säilytetään useiden kymmenien, jopa satojen vuosien ajan. Vaikka laitteet, ohjelmistot ja informaatio vanhentuisivat, itse informaatio säilytetään. (Avoin tiede ja tutkimus [Viitattu 7.7.2017].) Nyrkkisääntönä on, että dataa kannattaisi säilyttää 5 vuotta, mutta mielellään 10 tai enemmän. Muutaman vuoden välein aineistoa kannattaa käydä läpi ja karsia data, jota ei tarvita. (Briney 2015, 9.1.2.)

Paras tapa valmistella aineisto pitkäaikaissäilytykseen tutkimusprojektin aikana on varmistaa, että aineisto pysyy luettavassa ja tulkittavassa muodossa ja aineistonhallinta ja työstäminen toteutetaan siten, ettei dataa häviä (Briney 2015, 9.2).

Nykyisessä digitalisoituneessa ympäristössä voi moni asia aiheuttaa sen, ettei luotu data ole enää käytettävissä pitkän ajan kuluttua. Aineistolla on yleensä arvoa myös oman tutkimusprojektin ulkopuolella ja siksi on tärkeää, että aineistoa voidaan käyttää vuosienkin päästä. Ohjelmistojen, tiedostoformaattien, tallennusmedioiden ja järjestelmien vanheneminen vuoksi käytettävyyden varmistaminen on haastavaa ja on tärkeää miettiä pitkäaikaissäilytystä koskevia kysymyksiä. Pitkäaikaissäilytys käsittelee projektin jälkeisen ajan, jolloin data ei ole enää aktiivisessa käytössä. Datan säilytys helpottuu, mikäli datan säilyttämiseen liittyvät kysymykset ja päätökset käsitellään jo projektin aikana. Viimeistään siinä vaiheessa, kun projekti on päätymässä, on data valmisteltava tulevaisuutta varten. (Briney 2015, 9.)

Ensimmäisenä tulisi miettiä, mitä säilytetään, ja kauanko sitä säilytetään. Oma organisaatio tai hankkeen rahoittaja saattaa vaatia, että aineistoa täytyy säilyttää tietyn ajan projektin päätyttyä. Kannattaa siis selvittää, onko valmiiksi olemassa olevia säilytyskäytänteitä, jotka sitovat tuotettavaa aineistoa. Useimmissa korkeakouluissa selkeitä käytäntöjä ei kuitenkaan ole. Mikäli käytänteitä on, ne yleensä määrittelevät

vain kauanko dataa säilytetään, mutta ei sitä, että mitä säilytetään. (Briney 2015, 9.1.1.) Tämä onkin henkilön oman harkintakyvyn mukainen asia. Kun päätetään siitä, mitä dataa pitkäaikaissäilytykseen kannattaa viedä, on oltava valikoiva. Kaikkea dataa ei voi eikä kannata säilyttää, sillä tämä kuluttaa liikaa resursseja. Suurimmalle osalle aineistoa tulee piste, jonka jälkeen sitä ei enää tarvita. Data, joka edesauttaa julkaisemista tai raportointia kannattaa säilyttää. Data, joka ei ole toistettavissa tai sen toistettavuus on vaikeaa, kannattaa säilyttää. Data kannattaa säilyttää, mikäli ei osata päättää, säilytettäisiinkö se vai ei. Dataa ei välttämättä kannata säilyttää, mikäli se on helppo toisintaa, eikä sitä tulla suurella todennäköisyydellä käyttämään uudelleen. (Briney 2015, 9.1.2.)

Pitkäaikaissäilytyksen isoin haaste on teknologian nopea kehittyminen. Ohjelmistot ja laitteistot vanhenevat nopeasti, joten data kannattaa tallentaa sellaisessa tiedostomuodossa ja tallennusmediassa, joka on vuosienkin päästä käytettävissä. Saatavilla olevia tiedostomuotoja on varsinkin tutkimusdatalla olemassa lukuisia eri vaihtoehtoja. Tämä on ongelmallista pitkällä aikavälillä, sillä usein tietylle formaatille tarvitaan sille erikoistunut ohjelmisto tiedostojen avaamiseen ja muokkaamiseen. Yhden ohjelmiston eri versiot voivat lisäksi erota niin paljon toisistaan, että uudemmalla versiolla ei saa aikaisemman version tiedostoja auki. Hyvä tapa olisi valita parhaat formaatit projektin alussa, mutta mikäli tämä ei ole mahdollista, tiedostot voi konvertoida myöhemmin. Formaateja, joita kannattaa suosia, perustuvat avoimeen lähdekoodiin, ovat standardisoituja, laajasti käytettyjä ja hyvin dokumentoituja. Tuntemattomia ja suljetun lähdekoodin ohjelmistoja kannattaa välttää. Mikäli käytetyn formaatin avaamiseen on olemassa vain maksullisia ohjelmia, niin formaatin vaihtamista kannattaa harkita. Formaatin valinnassa ei ole olemassa ehdottomia sääntöjä, joten valinnassa on käytettävä omaa arviointikykyään. Mikäli formaattia on vaikea valita, kannattaa suosia vaihtoehtoja sitä, millä on suurempi käyttäjäkunta. Suositumman tiedostomuodon tuki löytyy todennäköisemmin tulevaisuudessakin. (Briney 2015, 9.2.1.)

Pitkäaikaissäilytysjärjestelmään viedään varsinainen tutkimusdata sekä tutkimusdatan metatiedot. Metatiedoilla ja säilytys suunnitelmalla varmistetaan, että in-

formaatio pysyy muuttumattomana. Tiedon hyödyntäminen vaatii informaation säilymisen lisäksi ymmärrettävyyden säilyttämistä sekä säilytysmenetelmien soveltamista. (Avoin tiede ja tutkimus [Viitattu 7.7.2017].)

## 7.2 Konvertointi

Tiedostomuoto ei ole säilytyksen aikana välttämättä lopullinen, tiedostomuotoa voi joutua muuntamaan toiseen muotoon eli konvertoimaan. Tämä voi tulla ajankoh- taiseksi ohjelmiston korvattaessa toisen. Konversio kannattaa tehdä ajoissa, ettei korvattava ohjelmisto ehdi täysin vanhentua, jolloin konversio ei enää välttämättä onnistu. Jos päätetään konvertoida tiedostot toiseen muotoon, kannattaa alkuperäinen data säilyttää vanhassa muodossa, sillä konversiossa voi kadota tietoa, esim. muo- toilu. Konversion tarpeen arviointi kannattaa tehdä silloin, kun suorittaa säännöllistä datan tarkistusta. (Briney, 9.2.1.) Useista ohjelmista löytyy export- tai save as -toi- minto, jonka avulla tiedosto voidaan tallentaa eri tiedostoformaattiin. Kyseiset toi- minnot eivät aina kuitenkaan muunna tiedoston sisältämää informaatiota täydelli- sestä. Tästä johtuen, kun tarvetta konversiolle ilmenee, tiedon katoamisen minimooi- miseksi tulisi tutustua eri tiedostomuotojen ja ohjelmistojen rajoituksiin. (Tietoarkisto 2014.)

Kansallinen digitaalinen kirjasto on arvioinut tiedostomuotojen soveltuvuuden pitkä- aikais säilytyksen ja käytettävyyden kannalta. Arviointi perustuu Yhdysvaltojen kong- ressin, Yhdistyneen kuningaskunnan kansallisarkiston, Alankomaiden kansalliskir- jaston ja Kanadan kansalliskirjaston ja -arkiston selvityksiin. (Kansallinen digitaali- nen kirjasto 2017.)

Arviointikriteerinä on toiminut:

- Avoimuus eli kuinka helposti tiedostomuodosta on saatavilla tietoa.
- Käyttö PAS-standardina eli onko formaattia muodollisesti hyväksytty kan- salliskirjastoissa kansallisarkistoissa ja muissa alan laitoksissa pitkäaikais- säilytyksen välineeksi.
- Vakaus / yhteensopivuus eli onko tiedostomuoto eteen- ja taaksepäin yh- teensopiva, onko korruptoimista vastaan suojauduttu ja miten usein tiedos- tomuodosta julkaistaan korvaavia versioita.

- Riippuvuudet / yhteentoimivuus eli onko formaatti laitteisto- tai ohjelmistosi-  
donnaainen
- Standardisuus eli onko formaattia käynyt läpi standardointiprosessin. (Kansallinen digitaalinen kirjasto 2017.)

