

Tiedeinstituuttien avoin tutkimusdata

Toimintamalli digitaaliseen aineistohallintaan

Harri Kiiskinen, Laura Nissin ja Manna Satama

Toimintamalli digitaalisen aineistohallintaan
2020

Harri Kiiskinen, Laura Nissin ja Manna Satama

Tiedeinstituuttien avoin tutkimusdata -hanke

Institutum Romanum Finlandiae sr

Suomen Ateenan-instituutin säätiö sr

Suomen Lähi-idän instituutin säätiö sr

Suomen Japanin-instituutin säätiö sr

DOI: 10.5281/zenodo.4113921

Asiasanat (YSA): avoin tieto; metadata; tutkimusaineisto; humanistiset tieteet; kuvailu;
projektit; pysyvät tunnisteet; tiedeinstituutit; tiedontuottajat; tieteellinen
julkaisutoiminta; tutkimusprojektit; tutkimustyö

CC BY 4.0

© 2020 Tiedeinstituuttien avoin tutkimusdata -hanke

Tiivistelmä

Tämä asiakirja on tehty osana opetus- ja kulttuuriministeriön rahoittamaa Tiedeinstituuttien avoin tutkimusdata -hanketta. Hankkeen toteuttajina toimivat Suomen ulkomaisten tiedeinstituuttien taustayhteisöt. Hanketta hallinnoi Institututum Romanum Finlandiae sr.

Asiakirja on hankkeen tutkijoiden ehdotus ja suositus tiedeinstituuteille, miten näiden olisi mahdollista järjestää ja kehittää omaa toimintaansa digitaalisten tutkimusaineistojen käsittelyssä, säilytyksessä ja julkaisemisessa. Asiakirjassa esitetyt näkemykset perustuvat hankkeen tutkijoiden selvityksiin. Sitä on käsitelty hankkeen ohjausryhmässä, joka on antanut sisältöön liittyviä ohjeita, kommentteja ja palautetta.

Asiakirjaan on kerätty palautetta erityisesti hankkeessa mukana olevilta instituuteilta (henkilöstö, säätiöiden hallitukset) kesällä ja syksyllä 2020. Palautteiden perusteella päivitetty versio laaditaan loppuvuonna 2020 ja julkaistaan versiona 1.1.

Asiakirja pyrkii ennen kaikkea antamaan tiedeinstituuttien henkilöstölle ja instituuttien toiminnasta vastaaville taustayhteisöille työvälineitä sekä pidemmän tähtäimen strategiseen suunnitteluun että instituuttien päivittäisen toiminnan järjestämiseen.

Sisällys

1. Yleistä	1
Perusteluita	1
Hyöty instituutille	2
Toimintamallin tavoitteet	2
Vähimmäisvaatimukset tutkijalle tarjottavista palveluista ja ratkaisuksista instituutissa	2
2. Tutkijan ja instituutin välisen aineistohallintasuhteen vaiheet	3
Rekrytointi	3
Työpaikkailmoitus	3
Työhaastattelu	3
Alku - Tutkija tulee instituuttiin / Tutkimushanke alkaa (instituutin näkökulmasta)	3
Työ- tai apurahasopimus	3
Aineistohallintasuunnitelma	3
Tunnisteet	4
Ennen työsuhteen alkamista	4
Keskikohta - Tutkija tekee tutkimusta	4
Tutkijan työsuhde tai tutkimushanke lähenee loppuaan	4
Tutkijan lähdettyä instituutista	5
3. Tekniset ratkaisut aineistohallinnat tarpeisiin	6
Tutkimuksen aikainen aineistohallinta	6
EUDAT-palvelut	7
Pilvipalvelu tiedostojen jakamiseen ("Dropbox")	7
Verkkolevy (+ oma pilvipalvelin)	8
Tutkimusaineistojen pitkäaikaissäilytys (PAS)	9
Tallentaminen erillisille medioille	10
Tallentaminen levypalvelimelle	10
Tutkimusaineistojen avaaminen digitaalisesti	11
Aineiston avaaminen kokonaisuuksien tasolla	11
Aineiston avaaminen sisältötasolla	12
Liite A: Tutkimusaineiston alkuinventaarior	13
Liite B: Tutkimusaineiston loppuinventaarior	14

