



Helsingin yliopiston kirjaston tutkimusdataprojekti

LOPPURAPORTTI VERSIO 1.2

TUTKIMUSDATATYÖRYHMÄ 8.3.2012

Helsingin yliopiston kirjasto

TEKIJÄ (T)

Eva Isaksson, Joonas Kesäniemi, Kimmo Koskinen, Mari Elisa Kuusniemi, Annikki Roos, Marjut Salokannel, Liisa Siipilehto.

TIIVISTELMÄ

Tämän pilottiprojektin tavoitteena oli pureutua tutkimusdatan hallinnan, käytön ja säilyttämisen problematiikkaan yhden Helsingin yliopistossa toimivan, kansainvälisen tutkimusryhmän kautta. Tarkoituksena oli, että projektin avulla selvitetään tutkijoiden tutkimusdataan tutkimusprosessin eri vaiheissa liittyviä käytäntöjä, kysymyksiä ja ongelmia. Työryhmän tuli pohtia kirjaston roolia jatkossa näissä kysymyksissä ja niiden ratkaisemisessa. Yhdeksi merkittäväksi tavoitteeksi pilottiprojektille asetettiinkin, että se antaa mahdollisuuden kartuttaa kirjastohenkilöstön aihepiiriin liittyvää osaamista. Lisäksi tavoitteena oli yleisemmällä tasolla pohtia, millaisin toimin käyttäjäystävällisen, integroidun informaatioympäristön rakentamista voitaisiin edistää.

Pilotin kautta paljastui varsin karu todellisuus siitä, miten tutkijat kamppailevat työssään puutteellisen ohjeistuksen ja suunnittelun sekä olemattomien datanhallintaresurssien kanssa, käytännöllisesti katsoen ilman mitään käytännön tukea. Tutkijat toivoivat tietoa siitä, mihin paikkaan, millaista aineistoa ja kuinka paljon dataa voi tallentaa. Datarakenteen ja tiedostokuvausten osalta oli tarvetta apuun ja tukeen jo tutkimusprojektin suunnitteluvaiheessa. Tukea tarvittiin myös lupa-asioiden hoitamiseen, kuten datan omistus- ja käyttöoikeuksiin mm. projektin käynnistys- ja lopetusvaiheiden kirjauksiin sopimuksista.

Tässä loppuraportissa työryhmä esittää useita ehdotuksia, joiden avulla tilannetta voitaisiin parantaa. Pilotin kuluessa kirjaston työryhmä on toiminut hyvässä yhteistyössä monien yliopiston sisäisten ja ulkoisten tahojen kanssa. Käsityksemme siitä, että tutkimusdataan liittyviä ongelmia ja käytännön toimintatapoja voi kehittää ainoastaan yhteistyössä monien eri tahojen kanssa, on entisestään vahvistunut. Tästä yhteistyöstä ja tehtäväjaosta eri toimijoiden kesken työryhmä tekee myös ehdotuksen.

Helsingin yliopiston kirjaston rooliksi työryhmä ehdottaa datan hallintapalvelujen integrointia, koordinoitua ja tukea, metadataan ja järjestelmien käyttöliittymiin (esim. haettavuus, linkitykset) liittyviä tehtäviä, tiettyä osaa koulutuksesta, viestintää, neuvontaa ja opetusaineiston valmistamista. Kirjasto toimii kaikilla kampuksilla ja on näin ollen lähellä kaikkien tieteenalojen tutkijoita. Siellä on kootusti eri tieteenalojen käytäntöjä tunteva ja tiedon järjestämisen, etsimisen ja tähän erikoistuneisiin tietojärjestelmiin perehtyneitä ammattilaisia. Olisi *toivottavaa*, että Helsingin yliopiston kirjasto entistä rohkeammin hakeutuisi yhteistyöhön tutkijoiden ja muiden yliopiston osajien kanssa. Tulosta syntyy arjessa, "solmumaisesti" työskentelemällä.

On tutkijoiden, tutkimustyön ja yliopiston etu, että työryhmän ehdottamiin toimiin tartutaan.