Taulukko 1. Säilytyskelpoisten tiedostomuotojen arviointitaulukko (Kansallinen digitaalinen kirjasto 2017.)

<b>A</b>	<b>Arviointikriteeri täyttyy hyvin</b>
<b>A<sup>€</sup></b>	Arviointikriteeri täyttyy hyvin; siihen liittyy kuitenkin kuluja (esim. maksullinen dokumentaatio)
<b>B</b>	Arviointikriteeri täyttyy kohtalaisesti
<b>C</b>	Arviointikriteeri ei täyty
<b>A/C</b>	Arviointikriteeri täyttyy hyvin yhdellä sektorilla (esim. elektroniset vapaakappaleet) muttei toisella sektorilla (esim. asiakirja-aineisto)
<b>A/B</b>	Arviointikriteeri täyttyy hyvin yhdellä sektorilla (esim. elektroniset vapaakappaleet) mutta vain kohtalaisesti toisella sektorilla (esim. asiakirja-aineisto)

Taulukko 2. Yhteenveto säilytyskelpoisten tiedostomuotojen arvioinnista (Kansallinen digitaalinen kirjasto 2017.)

## B.2. Yhteenveto säilytyskelpoisten tiedostomuotojen arvioinnista

Sisältö	Tiedostomuoto	Avoimuus	Käyttö PAS-standardina	Vakaus / yhteensopivuus			Riippumattomuus / yhteentoimivuus	Standardisuus
				Alas-/ylöspäin yhteensopivuuden taso	Korruptoitumisen tieto	Versiopäivitysten määrä		
TEKSTI	Comma Separated Values (CSV)	A	A	A		A	A	A
	Electronic Publications (EPUB)	A	B	B		A	A	A
	Extensible Hypertext Markup Language (XHTML)	A	B	B		A	A	A
	Extensible Markup Language (XML)	A	A	A		A	A	A
	Hypertext Markup Language (HTML)	A	A	B		A	A	A
	Open Document Format (ODF)	A	A	B		B	A	A
	PDF for long-term preservation (PDF/A)	A <sup>€</sup>	A	B		A	A	A
	Tekstitiedosto (plain text)	A <sup>€</sup>	A	B		A	A	A
ÄÄNI	Audio Interchange File Format (AIFF), PCM-koodattu	A	A	A		A	A	A
	Broadcast Wave Format (BWF)	A	A	A		A	A	A
	Free Lossless Audio Codec (FLAC)	A	B	A	A	A	A	A
	MPEG-4 AAC – Advanced Audio Coding (AAC)	A	B	A			A	A
	Waveform Audio Format (WAV)	A	A	A		A	A	A
ELÄVÄ KUVA	Digital Picture Exchange (DPX)	A <sup>€</sup>	A/C		A	A	A	A
	JPEG 2000 sekvenssi	A <sup>€</sup>	A	A	A	A	A	A
	MPEG-4	A	B	A		A	A	A
KUVA	Digital Negative (DNG)	A	B	B		A	A	C
	Joint photographic experts group (JPEG)	A <sup>€</sup>	A			A	A	A
	Joint photographic experts group JPEG 2000 (JP2)	A <sup>€</sup>	A			A	A	A
	Portable network graphics (PNG)	A <sup>€</sup>	A	A		A	A	A
	Tagged Image File Format (TIFF)	A	A			A	A	A
VERKKOARKISTO	Web ARChive Format (WARC)	A	A		B	A	A	A
PAIKKATIETOAINIESTOT	Geographic Tagged Image File Format (GeoTiff)	A	A			A	A	A
	Geography Markup Language (GML)							
	Keyhole Markup Language (KML)							
TUTKIMUSAINIESTOT	SPSS Portable (POR)	A <sup>€</sup>	A	A	A	A	A	C

## 8 Ohjeistuksen tekeminen ja esittely

Ohjeistuksen tekeminen alkoi palavereiden pitämisellä toimeksiantajan kanssa. Palavereissa kartoitettiin ja selvitettiin aihepiiriä. Tämän jälkeen alkoi tarkempi perehtyminen käsiteltyihin asioihin ja tiedonhaku työn teoriaosaan. Tiedonhaun jälkeen alkoi ohjeistuksen tekeminen. Ohjeistuksen sisältö perustuu opinnäytetyön tärkeimpiin teoriaosassa esiteltyihin tietoihin kohdennettuna SeAMKin henkilöstölle, tämän työn tekijän olemassa olevaan tietoon organisaation käytänteistä ja sovittuihin menetelmiin toimeksiantajan kanssa. Tavoitteena oli luoda mahdollisimman lyhyt ja selkeä ohjeistus, jossa kuitenkin olisi riittävä määrä tietoa parempaan datanhallintaan.

Ensimmäisessä luvussa kerrotaan yleisesti aineistohallintasuunnitelmasta ja suositellaan sellainen tekemään. Seuraavaksi kerrotaan yleisesti DMPTuuli-työkalusta, jota suositellaan vaihtoehdoksi aineistohallintasuunnitelman tekemiseen.

Toisessa luvussa esitellään, mitä hallintamenetelmiä tutkimusaineistolle on olemassa, mitä niistä kannattaa käyttää ja mitä niissä tulisi ottaa huomioon. Ohjeistuksessa annetaan ohjeita hakemistojen ja tiedostojen organisointiin, nimeämiseen, versionhallintaan ja dokumentaatioon.

Kolmannessa luvussa esitellään mahdollisia tutkimusdatan tallennusmedioita ja miten ne soveltuvat tutkimusdatan tallennukseen. Kantaa otetaan hyötyihin, haittoihin ja riskeihin.

Neljännessä luvussa kerrotaan, miten toteutetaan riittävä tutkimusdatan varmuuskopiointi ja miten käyttäjä pystyy itse palauttamaan tietoja SeAMKin verkkoasemalta.

Viidennessä kappaleessa esitellään, mitä on pitkäaikaissäilytys ja minkä vuoksi sitä on hyvä tehdä. Pitkäaikaissäilytyksen tärkeimmät kysymykset selvitetään ja annetaan ohjeita hyviin tapoihin.

## 9 Yhteenveto ja pohdintaa

Työssä perehdyttiin parhaisiin datanhallinnan käytänteisiin ja luotiin ohjeistus SeAMKin tutkimushenkilöstöä varten.

Työn haasteena oli aihealueen monitasoisuus: yhtä ongelmaa piti miettiä monesta näkökulmasta. Aiheen ja aihealueiden rajaaminen osoittautui haastavaksi niiden laajuuden vuoksi ja niiden painottaminen oli hankalaa. Vaihtoehtoisuuksien vuoksi ei kovin selkeitä linjauksia voinut ohjeistuksessa vetää. Tämän työn tekijällä ei ollut kokemusta tutkimusprojektityöstä, joten samaistuminen tutkimushenkilöstön arkeen, prosesseihin ja yleisiin toimintatapoihin oli haastavaa. Samaistuminen olisi ohjeistuksen laatimisen kannalta tärkeää. Työn vahva teoreettisuus oli tämän työn tekijälle myös haasteellinen.

Aihe on hyvin ajankohtainen: Tieteen piireissä on alettu korostamaan avoimuutta ja säilytyksen tärkeyttä, mutta samaan aikaan tietosuoja-asetus ja yksityisyyden korostuminen digitaalisessa maailmassa tuovat omat haasteensa.

Ohjeistuksen luomaa hyötyä ja käyttöastetta on vaikea arvioida etukäteen. Digitaalisessa maailmassa muutos on jatkuvaa. Kuten kaikki muukin, tämänkin ohjeistuksenkin tieto vanhenee nopeasti. Ohjeistuksen tuomat valmiudet eivät tule olemaan päteviä kovinkaan pitkään, mutta se luo pohjan, jolle on helppo rakentaa uusia käytänteitä sekä päivittää vanhentunutta tietoa.