1. Yleistä

Toimintamallin tavoitteena on tarjota instituuttien henkilöstölle toteutettavissa olevat ratkaisut tutkimusaineistojen hallintaan ja avaamiseen. Toimintamalli on luotu pitäen silmällä kansainvälistä ja kansallista keskustelua. Sen tarjoamat ratkaisut on laadittu siten, että niiden avulla olisi mahdollista pyrkiä kohti avoimen datan FAIR-periaatteiden mukaisia ratkaisuja. Tämän lisäksi toimintamalli pyrkii tarjoamaan ratkaisuja myös tutkimuksen aikaisen aineistohallinnan tarpeisiin tavalla, joka ottaa huomioon sekä instituuttien pienen koon, rajalliset resurssit että mahdolliset olemassa olevat ratkaisut. Tutkimuksen aikaisen aineistohallinnan ratkaisut ovat lisäksi sovellettavissa instituuttien muun hallinnon tarpeisiin.

Perusteluita

Keskeinen teema on sitoutuminen avoimen tieteen periaatteisiin ja käytäntöihin. Sellaisenaan tämä on yksinkertainen toteamus, mutta kun sitä puretaan käytännön elämään, esiin nousee useita ongelmallisia ja vaikeita seikkoja, joita voidaan pohtia tutkimukseen liittyvien tutkimuksen (osa)tekijöiden kautta.

- Tutkija: osaaminen, motivaatio, ajan ja rahan puute, ymmärrys, "ylhäältä sanelu", työsuhteet, aineiston omistajuus ja tekijänoikeudet, meritoituminen.
- Tutkimusyhteisö: osin kuten yllä – vaaditaan siis koko yhteisön tuki ja osallistuminen. On hyväksyttävä eräänlainen paradigman muutos, jossa tutkimuksen näkyvyys ei painotu julkaisuun vaan prosessiin.
- Rahoittaja: rahoitusmekanismi ohjaavat käytäntöjä (jos edellytetään datan hallintaa, sitä yleensä yritetään tehdä, jotta saataisiin rahoitusta), rahoittajan pitää investoida lisää yksittäisiin tutkimuksiin, jotta datan hallinta mahdollistetaan käytännössä (raha ei saa valua tutkimuksesta hallintoon)
- Tutkimusorganisaatio: johdon on sitouduttava avoimen tieteen periaatteiden ajamiseen ja myös budjetoitava tähän. Tämä edellyttää luottamusta tutkijoiden ja organisaation johdon välillä.

Hyöty instituutille

- Laatuarvioinnin tekijä: instituutissa tehtävä tutkimus dokumentoidaan – instituutti näkyväksi.
- Rahoittajan vaatimus: OKM:n rahoitukseen liittyvät velvoitteet avoimen tieteen ja tutkimuksen osalta
- Tutkijoiden auttaminen: tutkimuksen prosessi, vastuut ja velvollisuudet kaikille osapuolille selviä.

Toimintamallin tavoitteet

- Yhtenäinen linja tiedeinstituuteissa – pohdittava keskitettyjä ratkaisuja käytännössä siten, että rahoitus on turvattu.
- Hyötynä muille tutkimuslaitoksille: miten tutkimushallinto avoimen tutkimuksen periaatteet huomioiden voidaan toteuttaa pienessä organisaatiossa, jossa tutkimusinfrastruktuuri on olematon ja joka ei voi hyödyntää kansallisia palveluita
- Koko tutkimushallinnon aineistohallinta ja käsittelyprosessit järjestykseen.

Vähimmäisvaatimukset tutkijalle tarjottavista palveluista ja ratkaisuksista instituutissa

- Peruskäsikirja
- Malliasiakirjoja, esimerkiksi aineistohallintasuunnitelma, readme-pohja ja kuvailumalleja
- Tallennus/säilytysresurssi

Instituutin tutkijalla tarkoitetaan instituutin palkkalistoilla tai apurahalla olevaa henkilöä tai vastuuhenkilöä tutkimusprojektissa, jonka rahoitukseen ja/tai toteuttamiseen instituutti osallistuu.

2. Tutkijan ja instituutin välisen aineistohallintasuhteen vaiheet

Rekrytointi

Työpaikkailmoitus

Voidaan mainita työpaikkailmoituksessa instituutin lyhyessä kuvauksessa (esim. "NN instituutti on sitoutunut avoimen tieteen periaatteisiin").

Työhaastattelu

Tutkimusaineistojen omistajuuteen liittyvää keskustelua voi käydä jo työhaastattelun yhteydessä, jos tämä tuntuu tarpeelliselta esim. liittyen työntekijän tutkimushankkeeseen.