ASIASANAT

tutkimusaineistot, tutkimus, tiedonhallinta, metadata, tietoarkistot, roolit, toimijat, tieteelliset kirjastot

KOKONAISIVUMÄÄRÄ

38s

KIELI

suomi

Sisällysluettelo

1	Datatyöryhmän ehdotukset	3
1.1	Tutkimusdatatyöryhmä ehdottaa, että	3
2	Johdanto	5
2.1	Toimeksianto	5
3	Pilottitutkimus	6
3.1	Tutkimusryhmän valitseminen pilotiksi	6
3.2	Tutkimusryhmän esittely	6
3.3	Työskentely tutkimusryhmän kanssa	7
4	Yhteistyö muiden toimijoiden kanssa	8
4.1	Yhteiskuntatieteellinen tietoarkisto	8
4.2	Jyväskylän yliopisto	8
4.3	Tietotekniikkakeskus	8
4.4	Tieteen tietotekniikan keskus CSC	9
4.5	Kansalliskirjasto	9
4.6	Helsingin yliopiston Tutkimushallinnon yksikkö, tutkimuksen toimiala	9
5	Projektin tulokset	10
5.1	Havaitut ongelmat ja esteet datan hallinnassa	11
5.1.1	Aineistonhallintasuunnitelma	12
5.2	Oikeudellisia kysymyksiä	13
5.2.1	Primäärin tutkimusdatan tekijänoikeudellinen status	13
5.2.2	Tutkimusaineiston omistajuus	13
5.2.3	Tekijänoikeudellisen suojan monitasoisuus: tutkimusdataa sisältävien tietokantojen suoja	13
5.2.4	Työsuhteessa syntyneet tietokannat	14
5.2.5	Yksityisyyden suoja	14
5.2.6	Patentit	14
5.3	Tietojärjestelmät ja palvelut datanhallintaan	15
5.3.1	Tutkimusdatan varastointityyppejä	15
5.3.2	Tietojärjestelmäkartoitus	16
5.3.3	Tietojärjestelmien integraatioskenaarioita	16
5.3.4	Tietotekniikkakeskuksen hanke Tutkimusdatan säilyttämisen tekninen infrastruktuuri	18
5.4	Tutkimusdatan metadata	18
5.4.1	Standardit ja ohjeet	18
5.4.2	Tutkimusdatan pysyvät tunnisteet	19
5.5	Julkaisujen linkitys tutkimusdataan	19
5.6	Pikakysely ilmatieteen tutkijoille	20
5.7	Osaamisen kehittäminen	20
5.7.1	Kirjastot tarvitsevat datanhallinnan osaajia	20
5.7.2	Datanhallinnan taidot opiskelijalle ja tutkijalle	20

5.7.3	Opetusmateriaalin tuottaminen	21
5.8	Ehdotus kirjaston rooliksi ja eri toimijoiden tehtäviksi	22
5.8.1	Datatyöryhmän ehdotus eri toimijoiden rooleista tutkimusdatan ja -aineistojen hallintaan liittyvissä tehtävissä	22
5.8.2	Kirjaston rooli tutkijan tukena tutkimusdatan hallinnassa.....	23
5.8.3	Datatyöryhmän ehdotus Helsingin yliopiston kirjaston roolista.....	24
6	Lähdeluettelo	25
7	Liite Tietojärjestelmäkartoitus	27
7.1	SciDB - Open Source Data Management and Analytics Software for Scientific Research.....	27
7.2	KiWi - Open Source development platform for building Semantic Social Media Applications.....	27
7.3	Semanttinen Mediawiki.....	27
7.4	Terrastore	28
7.5	OrientDB.....	28
7.6	MongoDB.....	28
7.7	SenseiDB.....	28
7.8	CKAN - the open source data portal software.....	29
7.9	DSpace.....	29
7.10	hydra	29
7.11	Archivematica	30
7.12	iRODS.....	30
8	Liite Haastattelukysymykset.....	31
9	Liite Kysymyksiä pilottiryhmälle.....	35