## LÄHTEET

- Avoin tiede ja tutkimus. Ei päiväystä. Mitä avoimuus on? [Verkkosivu]. Opetus- ja kulttuuriministeriö. [Viitattu 17.7.2017]. Saatavissa: <http://avointiede.fi/mita-avoimuus-on>
- Avoin tiede ja tutkimus. Ei päiväystä. Mitä on pitkäaikaissäilytys? [Verkkosivu]. Opetus- ja kulttuuriministeriö. [Viitattu 7.7.2017]. Saatavissa: <http://avointiede.fi/mita-on-pitkaaikaissailytys>
- Avoin tiede ja tutkimus. Ei päiväystä. Tuuli – Aineistonhallintasuunnitelmatyökalu. [Verkkosivu]. Opetus- ja kulttuuriministeriö. [Viitattu 15.5.2017]. Saatavissa: <http://avointiede.fi/tuuli>
- Bauer, R. 2018. Hard Disk Drive (HDD) vs Solid State Drive (SSD): What's the Diff?. [Verkkosivu]. Blackblaze. [17.03.2019]. Saatavissa: <https://www.backblaze.com/blog/hdd-versus-ssd-whats-the-diff/>
- Beach, B. 2013. How long do disk drives last. [Verkkosivu]. Blackblaze. [Viitattu 24.02.2019]. Saatavissa: <https://www.backblaze.com/blog/how-long-do-disk-drives-last/>
- Briney, K. 2015. Data Management for Researchers: Organize, maintain and share your data for research success.[Verkkokirja]. Exeter: Pelagic Publishing. [Viitattu 21.11.2017]. Saatavana Ebsco Ebook Collection –tietokannassa. Vaatii käyttöoikeuden.
- DMPTuuli. Ei päiväystä. About DMPTuuli. [Verkkosivu]. Tuuli-projekti. [Viitattu 15.5.2017]. Saatavissa: [https://www.dmptuuli.fi/about\\_us](https://www.dmptuuli.fi/about_us)
- Kansallinen digitaalinen kirjasto. Ei päiväystä. KDK tarjoaa pitkäaikaissäilytyspalvelua. [Verkkosivu]. Opetus- ja kulttuuriministeriö. [Viitattu 6.7.2017]. Saatavissa: <http://www.kdk.fi/fi/pitkaaikaissailytys>
- Kansallinen digitaalinen kirjasto. 2017. Säilytys- ja siirtokelpoiset tiedostomuodot. [Verkkosivu]. Opetus- ja kulttuuriministeriö. [Viitattu 14.11.2017]. Saatavana: <http://www.kdk.fi/images/tiedostot/KDK-PAS-tiedostomuodot-v1.5.1.pdf>
- L 28.12.2018/1368. Ammattikorkeakoululaki.
- Microsoft. Ei päiväystä. Shadow Copies of Shared Folders. [Verkkosivu]. Microsoft Corporation. [Viitattu 9.11.2017]. Saatavissa: [https://technet.microsoft.com/en-us/library/cc771305\(v=ws.11\).aspx](https://technet.microsoft.com/en-us/library/cc771305(v=ws.11).aspx)

- Päälyysaho, S. & Latvanen, J. 8.3.2017. Avoimen TKI-toiminnan prosessimalli. [Verkkolehtiartikkeli]. @SeAMK 8.3.20197. [Viitattu 15.5.2017]. Saatavissa: <http://verkkolehti.seamk.fi/arkisto/maaliskuu-2017-seamk/avoimen-tki-toiminnan-prosessimalli/>
- Raivo, P. 2017. Innovaatioita, kehittämistoimintaa ja tutkimusta: Kaikki kirjaimet käytössä ammattikorkeakoulujen TKI-toiminnassa. [Verkkojulkaisu]. Helsinki: Arene. [Viitattu 11.06.2018]. Saatavana: [http://www.arene.fi/wp-content/uploads/Raportit/2018/arene\\_innovaatioita-kehittamistoimintaa-ja-tutkimusta\\_23032017.pdf?t=1526901760](http://www.arene.fi/wp-content/uploads/Raportit/2018/arene_innovaatioita-kehittamistoimintaa-ja-tutkimusta_23032017.pdf?t=1526901760)
- Seinäjoen ammattikorkeakoulu. 2018. Avoin TKI-toiminta ammattikorkeakouluissa - aineistohallinnan opas: Avoin TKI-toiminta. [Verkkosivu]. Seinäjoen ammattikorkeakoulu. [Viitattu 12.06.2018]. Saatavissa: <https://seamk.libguides.com/avointkitoimintaammattikorkeakouluissa/avoinTKI-toiminta>
- Seinäjoen ammattikorkeakoulu. Ei päivystä. Avoin TKI-toiminta [Verkkosivu]. Seinäjoen ammattikorkeakoulu. [Viitattu 11.06.2018]. Saatavissa: <https://www.seamk.fi/yrityksille/avoin-tki-toiminta/>
- Seinäjoen ammattikorkeakoulu. Ei päivystä. Tutkimus, kehittäminen ja innovointi SeAMKissa. [Verkkosivu]. Seinäjoen ammattikorkeakoulu. [Viitattu 10.06.2018]. Saatavissa: <https://www.seamk.fi/yrityksille/tutkimus-kehittaminen-ja-innovointi-seamkissa/>
- Tietoarkisto. 2014. Tiedostoformaatit ja ohjelmistot. [Verkkosivu]. Yhteiskuntatieteellinen tietoarkisto. [Viitattu 20.3.2018]. Saatavissa: <https://www.fsd.uta.fi/aineistohallinta/fi/tiedostoformaatit-ja-ohjelmistot.html>
- Tietoarkisto. 2015. Aineiston kuvailu ja metadata. [Verkkosivu]. Yhteiskuntatieteellinen tietoarkisto. [Viitattu 23.03.2019]. Saatavissa: <https://www.fsd.uta.fi/aineistohallinta/fi/aineiston-kuvailu-ja-metadata.html#tiedostot>
- Tietoarkisto. 2016a. Aineistohallinnan suunnittelu. [Verkkosivu]. Yhteiskuntatieteellinen tietoarkisto. [Viitattu 15.5.2017]. Saatavissa: <http://www.fsd.uta.fi/aineistohallinta/fi/aineistohallinnan-suunnittelu.html>
- Tietoarkisto. 2016b. Fyysinen säilytys. [Verkkosivu]. Yhteiskuntatieteellinen tietoarkisto. [Viitattu 16.5.2017]. Saatavissa: <http://www.fsd.uta.fi/aineistohallinta/fi/fyysinen-sailytys.html>
- University of Cambridge. Ei päivystä. Organising your data. [Verkkosivu]. University of Cambridge. [Viitattu 12.7.2017]. Saatavissa: <https://www.data.cam.ac.uk/data-management-guide/organising-your-data>

## LIITTEET

Liite 1. Ohjeistus datan hallinnasta SeAMKin tutkimushankkeissa

Jesse Tuhkanen

**Ohjeistus datan hallinnasta SeAMKin  
tutkimushankkeissa**

# SISÄLTÖ

<u>SISÄLTÖ</u> .....	1
<u>1 AINEISTOHALLINTASUUNNITELMA</u> .....	2
<u>DMPTuuli</u> .....	2
<u>2 Digitaalisen tutkimusdatan fyysiset hallintatavat</u> .....	3
<u>2.1 Hakemistot</u> .....	3
<u>2.2 Tiedostojen nimeäminen</u> .....	3
<u>2.3 Versiohallinta</u> .....	5
<u>2.4 Aineiston dokumentaatio</u> .....	6
<u>3 TALLENNUSMEDIAT</u> .....	10
<u>3.1 Oma työkone</u> .....	10
<u>3.2 Ulkoinen kiintolevy ja muut HDD-levyt</u> .....	11
<u>3.3 Tallentaminen organisaation verkkoasemalle (esim. M: -asema)</u> .....	11
<u>3.4 CD- ja DVD-levyt</u> .....	11
<u>3.5 Pilvitallentaminen</u> .....	12
<u>3.6 Muistitikut</u> .....	13
<u>3.7 Vanhentuneet tallennusvaihtoehdot</u> .....	13
<u>4 VARMUUSKOPIOINTI</u> .....	14
<u>4.1 Tiedostojen palauttaminen verkkoaseman varmuuskopiosta (Volume Shadow Copy Service)</u> .....	14
<u>5 PITKÄAIKAISSÄILYTYS</u> .....	18

## 1 AINEISTOHALLINTASUUNNITELMA

Rahoittajat vaativat nykyään yhä useammin, että rahoitushakemuksen liitteenä olevassa tutkimussuunnitelmassa kuvataan aineistohallintasuunnitelma.

Vaikka rahoittaja tai SeAMK eivät vaatisi aineistohallintasuunnitelmien tekemistä, kannattaa sellainen joka tapauksessa tehdä, varsinkin jos suunnittelee jakavansa aineiston avoimesti saataville. Tällöin varmistutaan, että datanhallinta on toteutettu kunnolla ja se on myös tapa osoittaa, että tutkimusaineistojen suhteen on noudatettu hyvää tieteellistä käytäntöä.