Alku – Tutkija tulee instituuttiin / Tutkimushanke alkaa (instituutin näkökulmasta)

Työ- tai apurahasopimus

Sovitaan työsuhteen aikana syntyvien aineistojen omistajuudesta ja käyttöehdoista. Tavoitteena on synnyttää asiakirja, jonka perusteella työntekijällä ja työnantajalla eli instituutilla on sama näkemys siitä, kuka työsuhteen aikana tuotetun tutkimusaineiston omistaa ja kenellä on siihen käyttöoikeus milläkin ehdoilla.

Aineistohallintasuunnitelma

Dokumentoidaan tutkimusaineistot, jotka tutkija tuo mukanaan.

Tutkija laatii tai päivittää aineistohallintasuunnitelman ja toimittaa sen instituuttiin. Suunnitelman päivityksessä se laaditaan vastaamaan instituutissa työskentelyn aikaa. Instituutissa tieto tallennetaan. On sovittava vastuhenkilö ja tallennus.

Pyydettävissä tiedoissa täytyy olla niin selkeät kysymykset, että instituutin henkilöstö ja tutkija pystyvät toimimaan ilman tutkimushallinnon tukea.

Tunnisteet

ORCID-tunniste on hyvä olla kaikilla instituuteissa tutkimusta tekevillä henkilöillä. (ISNI-tunnisteet muilla kuin tutkijoilla.) Nämä muistetaan kirjata!

Ennen työsuhteen alkamista

Instituutin aineistohallintavastaava tekee tutkijan kanssa tiedonhallintasuunnitelmassa sovitut ratkaisut: perustaa tarvittavat levyalueet, hankkii pilvipalvelut ja luo käyttäjätunnukset.

Apuna voi käyttää liitteenä olevaa datainventaarion malliasiakirjaa.

Keskikohta – Tutkija tekee tutkimusta

Aineistohallintasuunnitelman on tarkoitus olla käytännön toimia kuvaava asiakirja, ei hallinnon tarpeisiin tehty, työsuhteen alussa laadittu dokumentti. Niinpä aineistohallintasuunnitelmaa on päivitettävä työn kuluessa ja sen on toimittava ohjekirjana käytännön tasolla myös tutkijalle itselleen. Mikäli aineistohallintasuunnitelma ei vastaa konkreettiseen kysymykseen, vaatii se päivitystä – ja tämä on tutkijan itsensä vastuulla.

Päivitetty aineistohallintasuunnitelma käsitellään tutkimusraportoinnin yhteydessä.

Tutkija päivittää aineistohallintasuunnitelmaa. Vähintään kerran vuodessa tähän palataan esimerkiksi tutkimusraportoinnin yhteydessä instituutin tai säätiön toimintakertomukseen. Tarkoitus on huomioida aineistohallintaan ja käsittelyyn tehdyt muutokset ja niiden syyt, ongelmakohtat, tutkimuksen vaiheet jne.

Tutkijan työsuhde tai tutkimushanke lähenee loppuaan

Kun tutkijan työsuhde instituuttiin tai tämän vetämä tutkimushanke lähenee loppuaan, on syytä käydä yhdessä tutkijan kanssa läpi sekä se mitä työsuhteen alussa on

syntyvien aineistojen loppukäytöstä sovittu että se minkälaisia aineistoja tutkimuksen aikana on syntynyt.

Aineistojen kohdalta on syytä todeta niihin työsopimuksen ja lakien perusteella syntyneet tekijän- ja omistusoikeudet, ja mahdollisesti sopia lisäyksistä näihin esim. käyttöoikeuksien muodossa. Aineistoihin voidaan esim. myöntää määräaikaista yksinoikeudellisia käyttöjä, esim. siten, että vaikka aineiston omistajuus on instituutilla, työntekijä saa viiden vuoden ajan työsuhteen päättymisen jälkeen käyttää aineistoa yksinoikeudella. Aineiston omistajuus voi silti olla instituutilla ja tekijänoikeus tutkijalla.

Jos tutkimus jatkuu, vaikka tutkija ei ole enää instituutin työsuhteinen tai apurahalla toimiva tutkija, laaditaan loppuraportti, johon kirjataan aineiston tilanne (mitä aineistoa, mihin tallennettu, kuinka suuria määriä, jääkö instituuttiin jotain, jne.). Jos tutkimus päättyy, päätetään aineiston loppusijoitus (mitä, mihin, jne.).