1 Datatyöryhmän ehdotukset

1.1 Tutkimusdatatyöryhmä ehdottaa, että

- 1) Kirjasto edistää ja tukee omalta osaltaan tutkimusdatan löydettävyyttä, saatavuutta ja jatkokäyttöä eri tieteenaloilla.
 - Kirjasto panostaa metadata- ja ontologiaosaamiseen
 - Tutkimusdata- tai sitä vastaava työryhmä kirjastossa jatkaa, kunnes datan hallinnointiin liittyvä tuki saadaan vakiinnutettua osaksi kirjaston toimintaa.
- 2) Kirjasto tukee tutkijoita tutkimusdatan hallintaan liittyvissä kysymyksissä. Tutkimuskäytännöt eri tieteenaloilla vaihtelevat ja siksi myös tutkimusdatan hallinnointiin liittyvän tuen tarve vaihtelee tieteenaloittain.
 - Kirjaston henkilöstön osaamista kehitetään niin, että kullakin kampuksella on vähintään yksi tietoasiantuntija, joka pystyy antamaan tukea tutkijoille tutkimusdatan hallinnassa. Näistä asiantuntijoista yksi on tietoteknisesti painottunut, ns. data librarian.
 - Tutkimusdataa ja sen hallintaa koskeva yleinen osaaminen ja tietämys kuuluvat jatkossa jokaisen kirjaston tutkimuksen tukipalveluissa työskentelevän osaamisalueelle. Asiakaspalvelussa tunnetaan dataan liittyvät tiedonlähteet ja tahot, joihin tiedon tarvisija voi ottaa yhteyttä.
- 3) Kirjasto tekee jatkossa tiiviimpää yhteistyötä yliopiston tutkijoiden/tutkimusryhmien kanssa, yliopiston sisällä (Tietotekniikkakeskus, Tutkimuksen toimiala, Kansalliskirjasto), kansallisesti (CSC:n koordinoima TTA-projekti) ja kansainvälisesti (OpenAIREplus) tutkimusdatan hallinnan edistämiseksi.
- 4) Kirjasto toimii aktiivisesti kansainvälisillä foorumeilla ja on eurooppalaisen OpenAIREplus-projektin NOAD (National Open Access Helpdesk) Suomessa.
- 5) Kirjasto toteuttaa omalta osaltaan EU:n suositusta julkisen varoin tuotetun datan avaamiseksi.
- 6) Helsingin yliopistossa on tulevaisuudessa oma datakatalogi tai tietojärjestelmä, johon tutkimusdatan metadata voidaan tallentaa mm. kansalliseen datakatalogiin haravoinnin mahdollistamiseksi. Se voi kytkeytyä TUHAT-järjestelmään ja muihin järjestelmiin.
 - Kirjasto tekee Tietotekniikkakeskukselle aloitteen projektista, jossa selvitetään erilaisia mahdollisuuksia rakentaa Helsingin yliopistolle oma tai ottaa käyttöön jokin olemassa oleva tietojärjestelmä datan hallinnointiin.
- 7) Helsingin yliopisto ja Helsingin yliopiston kirjasto edistävät pysyvien tunnisteiden käyttöönottoa hankkeille, julkaisuille, datalle ja tutkijoille.
 - TUHAT-järjestelmässä tulisi olla nykyistä näkyvämpi mahdollisuus ilmoittaa datan sijainti, sillä tämä nostaisi datan esille ja myös muiden tutkijoiden tietoisuuteen.
 - Kirjaston tulisi voida vaikuttaa tehokkaammin TUHAT-järjestelmän kehittämiseen.
- 8) Helsingin yliopisto tarjoaa tutkijoille ja tutkimusryhmille useita eri mahdollisuuksia tallentaa tutkimusdatansa yliopiston palvelimille, aiheenmukaisiin, kansallisiin tai kansainvälisiin data-arkistoihin (esim. IDA, FSD, Pangaea).
- 9) Yliopistolla tulisi olla selkeät ohjeet siitä, miten eri tieteen alueilla datan omistajuus määräytyy ja kuka on vastuussa tekijänoikeudellisten lupien hankkimisesta. Yliopistolla pitäisi olla myös selkeät ohjeet henkilötietolainsäädännön edellyttämien tutkimuslupien hankkimiseksi ja rekisteriselosteen laatimiseksi.
- 10) Jokaiseen tutkimussuunnitelmaan tulisi jatkossa sisältyä aineistonhallintasuunnitelma, jossa kuvataan se, miten kyseisen tutkimuksen aineistot, mm. primaari data hankitaan, miten sitä käytetään, säilytetään ja suojataan ja