### **Aineistohallintasuunnitelma vastaa yleensä seuraaviin kysymyksiin:**

- Minkälaista aineistoa kerätään
- Miten aineisto säilytetään ja miten se turvataan
- Miten aineistoa hallinnoidaan projektin päätyttyä
- Miten uudelleenkäyttö mahdollistetaan

SeAMK:in tutkimusprojekteissa suositellaan, että aineistohallintasuunnitelman luomista varten käytetään DMPTuuli työkalua.

### **DMPTuuli**

DMPTuuli on työkalu aineistohallintasuunnitelman kirjoittamiseen. Työkalu auttaa tutkijaa suunnitelman kirjoittamisessa. Palvelussa voi valita kirjoittamisen avuksi tietyn organisaation ohjeistuksen aineistohallintasuunnitelman kirjoittamiseen sekä rahoittajan tai organisaatioon valmiin kysymyspohjan. Valittu pohja sisältää valmiit kysymykset, mihin kyseisen pohjan laatinut taho vaatii osaltaan vastaamaan. Aineistohallintasuunnitelma muodostuu, kun näihin kysymyksiin vastataan. Valmiin suunnitelman voi tämän jälkeen ladata itselleen.

DMPTuuli on saatavilla osoitteesta: <https://dmptuuli.fi/>

Palvelussa voi liittää Haka-tunnuksen omiin käyttäjätietoihin, joten palveluun pääsee kirjautumaan tämän jälkeen myös SeAMKin käyttäjätunnuksella.

## 2 DIGITAALISEN TUTKIMUSDATAN FYYSISET HALLINTATAVAT

### Hallintapojen valinnassa tulisi ottaa huomioon, että niiden tulisi:

- Integroitua sulavasti jokapäiväiseen tutkimustyöhön
- Olla ylläpidettävissä vaivattomasti ja rutiininomaisesti

Organisoinnin tärkeys korostuu tutkimusryhmässä. Tällöin tutkimusryhmän jäsenet sopivat yhteiset hallintamenetelmät.

### Hallintamenetelmiin kuuluu tiedostojen:

- organisointi
- nimeäminen
- versiohallinta
- dokumentaatio

### 2.1 Hakemistot

Tutkimusaineistolle luodaan oma hakemisto. Hakemistoon tallennetaan itse data-tiedostot ja siihen liittyvät kuvailutiedot. Kansiorakenteen tulisi olla looginen. Tämä saavutetaan siten, että tiedostot jaotellaan kansioiden avulla aihekokonaisuuksiin. Alikansioita käytetään aihepiirin tarkentamiseen.

### Sopivia aihekokonaisuuksia on esimerkiksi:

- Projekti
- Tutkija
- Päivämäärä
- Datatyyppi

### Tiedostojen nimeäminen

Johdonmukainen nimeämistapa lisää järjestystä kansioiden sisällä ja varmistaa sen, että tiedostot ovat helposti löydettävissä.

Nimeämiskäytänteet päätetään tutkimusprojektin alussa tai jo ennen sitä. Tämän jälkeen käytänteistä pidetään kiinni koko tutkimusprojektin ajan, jotta tiedostot on nimetty johdonmukaisesti.

Paras nimeämistapa saattaa vaihdella aineistokohtaisesti, joten eri tapojen käyttäminen on sallittua. Tällöin kannattaa tarkasti dokumentoida käytännöt aineistokohtaisesti. Aiheesta lisää Aineiston dokumentointi –kappaleessa.

#### **Hyvän nimeämistapa:**

- on tarkoituksellinen koko työryhmälle
- helpottaa tiedostojen löytämistä
- yksilöi tunnistettavasti tiedostoja

#### **Nimessä voi esiintyä esimerkiksi:**

- tutkimuksen tyyppi tai numero
- tutkijan nimi tai nimikirjaimet
- näytteen tyyppi tai numero
- analyysin tyyppi
- päivämäärä
- sijainnin nimi
- versionumero

#### **Työryhmän kannattaa sopia myös:**

- Sanastosta
- välimerkkien käytöstä
- päivämäärien käytöstä

Hyvän nimeämistavan löytää helposti, kun miettii mistä nimessä esiintyvistä tiedoista on eniten hyötyä itselle tiedoston löytymiseksi.

#### **Esimerkki:**

Brineyn teoksessa (Briney 2015, 5.2.1.) esitellään esimerkkejä nimeämistavoista. Päivämäärät kannattaa esittää toisin päin, eli käyttää muotoa päivä-kuukausi-vuosi. Esimerkeissä olevalla tavalla kannattaa nimeämismenetelmän säännöt dokumentoida kuvailutiedostoon. Aiheesta lisää Aineiston dokumentointi –kappaleessa.



1. YYYY-MM-DD\_site (date + site name)
2. YYYYMMDD\_ExpmtNum (date + experiment type + experiment number)
3. Species-expmt-num (species name + experiment type + experiment number)
4. Expmt\_Sample (experiment type + sample name)
5. YYYYMMDD\_source\_sample (date + sample source + sample name)
6. ChXXvXX (chapter + version)

Those systems translate into the following file names:

1. “2014-02-15\_Plymouth” and “2014-03-17\_Salem”
2. “19991021\_WesternBlot07” and “20010516\_WesternBlot02”
3. “Vlagopus-obsrv-112” and “Vlagopus-count-067”
4. “UVVis\_KMnO4” and “IRspec\_CH4”
5. “20040812\_AppleLake\_35cm” and “20050812\_AppleLake\_50cm”
6. “Ch04v12” and “Ch05v03”

## 2.2 Versiohallinta

Versionhallinnan ja numeroinnin ansiosta tiedetään aina mistä tiedoston versiosta on kyse. Tämä helpottaa tiedostojen työstämistä ja jakamista tutkimusryhmässä. Palaaminen tiedoston aikaisempaan tilanteeseen helpottuu.

Yksinkertainen, mutta tehokas tapa on merkitä merkittävät muutokset numeroin, esim. v01 ja v02. Pienemmät muutokset voidaan merkitä desimaaliluvuin esim. v01\_01 a v01\_02. Valmiiden tiedostojen nimeen kannatta sisällyttää valmis-merkintä.

### **Esimerkki:**

Brineyn teoksessa (Briney 2015, 5.2.2.) esitellyt esimerkit versionhallinnasta:

Here are a few examples of versioned file names:

- “PlasmaPaper\_v01”, “PlasmaPaper\_v02”, ..., “PlasmaPaper\_v15”, “Plasma-Paper\_FINAL”
- “sample31Spectral1”, “sample31Spectra2”, ..., “sample31SpectraFINAL”
- “RefLib-01”, “RefLib-02”, ..., “RefLib-FINAL”

Kuvailutiedostoon kannattaa luetella mitä muutoksia missäkin versiossa on tehty. Tästä lisää Aineiston dokumentaatio -kappaleessa

### **2.3 Aineiston dokumentaatio**

Digitaalista aineistoa pitää dokumentoida, jotta lukija ymmärtäisi sen sisällön tai, että hän pystyisi aineistoa työstämään tai uudelleenkäyttämään. Dokumentaatio onkin erittäin tärkeää, mikäli aineisto jaetaan avoimesti saataville tai aineistoa työstää tulevaisuudessa joku muu. Mikäli dokumentaatiota ei ole, niin tutkimusdata saattaa olla täysin merkityksetöntä sen lukijalle.

Kun hallintamenetelmät on päätetty, ne on syytä kuvailla. Tätä varten voi luoda erillisen kuvailutiedoston mikä tallennetaan aineiston ylätasolle. Erillinen kuvailutiedosto on helppo ja tehokas tapa dokumentoida ja kuvailla tutkimusaineistoa. Kuvailutiedostoon kuvaillaan tiedostojen sisältö ja järjestys pääpiirteittäin. Se kannattaa luoda aina, mikäli data tarvitsee ymmärtääkseen taustatietoja.