Aineistosta on syytä tehdä loppuinventaario, josta on malliasiakirja toimintamallin liitteenä.

Tutkijan lähdettyä instituutista

Instituutin aineistohallintovastaava käy lävitse tutkijan jälkeensä jättämän aineiston ja huolehtii sen säilytyksestä instituutin aineistohallintapolitiikan mukaisesti.

3. Tekniset ratkaisut aineistohallinnat tarpeisiin

Tutkimuksen aikainen aineistohallinta

Tutkimuksen aikaisessa aineistohallinnassa keskeistä on aineistojen saatavuus tutkimusryhmän ja hankkeen sisällä. Tutkimushankkeen aineistohallintasuunnitelmassa (*Data Management Plan*) kerrotaan tarkemmin hankkeessa käytetyistä aineistoista ja niiden hallinnasta ja järjestämisestä. Aineistohallintasuunnitelman perusteella ja pohjalta on mahdollista tehdä ratkaisuja valittuihin teknisiin järjestelmiin perustuen kiinnittämällä huomiota mm. seuraaviin kysymyksiin:

1. Onko hankkeessa useampia tutkijoita, joiden pitää pystyä käyttämään samaa aineistoa samanaikaisesti?

- a. Onko aineisto staattista tiedostoaineistoa (vrt. valokuvat) vai dynaamista, hankkeen aikana muuttuvaa ja kehittyvää aineistoa (vrt. tietokanta) – vaikuttaa siihen, onko tiedostonhallintaratkaisu riittävä vai tarvitaanko esim. tietokantasovellus.
- b. Ovatko kaikki hankkeen tutkijat instituutin palkkalistoilla (riittääkö sisäinen järjestelmä) vai tarvitaanko aineistoon pääsy myös instituutin ulkopuolisille osallistujille?

2. Kerätäänkö hankkeessa uutta alkuperäisaineistoa, joka on tarkoitus tallentaa pysyväissäilytykseen hankkeen jälkeen?

3. Onko hankkeessa kerättyä aineistoa tarkoitus julkaista sellaisenaan jo hankkeen aikana?

Seuraavassa esitetään erilaisia ratkaisuja tutkimuksenaikaisen aineistohallinnan tarpeisiin. On huomattava, että nämä eivät välttämättä ole toisiaan poissulkevia valintoja, vaan instituutissa voidaan hyvin ottaa käyttöön useampiakin näistä hankkeiden ja toimintojen erilaisten vaatimusten mukaisesti.

EUDAT-palvelut

EUDAT tarjoaa työvälineitä tutkimuksen aikaiseen tiedostojen ja aineistojen hallintaan. Aineistoja on mahdollista jakaa tutkimusryhmän jäsenten kesken, minkä lisäksi niitä on mahdollista julkaista ja siirtää pitkäaikaistallennukseen.

EUDAT-infrastruktuurin tuottavat yhdessä joukko eurooppalaisia suuria datakeskuksia ja tutkimusorganisaatioita.

Pilvipalvelu tiedostojen jakamiseen ("Dropbox")

Käytännöllisin ratkaisu monen tutkijan hankkeissa on jonkinlainen pilvipohjainen tiedostonhallinta, jotka tunnetaan usein nimellä "Dropbox" ensimmäisen laajalle levinneen sovelluksen mukaan. Dropbox ei kuitenkaan ole soveltuva tieteellisten hankkeiden käyttöön, vaikka ilmaisena helposti saatavissa onkin. Dropboxin keskeisinä ongelmina voidaan pitää mm. sitä, että Dropbox-sovelluksen kommunikatio pilven kanssa on salaamatonta ja tiedostot säilytetään pilvessä salaamattomina. Minkäänlaisen sensitiivisen informaation käyttöön Dropboxia ei missään tapauksessa voi käyttää. (Tämä koskee erityisesti ilmaisversiota, maksullinen lisenssi saattaa sisältää parannuksia tietoturvaan.) Toinen ongelma koskee palvelinten sijaintia. Esim. eurooppalainen lainsäädäntö rajoittaa sitä, mihin tietoa saa siirtää, ja ainakin Rooman ja Ateenan kannalta on relevanttia pitäytyä sellaisissa palveluissa, jotka takaavat tiedon pysymisen EU-alueen sisäpuolella.