miten sen myöhempi käyttö mahdollistetaan. Yliopistossa tulisi huolehtia siitä, että tutkijoilla on tiedossaan sellaiset kirjatut, selkeät käytännöt, että suunnitelman laatiminen on mahdollista. Tämä edellyttää useiden tahojen yhteistyötä.

2 Johdanto

Helsingin yliopiston datatyöryhmä (HY:n tutkimusdatan ja -aineistojen saatavuutta selvittävä työryhmä. [Rehtorin päätös 6.6.2008, 137/2008](#)) teki ehdotuksen, että yliopisto asettaisi toimielimen, jossa koottaisiin yhteen keskeiset tutkimusdatan hallinnointiin, käytön edistämiseen ja säilyttämiseen liittyvät toimijat ([Helsingin yliopiston datapolitiikka 2011](#)). Työryhmä esitti, että toimielimen tavoitteeksi asetettaisiin sellaisten toimenpiteiden valmistelu, joilla Helsingin yliopisto edistäisi integroidun informaatioympäristön toteutumista. Työryhmän toimielimen vetovastuuta ehdotettiin kirjastolle. Kirjastossa käynnistettiin tutkimusdataprojekti vuoden 2011 alkupuolella tähän ehdotukseen pohjautuen.

Kirjastossa on selvitetty tutkimusdatan hallintaan liittyviä haasteita jo aiemmissa projekteissa. Kirjasto koordinoi vuosina 2007-2009 Helsingin yliopiston keskustakampuksella mittavaa [Tieteellisten aineistojen verkkopalveluhanketta](#), jossa selvitettiin laitosten tutkimusaineistojen digitoinnin ja jatkokäytön edistämistä (Ristikartano 2010).

Vuonna 2009 Helsingin yliopiston kirjasto oli valmistelemaan [Helsingin yliopiston infrastruktuuria](#) vuosille 2010-2012 Tietotekniikkaosaston ja Kansalliskirjaston kanssa.

[Solmu-hanke](#) oli kirjaston tutkimus- ja kehittämishanke vuosina 2009-2011. Solmu-hankkeessa selvitettiin mm. sitä, voisiko kirjasto tukea tutkijaa tai tutkimusryhmää aineistonhallintasuunnitelman tekemisessä. Hankkeessa tehdyn tutkimuksen tehtävänä oli mallintaa, testata käytännössä ja arvioida kirjastotyöntekijöiden ja tutkimusryhmien uudenlaisen kumppanuustoiminnan, ns. solmutyöskentelyn toteuttamistapoja ja vaatimuksia (Engeström ym. 2010, Engeström ym. 2011)

Lisäksi niin kansallinen kuin kansainvälinen kehitys e-tutkimuksen ja tutkimusdatan aihealueilla on ollut kimmokkeena kehittää niin kirjaston kuin kirjastoammattilaisten osaamista tutkimusdatan hallinnassa. Kirjastossa ymmärretään, että tutkimuksen keskiössä on viime kädessä aina yksittäinen tutkija tai tutkimusryhmä. Kirjaston tavoitteena on, että se omalta osaltaan, yhteistyössä muiden toimijoiden kanssa kaikin tavoin helpottaa tutkijan työtä. Tässä tarkoituksessa kirjasto haluaa myös olla mukana edistämässä niitä toimia, joilla tutkimusdatan hallinnointiin liittyviä periaatteita ja käytäntöjä yliopistossa selkiytetään.