**Projektin yleisen tiedot kannattaa kirjata ylätason kuvailutiedostoon. Esimerkiksi:**

- projektin nimi
- projektin yhteenveto
- rahoitustiedot
- yhteyshenkilön yhteystiedot
- oma nimi ja titteli
- muiden projektissa työskentelevien henkilöiden tiedot
- datan ja sitä tukevien tietojen sijainnit
- hallinta ja nimeämiskäytännöt

#### **Esimerkki:**

Brineyn teoksessa (Briney 2015, 5.3.1.) olevassa esimerkissä ylätason esimerkkikuvailutiedostossa esitellään aluksi, miten kansiorakenne on toteutettu ja kuvataan lyhyesti mitä mistäkin kansioista löytyy. Seuraavaksi esimerkissä esitellään millaisia käytänteitä tiedostojen nimeämisessä on käytetty. Nimeämisessä käytetty kaava

avataan ja käytetyt lyhenteet määritellään. Esimerkissä kuvataan myös käytetty versiohallintatapa. Esimerkissä on myös kerrottu mistä löytyy tiedostojen tarkempi kuvailutiedosto

I use the following conventions for my datasets.

**ORGANIZATION** Datasets are organized into two main folders: “RawData” and “AnalyzedData”.

“RawData” contains the master copies of the data, which is further divided into folders by collection site and then by date.

Any changes to and analysis of the raw data are done on copies of the data in the “AnalyzedData” folder. Data in this folder is grouped by analysis type. README.txt files in each subfolder record which datasets are analyzed, noting both the site and date for the data.

#### NAMING

Raw data files use the following naming convention: “SiteAbbrYYYYMMDD-Num” (site abbreviation + date + sample number).

Examples: “BLA-20000710-2”, “MAM-20010508-1”, “EME-20010627-5” I use these site abbreviations:

“AZU” = Azure Spring

“BLA” = Black Pool

“BOP” = Black Opal Pool

“EME” = Emerald Spring

“MAM” = Mammoth Hot Springs

“SUR” = Surprise Pool

Analyzed files use the following naming convention: “Analysis-Description-vX” (analysis type + brief description of data + version number of file). Analyzed datasets will be further described using README.txt files in relevant subfolders.

Examples: “AtomicEmission-FeContent-v2”, “pH-SURsamples-v3”, “Temp-200105data-FINAL”

In the course of analysis, files should periodically be saved to a new version with an updated version number; final versions should be labeled “FINAL”.

Alakansoiden kuvailutiedostot kuvailevat tiedostojen sisältöä ja järjestystä.

**Yksittäisestä tiedostosta voidaan tarvittaessa kuvailla:**

- tiedoston nimi
- missä tiedosto sijaitsee (tiedostopolku)
- tiedoston koko
- tiedostoformaatti
- millä ohjelmalla tiedosto on tuotettu
- milloin tiedosto on tuotettu
- kuka tiedoston on tuottanut
- tiedoston versio
- tiedoston käyttöoikeudet (Tietoarkisto 2015).

**Esimerkki:**

Brineyn teoksessa (Briney 2015, 4.3.1.) alatasen kuvailutiedoston esimerkissä kerrotaan mitä kyseinen kansio sisältää ja sen jälkeen luetellaan tiedostonimi ja siitä lyhyt kuvaus mitä yksittäinen tiedosto sisältää.

This folder contains sample information from the Green Lake site, summer of 2009. This work was paid for by the NSF.

Subfolders contain the following information:

- “2009GL\_Analysis”: Chemical analysis data from all samples
- “2009GL\_Notes”: Digital scans of the relevant field notes
- “2009GL\_Raw”: Raw data
- “2009GL\_Weather”: Temperature data from all days on site

The file “conventions.txt” explains all naming conventions for the files in this folder.

Mikäli aineisto sisältää paljon muuttujia sisältävää tietoa niin muuttujien tiedot on hyvä kirjata tietohakemistoon. Tietohakemiston tiedot voi halutessaan kirjata myös erilliseen kuvailutiedostoon.

**Tietohakemisto voi sisältää esimerkiksi muuttujan seuraavat tiedot:**

- Muuttujan nimi
- muuttujan määritelmä
- kuinka muuttuja mitattiin
- datayksiköt
- minimi ja maksimiarvot
- koodatut arvot ja niiden tarkoitus
- puuttuvien arvojen määrittely
- mittauksen tarkkuus
- tunnetut ongelmat
- toisten muuttujien suhde
- muut muistiinpanot datasta

**Esimerkki:**

Brineyn teoksen (Briney 2015, 4.3.3.) esimerkissä on kerrottu muuttujan nimi, määritelmä, datayksikkö, tarkkuus, minimi ja maksimiarvo sekä muuta hyödyllistä tietoa.

Variable name: "PMT\_cali"

Variable meaning: Calibrated voltage measurements from the spectrometer's PMT detector. These values correlate to the sample concentration over time

Units: millivolts

Precision: PMT is precise to 0.002 mV

Minimum: minimum value is set as 0 mV

Other information: The PMT is calibrated daily, with raw values given in the "PMT\_raw" column. Sample concentration calculated from the PMT values are in the "SmplConc" column of the spreadsheet

### 3 TALLENNUSMEDIAT

Ennen tallentamisen aloittamista arvioidaan karkeasti, kuinka paljon tilaa aineiston tallentamiseen tarvitaan. Kun tämä on selvillä, niin mietitään mitä tallennusvaihtoehtoja on valmiiksi saatavilla. Mikäli valmiiksi saatavat tallennusmediat eivät riitä kattamaan tutkimuksessa syntyvän datan kapasiteettia tai sen vaatimaa tietoturvan tasoa, tehdään tarvittavat hankinnat / mietitään vaihtoehdot.

Tarjolla olevista tallennusmedioiden vaihtoehdoista tulisi käyttää vähintään kahta. Aiheesta lisää **Varmuuskopiointi** –kappaleessa.

#### 3.1 Oma työkone

- Nopein ja helpoin vaihtoehto
- Dataa voi olla vaikea pitää hyvässä järjestyksessä
- Muistin määrä usein rajoittava tekijä
- Ei kovin tietoturvallinen – alttiina mm. varkauksille ja vikaantumiselle
- Käytettävä tallentamisen ohella muitakin tallennusmedioita

Oma tietokone on helpoin ja nopein vaihtoehto, sillä se on melkein kaikissa tilanteissa valmiiksi saatavilla ja sitä käytetään itse datan työstämiseen. Kuitenkin dataa on vaikea pitää hyvässä järjestyksessä johtuen muista tietokoneen tiedostoista hakemistoista. Muistin määrä on usein rajoittava tekijä. Data itsessään on turvassa käyttäjätunnusten takana, mutta tietokone on alttiina varkauksille ja vikaantumiselle ja täten sitä ei voi pitää kovin tietoturvallisena ratkaisuna. Tämän vuoksi data on tallennettava myös toiselle tallennusmedialle.

Seinäjoen ammattikorkeakoulun työkoneet ovat pääosin kannettavia tietokoneita, joissa on nykyään melkein kaikissa SSD-levyt. Tallennustilana tämä levy on erittäin nopea ja luotettava, joten se on ihanteellinen datan työstämiseen ja tallentamiseen hankkeen aktiivivaiheessa. SSD-levyt ovat normaalissa käytössä pääosin erittäin pitkäikäisiä ja koska työkoneiden vuokra-aika on yleensä neljä vuotta niiden vikaantumisesta tai datan säilyvyydestä ei yleensä tarvitse huolehtia. Työkone vaihtuu suurella todennäköisyydellä ennen kuin nämä asiat tulevat ajankohtaiseksi.

### 3.2 Ulkoinen kiintolevy ja muut HDD-levyt

- Suuri muistikapasiteetin määrä pienellä kustannuksella
- Vikaantuvat ennen pitkää – elinkaaren pituus aktiivikäytössä n. viisi vuotta.

Ulkoinen kiintolevy tai muut HDD-levyt ovat hyvä vaihtoehto datan tallentamiseen, sillä ne ovat tallennustilaansa nähden halpoja vaihtoehtoja. Kuitenkin ne ovat alttiina vikaantumiselle ja perinteisen pyörivän kiintolevyn elinkaarena on hyvä pitää aktiivikäytössä viisi vuotta. Tämä kannattaa pitää mielessä, mikäli kovalevy on ollut käytössä pitkään. Ulkoisen kovalevyn ohella on siis käytettävä muitakin tallennusvaihtoehtoja.