Muita vastaavia palveluita kuitenkin löytyy. Hankkeen koekäytössä on ollut saksalainen TeamDrive, minkä lisäksi esim. joissain suomalaisissa yliopistoissa käytetään Seafile-järjestelmää. Näiden lisäksi ainakin ownCloud tarjoaa vastaavan toiminnallisuuden. Näissä kaikissa tietoturva on korkealla tasolla, ja ainakin TeamDrive, ownCloud ja Seafile tarjoavat mahdollisuuden tyypillisimpien toimisto-ohjelma-asiakirjojen yhteiskäyttöön ja -muotoiluun (vrt. Google Docs).

Näitä valittaessa on kiinnitettävä huomiota seuraaviin seikkoihin:

- Halutaanko ostaa pilvipalvelu ja palvelintila valmiina vai asentaa pilvipalvelu omille palvelimille?

- Valmiina ostettu palvelu sisältää ulkopuolisen palveluntarjoajan ylläpitämän pilven, johon liittyy erilaisia tietoturva- ja käyttöoikeusominaisuuksia sekä aineistojen varmuuskopiointi. Palvelimien fyysinen sijainti saattaa asettaa rajoituksia esim. aineistolle, jota ei saa viedä pois maasta tai EU:sta. Valmiina ostettu pilvi on yleensä helppo ottaa käyttöön, mutta integrointi olemassa oleviin järjestelmiin voi vaatia työtä.
 - Asennus omille palvelimille. (Ks. seuraava pääkohta)
- Miten käytetty järjestelmä käsittelee ristiriitatilanteet, esim. sen, että kaksi käyttäjää on muokannut samaa tiedostoa samanaikaisesti, tai yksi käyttäjä on muuttanut tiedoston nimen ja toinen samaan aikaan poistanut sen? Esim. samanaikaisten muutosten tilanteessa Dropbox kirjoittaa huoletta aina edellisen muutoksen ylitse asiasta mitään huomauttamatta, joten jommankumman tekemä työ menee täysin hukkaan. TeamDrive sisältää tiedostojen lukitsemisominaisuuden, jolloin kahden käyttäjän ei ole periaatteessa mahdollista muokata yhtä tiedostoa samanaikaisesti. Jos näin käy, syntyy ristiriita, joka näkyy TeamDrive-asiakasohjelmassa.

Verkkolevy (+ oma pilvipalvelin)

Mikäli koko tutkimushanke toteutetaan saman fyysisen rakennuksen sisällä ja valtaosa työstä tehdään samassa tilassa, instituutin tiloihin rakennettu fyysinen verkko verkkolevyineen täyttää hankkeen tallennustilan vaatimukset. Tällöin verkkolevyjärjestelmän varmuuskopiointinista on huolehdittava säännöllisesti siten, että aineisto kopioidaan tai se kopioituu myös toiseen fyysiseen sijaintiin (palovahingot jne.)

Mikäli verkkolevyjärjestelmän yhteyteen on mahdollista rakentaa oma pilvipalvelin (esim. TeamDrive ja ownCloud tarjoavat mahdollisuuden rakentaa pilvipohjainen tiedostonhallinta instituutin omien palvelimien ympärille) on omia verkkolevyjä mahdollista käyttää aineistonhallintaan myös sellaisissa hankkeissa, joissa osallistujat työskentelevät säännöllisesti instituutin ulkopuolella. Yleensä tällaiset järjestelmät sisältävät muitakin ryhmätyöskentelyä tukevia ominaisuuksia.

Koska joka tapauksessa on suositeltavaa, että instituuteissa siirryttäisiin käyttämään erilaisia pilvipohjaisia järjestelmiä tiedonhallintaan (ks. suositukset instituuteille), muodostaa tämä relevantin ratkaisun instituuteille, joilla on jo käytössä omia it-ratkaisuja.

Tutkimusaineistojen pitkäaikaissäilytys (PAS)

Digitaalisten tutkimusaineistojen pitkäaikaissäilytyksen tekniset vaatimukset eivät todennäköisesti ole yhdenkään yksittäisen instituutin ratkaistavissa. Siinä missä materiaalien tutkimusaineistojen säilytysratkaisut ovat helposti ymmärrettävissä ja kohtuullisella vaivalla toteutettavissa tavalla, joka varmistaa aineistojen säilymisen olemassa olevien ja tunnettujen arkistovaatimusten mukaisesti esim. lämpötilaa, kosteutta ja valon määrää ja laatua säätelemällä, digitaalisten tutkimusaineistojen pitkäaikaissäilytykseen ei vastaavia, selkeästi ja varmasti toteutettavia ratkaisuja ole vielä olemassa.