2.1 Toimeksianto

Helsingin yliopiston kirjaston ylikirjastonhoitaja Kaisa Sinikara asetti 31.5.2011 (Ylikirjastonhoitajan päätös 13/2011) työryhmän käynnistämään pilottihankkeen, jonka avulla selvitettiin tutkijoiden ja tutkimusryhmien ongelmia, esteitä ja kannustimia tutkimusdatan hallinnoimisessa. Työryhmän tuli koota yhteen keskeiset tutkimusdatan hallinnointiin, käytön edistämiseen ja säilyttämiseen liittyvät toimijat Helsingin yliopistossa ja tehdä ehdotus kirjaston roolista ja tehtävistä tässä kokonaisuudessa. Lisäksi ryhmän tuli tutkia ja selvittää niitä edellytyksiä, joilla integroitua informaatioympäristöä voidaan edistää ja tehdä ehdotuksia toimenpiteiksi, jotka mahdollistavat tämän päämäärän toteuttamisen.

Projektiryhmän kokoonpano:

Liisa Siipilehto, tietoasiantuntija, projektipäällikkö
Annikki Roos, kampuskirjastonjohtaja, projektin johtaja
Tiina Onttonen, tietoasiantuntija, projektisuunnittelija
Ilona Pietilä, projektisuunnittelija (2-10/2011)
Joonas Kesäniemi, tietojärjestelmäasiantuntija, pysyvä asiantuntija syksystä 2011 alkaen
Mari Elisa Kuusniemi, tietoasiantuntija, pysyvä asiantuntija syksystä 2011 alkaen
Marjut Salokannel, tutkimusjohtaja, oikeudellinen asiantuntija
Maria Forsman, johtava tietoasiantuntija
Eva Isaksson, kirjastonhoitaja
Kimmo Koskinen, kehityspäällikkö
Maria Kovero, erikoissuunnittelija
Monica Allardt, tietoasiantuntija
Pekka J. Salminen, kirjastos sihteeri

3 Pilottitutkimus

3.1 Tutkimusryhmän valitseminen pilotiksi

Projektissa päätettiin ottaa yksi tutkimusryhmä pilotiksi, jonka avulla selvittäisiin tutkijoiden ja tutkimusryhmän ongelmia, esteitä ja kannustimia tutkimusdatan hallinnassa. Sopivia tutkimusryhmiä pyydettiin ilmoittautumaan ja kolmesta ehdokkaasta valittiin Kumpulan kampukselta tutkimusaineistoltaan monipuolinen ja kooltaan keskikokoinen hanke. Hankkeen tutkimuksen tarkoituksena on selvittää Pyhäjärveen purkautuvaa pohjavettä, sen määrää, laatua ja vaikutusta järven pohjan kasvillisuuteen, vesitaseeseen, ravinnetaseeseen ja sitä kautta myös järven veden tilaan ja mm. kalastoon. Lisäksi tutkimusryhmän valintaa puolsi ryhmän halu olla tosissaan mukana ja varautuminen käyttämään siihen työaikaa.

3.2 Tutkimusryhmän esittely

Pilotiksi valitun tutkimusryhmän nimi on [Groundwater-surface water interaction at Lake Pyhäjärvi, SW Finland](#). Hanke on käynnistynyt vuonna 2008 ja päättyy 2013 ja oli siten haastatteluvaiheessa (keväänä 2011) noin puolessa välissä hankkeen kokonaiskestoajaksi. Tutkimusryhmään kuuluu yhteensä 11 tutkijaa/opiskelijaa neljästä eri organisaatiosta:

FT Kirsti Korkka-Niemi (Helsingin yliopisto, Geotieteiden ja maantieteen laitos, vastuullinen tutkija)
FM Anne Rautio (väitöskirja) (Helsingin yliopisto)
FM Päivi Niemistö (Turun Yliopisto, Geologian laitos)
Prof. Juha Karhu (Helsingin yliopisto)
Ossi Isola (maisterin tutkielma) (Helsingin yliopisto)
Andrew Wiebe (M.Sci) (University of Waterloo, Kanada)
Prof. Brewster Conant (University of Waterloo, Kanada)
Prof. David Rudolph (University of Waterloo, Kanada)
FT Anne-Mari Ventelä (Pyhäjärvi-Instituutti)
FT Marjo Tarvainen (Pyhäjärvi-Instituutti)
FM Teija Kirkkala (Pyhäjärvi-Instituutti)

Projektin työryhmä sai käyttöönsä tutkimusryhmän aineistot, tiedot ja haastattelut vastuulliselta tutkijalta Kirsti Korkka-Niemeltä ja väitöskirjan tekijältä Anne Rautiolta.