### 3.3 Tallentaminen organisaation verkkoasemalle (esim. M: -asema)

- Kustannustehokas
- Vastuu varmuuskopiointista ja käyttökuntoisuudesta organisaatiolla
- Rajallinen tallennuskapasiteetti

Verkkoasemalle tallentamisessa on lukuisia hyötyjä. Verkkoaseman data on turvassa käyttäjätunnusten takana. Verkkoasemien käyttökuntoisuudesta ei tarvitse vastata itse vaan niistä vastaa SeAMKin tietohallinto. SeAMKissa etätyöyhteys mahdollistaa pääsyn verkkoasemiin käsiksi myös ulkoverkosta. Tähän tarvitaan vain oma työkone, toimiva verkkoyhteys ja etätyöyhteysmäärittäminen. Etätyöyhteysmäärittäminen tehdään oletuksena kaikkiin SeAMKin työkoneisiin. Jaetut verkkoasemat ovat myös erittäin käteviä, kun niitä ylläpitää pätevä henkilö. SeAMKin verkkoasemat ovat varmuuskopioitu ja käyttäjä voi itse palauttaa tästä varmuuskopiosta tietojansa. Aiheesta kerrotaan lisää kappaleessa **Varmuuskopiointi**.

### 3.4 CD- ja DVD-levyt

- Hyvä vaihtoehto tiedon säilyvyyden kannalta
- Erittäin työläitä käytettävyyden ja erityisesti tallentamisen kannalta.

CD- ja DVD-levyt säilyttävät tiedot hyvin, mikäli niitä pidetään fyysisesti hyvässä kunnossa. Kuitenkin datan lukeminen ja etenkin niille tallentaminen on erittäin työlästä. Tämän lisäksi levyjä lukevat asemat alkavat olla harvinaisia etenkin kannettavissa tietokoneissa, eikä levyjä juurikaan SeAMKissa enää käytetä. Tämä vaikeuttaa työstämistä entisestään. Täten CD- ja DVD-levyjen käyttöä ei suositella.

### 3.5 Pilvitalentaminen

- Käyttöehtoihin tutustuttava
- Ei sovi kaikelle datalle

Parhaimmillaan turvallinen ja luotettava tapa tallentaa aineisto, mutta pahimmillaan tämän vastakohta. Erityisen tarkkana pitää olla käytettävän palveluntarjoajan käyttöehtojen kanssa. Pilvipalveluun ladatessa tiedostoja voidaan esimerkiksi antaa palveluntarjoajayhtiölle laajat käyttöoikeudet omaan dataan. Pilvipalveluun ei saa ladata henkilötietoja sisältävää dataa.

SeAMKin henkilöstöllä on valmiiksi saatavilla Office365 pilvipalvelu, jonka OneDrive-sovelluksessa on saatavilla suuri tallennustila 1024 Gt. Palveluun ladatut tiedostot voidaan myös synkronoida oman tietokoneen muistiin, jolloin ne ovat saatavilla kätevästi kansiorakenteessa. Tiedostoihin pääsee käsiksi mistä vain ja miltä laitteelta tahansa. Palvelussa on laajat ja reaaliaikaiset työkalut yhteistyölle ja tiedoston jakamiselle. Palvelu tallentaa versiohistoriaa, joka mahdollistaa versiohistorian tarkastelun tiedostokohtaisesti ja palauttamaan tiedoston aikaisempaan tilanteeseen. Koko OneDriven sisällön pystyy palauttamaan aikaisempaan tilanteeseen.

Ohjeen kirjoitushetkellä Microsoft lupaa, että tietojen oikeudet, omistusoikeus ja in-tressi säilyvät käyttäjällä. Käyttöehdot kannattaa aina kuitenkin lukea ennen palvelun käyttöä.



### 3.6 Muistitikut

Muistitikut eivät sovi lainkaan aineiston säilyttämiseen. Ne ovat erittäin alttiina varkauksille ja hukkumisille. Niitä kannattaa käyttää ainoastaan data siirtämiseen.

Mikäli töitä viedään työpaikan ulkopuolelle, Seinäjoen Ammattikorkeakoulu edellyttää organisaation työntekijöitä tietosuojasyistä suojaamaan USB-muistivälineet BitLocker-salauksella. Tällä saavutetaan myös tutkimusdatalle tarvittava suojauksen taso, joten tikku kannattaa suojata joka tapauksessa. Ohjeet USB-muistivälineen salaukseen löytyy tältä sivulta: [https://share.seamk.fi/tyoryhmat/eu\\_tietosuoja/layouts/15/WopiFrame.aspx?sourcedoc={9882115C-E80F-492C-8175-380D5F85522E}&file=Pieni%20USB%20saattaa%20aiheuttaa%20suuria%20ongelmia%20SeAMK.pdf&action=default](https://share.seamk.fi/tyoryhmat/eu_tietosuoja/layouts/15/WopiFrame.aspx?sourcedoc={9882115C-E80F-492C-8175-380D5F85522E}&file=Pieni%20USB%20saattaa%20aiheuttaa%20suuria%20ongelmia%20SeAMK.pdf&action=default)

### 3.7 Vanhentuneet tallennusvaihtoehdot

Mikäli aineistoa löytyy vanhentuneilta tai vanhenemassa olevilta tallennusmedioilta (esim. levyke), niin data kannattaa siirtää nopeasti turvaan nykyisille medioille. Mitä pidempään asiaa pitkittää, niin sitä epätodennäköisemmäksi ja vaikeammaksi tietojen avaaminen ja siirtäminen käy.

## 4 VARMUUSKOPIINTI

### Oikein toteutettu varmuuskopiointi suojaa aineistoa:

- tahattomilta muutoksilta tai poistamiselta
- tietyn tallennusmedian tai ohjelmiston ongelmilta tai vikaantumiselta
- Viruksilta, hakkereilta, kryptolockereilta
- luonnonkatastrofeilta, tulipaloilta tms.

Varmuuskopiot ovat erittäin tärkeitä tutkimushankkeen kannalta, jotta välttyttäisiin tahattomalta tiedostojen katoamiselta tai muuttamiselta. Mikäli käyttäjä tallentaa SeAMKin verkkoasemalle hänen ei tarvitse huolehtia sen sisältämän tiedon varmuuskopioinnista. Mikäli käytetään muita tallennusratkaisuja, pitää varmistua siitä toteutuuko riittävä varmuuskopiointin taso ja seurata hyviä varmuuskopiointin käytänteitä.

Varmuuskopioita tulee tallentaa säännöllisesti ja vanhoja varmuuskopioita ei kannata ylikirjoittaa uusilla, mikäli tilaratkaisu sen vain suinkin sallii.

Sopiva varmuuskopioiden määrä on kolme. On hyvä käyttää vähintään kahta eri tallennusmediaa ja yhtä kopiota pidetään fyysisesti erillään muista kopioista. Tämä suojaa tietyn tallennusmedian riskeiltä ja fyysisesti erillään oleva kopio suojelee esim. luonnonkatastrofeilta ja tulipaloilta.

Yksinkertaisimmillaan riittävä varmuuskopiointin taso saavutetaan siis esimerkiksi tallentamalla aineisto omalle koneelle, ulkoiselle kovalevyllä ja pilvipalveluun.

Tiedostojen palautustoiminnot (verkkoasema, pilvipalvelu) kannattaa opetella ja varmuuskopioiden käyttökuntoa on seurattava aktiivisesti.