Analogisessa muodossa olevien aineistojen hävittämistä digitalisoitujen aineistojen tieltä ei arvokkaan tutkimusaineiston yhteydessä voi vielä suositella, vaikka hallinnollisten asiakirjojen tapauksessa järjestelmällinen tiedonhallinta tämän voisikin mahdollistaa. Pysyvästi säilytettävien hallinnollistenkin asiakirjojen säilyttäminen paperisina on kuitenkin lähtökohta.

Ehdottomasti suositeltavinta on järjestää pitkäaikaissäilytys hyödyntäen julkisin resurssein varmistettuja infrastruktuureja. Tällaisia ovat mm. Suomessa Tietoarkisto, CSC:n tuottamat fairdata-palvelut sekä esim. CERN:in ylläpitämä Zenodo.org. Lisäksi EUDAT-palveluiden yhteydessä toimiva B2SAFE voi soveltua instituuttien aineistojen ulkopuoliseksi varmuuskopioratkaisuksi.

Tästä huolimatta on tieteellisesti perusteltua säilyttää aineistoa myös instituuteissa. Tässä esitetyt säilytykseen liittyvät ohjeet koskevat juuri instituutin omaa aineistonhallintaa, mutta ne eivät vastaa aineiston pitkäaikaissäilytyksen vaatimuksiin.

Tallentaminen erillisille medioille

Suhteellisen lupaavilta näyttävät uudet M-DISC-tekniikkaan pohjautuvat DVD- ja Blu-Ray -levyt, joiden säilyvyys on nähtävästi huomattavasti parempi kuin perinteisten DVD- ja Blu-Ray-levyjen. Suurin ero näiden välillä on siinä, että M-DISC-tekniikassa käytetty väriaine on epäorgaanista, ja sikäli huomattavasti vähemmän altista olosuhteiden aiheuttamille vaurioille. Onkin suositeltavaa, että paikallinen, tallennettava aineisto tallennetaan M-DISC-standardin mukaisille levyille siten että fyysiset tallennusyksiköt sijoitetaan seuraavasti.

- Arkistoversio, johon ei kosketa. Tallennetaan pimeään, lämpötilaltaan tasaiseen ympäristöön, jonka ilmankosteus on vakio
- Käyttöversio, joka tallennetaan samoin, mutta jota voidaan tarvittaessa käyttää tiedon palauttamiseen
- Toiseen fyysiseen sijaintiin sijoitettu varmuuskopio. Instituuttien tapauksessa toinen sijainti voisi hyvin olla Suomessa, esim. säätötoimistotilojen yhteydessä.

Näiden formaattien lisäksi on syytä laatia aineistolle tarkastus- ja päivityssuunnitelma, jonka perusteella aineiston luettavuus varmistetaan ja jonka mukaan aineisto siirretään uusille medioille **ennen sen muuttumista lukukelvottomaksi**. M-DISC-tekniikan yhteydessä voidaan ehkä ajatella aineiston kontrolloinnille 10 vuoden väliä. Siirtämistä uudelle medialle on vaikea arvioida; M-DISC-tekniikan kehittäjä väittää levyjen kestävän 1000 (tuhat) vuotta, mutta konkreettisia esimerkkejä tästä ei tietystikään ole olemassa.

Tallentaminen levypalvelimelle

Toinen vaihtoehto on huolehtia aineiston arkistoinnista pitämällä se tallessa instituutin levypalvelimella. Tällöin levypalvelimen on oltava toteutettu niin, ettei satunnainen laitteiston vaurioituminen, tulipalo, vesivahinko tai maanjäristys tuhoa arkistoitua tietoa kokonaisuudessaan. Käytännössä kyseeseen tulee palvelinratkaisu, josta on säännöllinen peilitys ulkopuoliseen palvelimeen, esim. erilaisia pilvitekniikoita hyödyntäen. Tällainen ratkaisu vastaa aineiston pitkäaikaistallentamisen vaatimuksiin, mutta soveltuu käytännössä vain tutkimushankkeiden aineistojen

sisäisen arkistoinnin toteuttamiseen – aineistojen avaaminen on käytännössä tehtävä toisin.

Tutkimusaineistojen avaaminen digitaalisesti

Tutkimusaineistojen digitaaliseen avaamiseen on käytännössä kaksi eri vaihtoehtoa.

- Aineisto avataan kokonaisuuksien tasolla.
- Aineisto avataan yksittäisten tiedoston tai jopa tietokannan tietueiden tasolla.

Ensimmäinen vaihtoehto käytännössä tarkoittaa, että hankkeen aineisto tai siitä julkaistavaksi valittu osa paketoidaan yhdeksi kokonaisuudeksi, joka asetetaan sopivaan paikkaan julkisesti saataville.

Toisessa tapauksessa tiedostojen tai tietokannan ympärille rakennetaan jonkinlainen käyttöliittymä, joka mahdollistaa aineiston selailun, järjestämisen ja lataamisen pienemmissä osissa, yleensä ilman erityistä teknistä osaamista.

Näiden kahden ratkaisun tekniset vaatimukset instituutille ovat hyvin erilaiset.

Aineiston avaaminen kokonaisuuksien tasolla

Aineiston avaaminen kokonaisuuksien tasolla kannattaa yleensä toteuttaa olemassa olevia julkisia resursseja hyödyntäen. Paras vaihtoehto olisi käyttää kansallisia resursseja, joiden tuottamisesta vastaa CSC (ks. avoindata.fi), mutta koska näiden hyödyntäminen ei tällä hetkellä ole instituuteille mahdollista, on aineistotason avaaminen syytä toteuttaa muuten. Yksinkertaisin ratkaisu on julkaista aineistotason kokonaisuudet hyödyntäen CERN:in ylläpitämää Zenodo.org -palvelua.

Jokaisen instituutin kannattaa tätä varten tehdä Zenodo.org -palveluun oma ”yhteisö” (*community*) jonka ylläpitäjänä (*curator*) toimii instituutin aineistohallintovastaava. Tällöin instituutin piirissä tuotetut aineiston on mahdollista kerätä yhteen paikkaan ja niihin on mahdollista viitata esim. instituutin www-sivuilta. Yhteisön ylläpitäjä vastaa siitä, mitkä aineistot voidaan liittää osaksi instituutin yhteisöä, mutta

itse aineistojen siirto palveluun on lähtökohtaisesti niiden omistajien tai haltijoiden vastuulla.

Aineiston avaaminen sisältötasolla

Aineiston avaamisessa sisältötasolla on kysymys siitä, että rakennetaan kunkin aineiston tarpeen soveltuva käyttöympäristö, jossa aineiston sisältöä on mahdollista selata, järjestää, hakea ja ladata käytettäväksi edelleen. Tällöin kyseeseen tulee käytännössä jonkinlaisen aineistonhallinto- tai arkistojärjestelmän käyttöönotto, joka ei ole teknisesti triviaali toimenpide.

Mikäli aineistoja halutaan avata tällä tasolla, on varauduttava siihen, että valittu ratkaisu tulee synnyttämään kiinteitä kustannuksia niin kauan kuin aineisto halutaan pitää saatavilla. Toisaalta kulut voivat olla yhdistettävissä instituuttien it-palveluista joka tapauksessa syntyviin kuluihin, mutta tämä on tapauskohtaista.¹

Toinen vaihtoehto on toteuttaa avaaminen yhteistyössä jonkin toimijan kanssa, jolla on jo valmiiksi käytettävissä soveltuva alusta. Tällöin mahdollisten kustannusten synty riippuu täysin valitusta kumppanista ja tämän kanssa tehdystä sopimuksesta.²

¹Datahankkeessa tätä on testattu sekä Rooman Bolli Doliari että Ostia-pilottien että Ateenan Gouriza-pilotin yhteydessä.

²Suomen Lähi-Idän instituutin Hilma Granqvist -aineiston julkaisu on toteutettu tällä tavalla yhteistyössä SLS:n kanssa.

Liite A: Tutkimusaineiston alkuinventaarior

Kysymyksiä, kun johtaja/tutkijaopettaja/tutkija aloittaa x-instituutin säätiön työntekijänä tai työskentelyn säätiön rahoittaman tai hallinnoiman tutkimuksen parissa.

x-instituutin säätiö on itsenäinen, tutkimusta ja tiedettä tukeva organisaatio, joka ylläpitää x-instituuttia. Tutkimusaineistot ovat arvokasta omaisuutta tutkimustarkeituuksiin, ja niiden asianmukaisen hallinnan turvaaminen kuuluu säätiön velvollisuuksiin. x-instituutin säätiö on vastuussa sen hallussa olevista henkilötiedoista, jotka voivat liittyä joko tutkimusaineistoon tai tutkimushankkeen hallinnointiin.