### 4.1 Tiedostojen palauttaminen verkkoaseman varmuuskopiosta (Volume Shadow Copy Service)

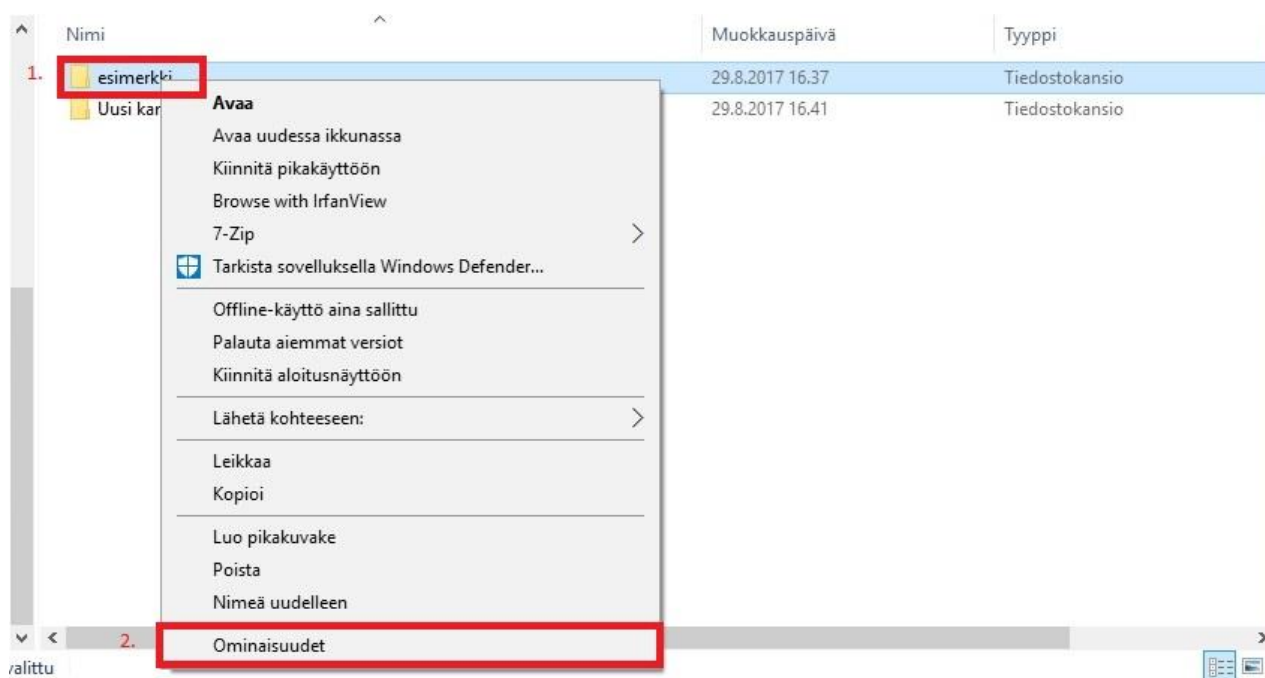
Jos palautettavat tiedostot ovat tallennettu SeAMKin verkkoasemalle, voi käyttäjä yrittää itse palauttaa tiedostoja palvelimella olevan palvelun ansiosta. Yleisimmät tilanteet ovat tiedostojen tahaton poistaminen tai ylikirjoittaminen. Se kuinka pitkältä

aikaväliltä palauttaminen on mahdollista, riippuu kyseisen palvelimen kokonaisu-  
tämäärästä, minne tiedostot ovat tallennettu. Yleensä kuukauden takaisten tiedos-  
tojen palauttaminen on mahdollista.

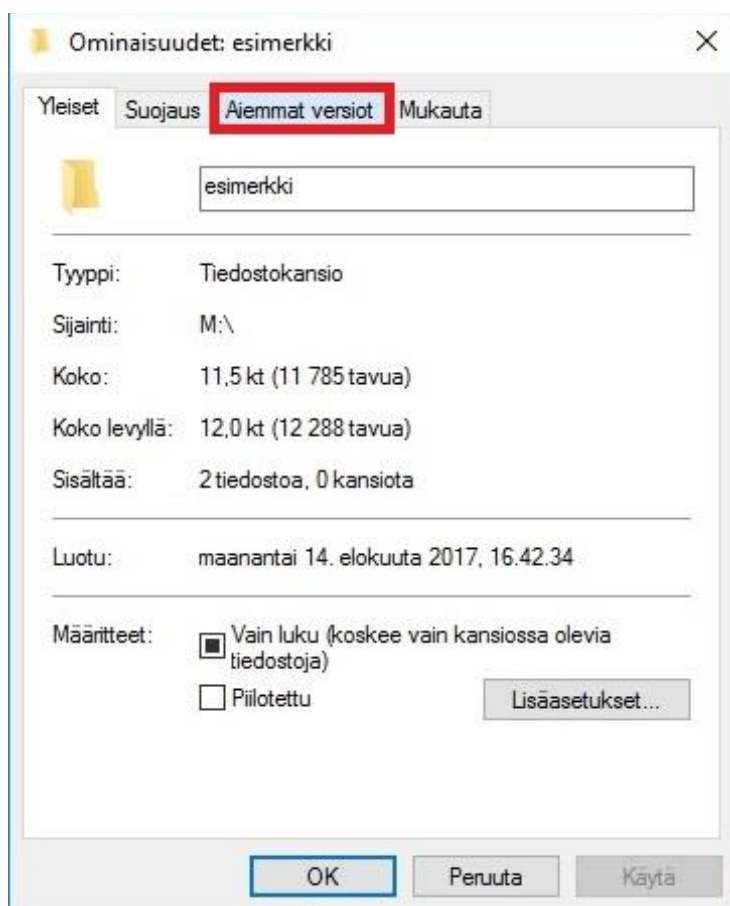
### Palautus tapahtuu näin:

1. painetaan verkkoasemalle tallennetun kansion päältä hiiren oikealla painikkeella

2. Valitaan **Ominaisuudet**

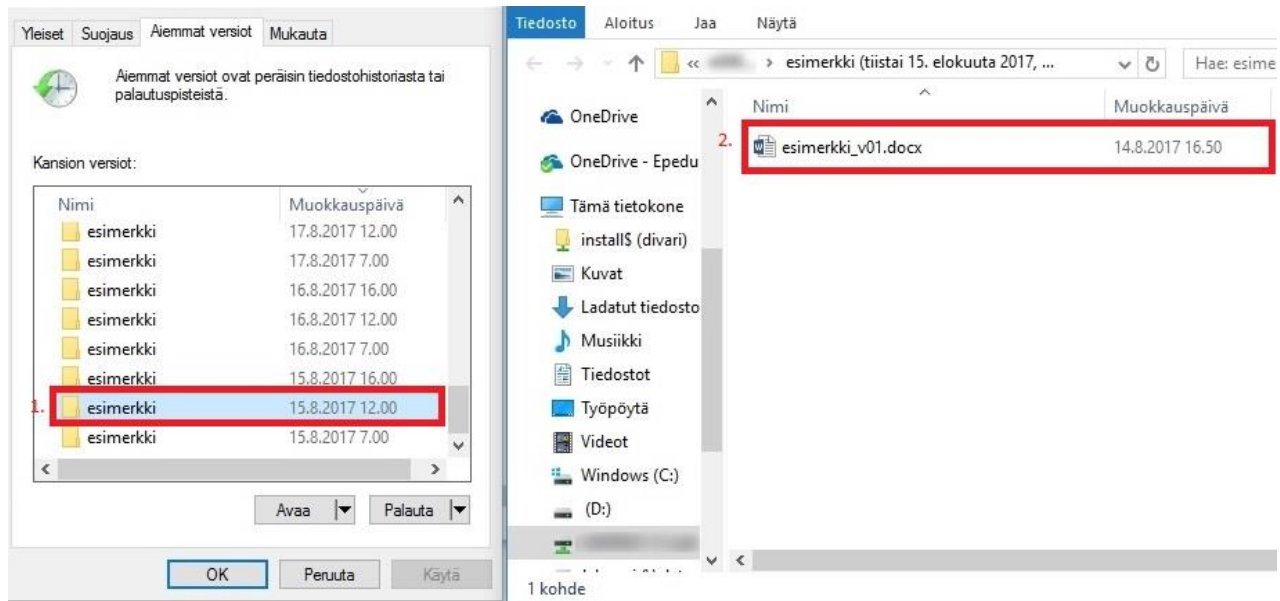


3. Ominaisuudet ikkuna avautuu, josta valitaan **Aiemmat versiot** -välilehti auki.



Aiemmat versiot -välilehdeltä pystyy valitsemaan mistä ajankohdasta tiedosto palautetaan. Shadow Copy tekee tilannevedoksen 3 kertaa päivässä. Ajankohdat ovat klo 7.00, klo 12.00 ja klo 16.00.

1. Kun haluttu ajankohta on löydetty, kansiota kaksoisnapsautetaan tai valitaan Avaa -painike
2. Uusi ikkuna avautuu mistä löytyy kansion sisältö valitusta ajankohdasta. Täältä tiedostot voi avata tarkastelua varten tai siirtää ne talteen.



Mikäli palautusta tarvitaan taaemmalta ajalta, kuin mitä **Aiemmat versiot** –välilehdeltä on saatavilla, otetaan yhteyttä tietohallintohenkilöstöön ja tiedustellaan, onko tiedostoja mahdollista palauttaa taaemmalta ajalta.

## 5 PITKÄAIKAISSÄILYTYS

Aineistolla on arvoa usein oman tutkimusprojektin ulkopuolella. Pitkäaikaissäilytyksen tarkoituksena on mahdollistaa aineiston jatkokäyttö pitkänkin ajan kuluttua. Pitkäaikaissäilytys käsittää projektin jälkeisen ajan, kun data ei ole enää aktiivisessa käytössä. Kun pitkäaikaissäilytykseen liittyvät kysymyksen käsitellään projektin aikana, datan säilytys helpottuu.