Jotta säätiö voi täyttää velvollisuutensa tutkimuslaitoksena tutkimusaineiston ja henkilötiedon osalta, säätiöllä on oltava tieto sen toimintaan kuuluvasta tutkimusaineistosta. Kysymykset auttavat myös suunnittelemaan ja varautumaan tarvittaviin teknisiin ratkaisuihin (esim. säilytys).

Nimi

Aiempi tai nykyinen organisaatio (jos eri kuin x-instituutin säätiö)

Yhteystiedot

Mitä aineistoa tuot x-instituuttiin (tai x hankkeeseen)?

Pysyvä tunniste (PID eli persistent identifier, esim. DOI):

Aineiston nimi (jos on):

Aineiston kuvaus (miksi aineisto on kerätty):

Aineiston sijainti:

Aineiston luoja (nimi ja ORCID):

Aineiston omistaja/hallinnoija (nimi ja ORCID):

Aineiston tyyppi (teksti, numeromuotoinen, tietokanta, tms.):

Aineiston luontiajankohta:

Onko aineistossa henkilötietoja tai arkaluonteista tietoa (ks. käsikirja):

Tutkimusala:

Asiasanat:

Pääsy aineistoon (avoin/rajattu/ei pääsyä):

Liittyykö aineistoon lisenssi tai sopimus?

Kyllä -> Mitä rajoituksia käyttöön sisältyy?

Kyllä -> Datan lisenssi:

Ei

En tiedä

Käytän aiemmin kerättyä ja arkistoitua dataa, jonka pysyvä tunniste on xxxx.

Liite B: Tutkimusaineiston loppuinventory

Kysymyksiä, kun johtaja/tutkijaopettaja/tutkija lopettaa x-instituutin säätiön työntekijänä tai työskentelyn säätiön rahoittaman tai hallinnoiman tutkimuksen parissa.

x-instituutin säätiö on itsenäinen, tutkimusta ja tiedettä tukeva organisaatio, joka ylläpitää x-instituuttia. Tutkimusaineistot ovat arvokasta omaisuutta tutkimustarkeituksiin, ja niiden asianmukaisen hallinnan turvaaminen kuuluu säätiön velvollisuuksiin. x-instituutin säätiö on vastuussa sen hallussa olevista henkilötiedoista, jotka voivat liittyä joko tutkimusaineistoon tai tutkimushankkeen hallintaan.

Jotta säätiö voi täyttää velvollisuutensa tutkimuslaitoksena tutkimusaineiston ja henkilötiedon osalta, säätiöllä on oltava tieto sen toimintaan kuuluvasta tutkimusaineistosta tutkijan lähtiessä x instituutista (mitä aineistoa työsuhteen aikana on luotu ja julkaistu). Kysymykset auttavat myös suunnittelemaan ja varautumaan tarvittaviin teknisiin ratkaisuihin (esim. säilytys).

Nimi

Nykyinen ja tuleva organisaatio

Yhteystiedot

Jos olet ollut mukana luomassa tutkimusaineistoa, vastaathan seuraaviin kysymyksiin, jotta varmistat omalta osaltasi tutkimusryhmän muiden jäsenten pääsyn aineistoon.

Missä aineistoa säilytetään?

Jos olet ainoa, jolla on pääsy aineistoon, millä tavoin jaat aineiston muulle tutkimusryhmälle?

Mitä aineistoa olet toimittanut datarepositorioihin?

Pysyvä tunniste (PID eli persistent identifier, esim. DOI):

Aineiston nimi (jos on):

Aineiston kuvaus (miksi aineisto on kerätty):

Aineiston sijainti:

Aineiston luoja (nimi ja ORCID):

Aineiston omistaja/hallinnoija (nimi ja ORCID):

Aineiston tyyppi (teksti, numeromuotoinen, tietokanta, tms.):

Aineiston luontiajankohta:

Onko aineistossa henkilötietoja tai arkaluonteista tietoa:

Tutkimusala:

Asiasanat:

Pääsy aineistoon (avoin/rajattu/ei pääsyä):