Suurin haaste pitkäaikaissäilytykselle on ohjelmistojen, tiedostoformaattien, tallennusmedioiden ja järjestelmien vanheneminen.

Pitkäaikaissäilytyksen liittyvät kysymykset

### **Kauanko säilytetään**

Ensimmäisenä tulee selvittää, onko rahoittajalla tai omalla organisaatiolla vaatimuksia siitä, kauanko dataa tulisi säilyttää. Dataa kannattaisi joka tapauksessa säilyttää vähintään viisi vuotta, mutta mielellään 10 tai enemmän. Säilyttämiselle ei ole ylärajaa.

### **Mitä säilytetään**

Kaikkea projektissa syntynyttä dataa ei kannata välttämättä pitkäaikaissäilyttää, sillä säilytettävä data vaatii seurantaa ja työstämistä, jotta se säilyy luettavana. Jos koko aineisto säilytetään, resursseja kuluu tämän varmistamiseen liikaa. Rajoittavia tekijöitä ovat mahdollisesti myös pitkäaikaissäilytysjärjestelmän koko.

### **Data kannattaa säilyttää mikäli:**

- se edesauttaa julkaisemista tai raportointia
- se ei ole toistettavissa tai sen toistettavuus on hankalaa
- ei osata päättää säilytetäänkö se vai ei

### **Dataa ei välttämättä kannata säilyttää mikäli:**

- Dataa mikä on helposti toistettavissa

- Sitä ei tulla suurella todennäköisyydellä käyttämään uudelleen

### **Missä muodossa säilytetään**

Hyvä tapa olisi valita parhaat tallennusformaatit tutkimusprojektin alussa. Valinnassa on käytettävä omaa arviointikykyä, sillä formaattien valintaan ei ole olemassa ehdottomia sääntöjä tai tiettyjä parhaita formaatteja. Ohjelmistot ja formaatit nimitäin uusiutuvat jatkuvasti.

### **Ohjelmistoa tai tallennusformaattia kannattaa suosia, kun se:**

- Perustuu avoimeen lähdekoodiin
- On standardisoitu
- laajasti käytetty
- hyvin dokumentoitu

### **Ohjelmiston tai tallennusformaatin käyttöä kannattaa välttää mikäli:**

- tuntematon
- suljettuun lähdekoodiin perustuva
- Mikäli olemassa vain maksullisia sovelluksia datan avaamiseen / käsittelyyn

Mikäli ei osaa päättää eri formaattien väliltä, kannattaa valita niistä suositumpi, sillä sen tuki jatkuu todennäköisesti pidempään.

Alla olevaa taulukkoa voi käyttää apuna suuntaa antavasti formaatin valintaan. Jokainen aineisto on kuitenkin erilainen ja ohjelmistot ja formaatit muuttuvat, joten valinta tulee tehdä aina tapauskohtaisesti ja harkiten.

Kansallinen digitaalinen kirjasto on arvioinut tiedostomuotojen soveltuvuuden pitkäaikaissäilytyksen ja käytettävyyden kannalta. Arviointi perustuu Yhdysvaltojen kongressin ja Yhdistyneen kuningaskunnan kansallisarkiston, Alankomaiden kansalliskirjaston ja Kanadan kansalliskirjaston ja arkiston selvityksiin.

Arviointikriteerinä on toiminut:

- Avoimuus, eli kuinka helposti tiedostomuodosta on saatavilla tietoa.

- Käyttö PAS-standardina, eli onko formaattia muodollisesti hyväksytty kansalliskirjastoissa, kansallisarkistoissa ja muissa alan laitoksissa pitkäaikais säilytyksen välineeksi.
- Vakaus / yhteensopivuus, eli onko tiedostomuoto eteen- ja taaksepäin yhteensopiva, onko korruptoimista vastaan suojauduttu ja miten usein tiedostomuodosta julkaistaan korvaavia versioita.
- Riippuvuudet / yhteentoimivuus, eli onko formaatti laitteisto- tai ohjelmistosiidonainen
- Standardisuus, eli onko formaattia käynyt läpi standardointiprosessin (Kansallinen digitaalinen kirjasto 2017.)

<b>A</b>	<b>Arviointikriteeri täyttyy hyvin</b>
<b>A<sup>€</sup></b>	Arviointikriteeri täyttyy hyvin; siihen liittyy kuitenkin kuluja (esim. maksullinen dokumentaatio)
<b>B</b>	Arviointikriteeri täyttyy kohtalaisesti
<b>C</b>	Arviointikriteeri ei täyty
<b>A/C</b>	Arviointikriteeri täyttyy hyvin yhdellä sektorilla (esim. elektroniset vapaakappaleet) muttei toisella sektorilla (esim. asiakirja-aineisto)
<b>A/B</b>	Arviointikriteeri täyttyy hyvin yhdellä sektorilla (esim. elektroniset vapaakappaleet) mutta vain kohtalaisesti toisella sektorilla (esim. asiakirja-aineisto)



## B.2. Yhteenveto säilytyskelpoisten tiedostomuotojen arvioinnista

Sisältö	Tiedostomuoto	Avoimuus	Käyttö PAS-standardina	Vakaus / yhteensopivuus			Riippumattomuus / yhteentoimivuus	Standardisuus
				Alas-/ylöspäin yhteensopivuuden taso	Korruptoitumisen sieto	Versiopäivitysten määrä		
TEKSTI	Comma Separated Values (CSV)	A	A	A		A	A	A
	Electronic Publications (EPUB)	A	B	B		A	A	A
	Extensible Hypertext Markup Language (XHTML)	A	B	B		A	A	A
	Extensible Markup Language (XML)	A	A	A		A	A	A
	Hypertext Markup Language (HTML)	A	A	B		A	A	A
	Open Document Format (ODF)	A	A	B		B	A	A
	PDF for long-term preservation (PDF/A)	A <sup>ε</sup>	A	B		A	A	A
	Tekstitiedosto (plain text)	A <sup>ε</sup>	A	B		A	A	A
ÄÄNI	Audio Interchange File Format (AIFF), PCM-koodattu	A	A	A		A	A	A
	Broadcast Wave Format (BWF)	A	A	A		A	A	A
	Free Lossless Audio Codec (FLAC)	A	B	A	A	A	A	A
	MPEG-4 AAC – Advanced Audio Coding (AAC)	A	B	A			A	A
	Waveform Audio Format (WAV)	A	A	A		A	A	A
ELÄVÄ KUVA	Digital Picture Exchange (DPX)	A <sup>ε</sup>	A/C		A	A	A	A
	JPEG 2000 sekvenssi	A <sup>ε</sup>	A	A	A	A	A	A
	MPEG-4	A	B	A		A	A	A
KUVA	Digital Negative (DNG)	A	B	B		A	A	C
	Joint photographic experts group (JPEG)	A <sup>ε</sup>	A			A	A	A
	Joint photographic experts group JPEG 2000 (JP2)	A <sup>ε</sup>	A			A	A	A
	Portable network graphics (PNG)	A <sup>ε</sup>	A	A		A	A	A
	Tagged Image File Format (TIFF)	A	A			A	A	A
VERKKOARKISTO	Web ARChive Format (WARC)	A	A		B	A	A	A
PAIKKATIETOAINIESTOT	Geographic Tagged Image File Format (GeoTiff)	A	A			A	A	A
	Geography Markup Language (GML)							
	Keyhole Markup Language (KML)							
TUTKIMUSAINIESTOT	SPSS Portable (POR)	A <sup>ε</sup>	A	A	A	A	A	C

Tiedostoja voi joutua muuttamaan projektin tai pitkäaikaissäilytyksen aikana toiseen tiedostomuotoon eli konvertoimaan. Tämä tulee ajankohtaiseksi, mikäli edellinen tiedostomuoto vanhenee tai sen korvaa joku toinen tiedostomuoto. Tällöin konvertointi kannattaa tehdä hyvissä ajoin ennen kuin ohjelmisto vanhenee täysin. Yleisin konversion muoto on käyttää eri ohjelmien Save as tai export –toimintoja. Konvertointi ei välttämättä muuta tietoa täydellisesti. Esimerkiksi muotoilu voi kärsiä. Tämän vuoksi kannattaa säilyttää myös alkuperäinen data, mikäli mahdollista.