Tutkimusryhmä tekee vesistötutkimusta Säkylän Pyhäjärvellä ja siihen liittyvissä jokiympäristöissä. Pyhäjärvi on Lounais-Suomen suurin järvi Säkylän, Euran ja Yläneen kuntien alueella. Sen pinta-ala on 154 km² ja rantaviivaa on 80 km. Järven keskisyvyys on 5,4 metriä ja suurin syvyys 26 m. Ryhmä tutkii pintavesien ja pohjavesien vuorovaikutusta, ja sen vaikutusta vesistöön. Harjumuodostumissa on pohjavesivarastoja, jotka ovat eri tavoin yhteydessä alueen pintavesiin, ja näitä vuorovaikutussuhteita pyritään mallintamaan. Hanke on Helsingin yliopiston koordinoima, maantieteen, biologian, geologian ja hydrogeologian poikkitieteellinen hanke. Lisäksi hankkeessa on kansainvälisiä osallistujia. Hankkeelle on tehty tutkimussuunnitelma, data on pääosiltaan kerätty ja joitain artikkeleita on hankkeessa jo julkaistu konferenssipostereiden ja -esitysten lisäksi.

Hankkeessa on tehty helikopterivideokuvauksia ja lämpökamerakuvauksia FLIR-kameralla. Näihin kuvauksiin liittyy myös GPS-paikkatietoja. Vesinäytteitä otetaan kuukausittain, ja näytteitä otetaan myös pohjavedestä järven alta. Vesinäytteistä tutkitaan veden laatu, kemiallinen koostumus ja isotootit. Vesinäytteitä otetaan useista eri syvyyksistä, ja niistä muodostuu pitkiä aikasarjoja. Lisäksi tehdään veden virtaamamittauksia ja sedimentin lämpötilamittauksia eri syvyyksistä. Kairaus tietoja ja maaperätietoja on kerätty 30-40 m syvyydestä. Kairausaineisto on paperimuodossa kartoilla. Pyhäjärvi-instituutti on alkanut tehdä myös digitaalisia julkaisuja.

Hankkeessa kerätään HydroGeoSphere-mallin mukaista aineistoa, ja siinä käytetään myös Suomen ympäristökeskuksen CatchLake-hankkeen aineistoa. Aineistoon kuuluu myös säättietoja ja kasvillisuuskartoituksia. Määrällisesti eniten mittaustuloksia on vesinäytteistä ja lämpötilamittauksista, mutta eniten tallennustilaa vievät lämpökamerakuvaukset. Kenttäoloissa tehtäviä mittauksia ovat happi, pH, sähkönjohtavuus ja lämpötila. Laboratoriossa vesinäytteistä tutkitaan lisäksi anionit, sulfaattit, silikaatit ja ravinteet. Laboratorion mittalaitteet tallentavat tulokset Excel-, csv- tai PDF-muodossa. Tutkimusdataan kuuluu myös laserkeilausaineisto, joka on maanpintaa kuvaava kolmiulotteinen pistemäinen

aineisto. Koko aineiston perusteella tehtyjä valmiita tuotoksia ovat posterit, abstraktit, julkaisut ja mallit. Tutkimusryhmän datamäärä syksyllä 2011 oli yhteensä n. 25-30 Gt.

3.3 Työskentely tutkimusryhmän kanssa

Tutkimusryhmän tiedonhallinnan ongelmiin ja haasteisiin paneuduttiin haastattelemalla tutkimusryhmää. Haastattelutilaisuudessa tutkimusryhmä esitteli pilotiksi valitun tutkimushankkeensa kirjaston edustajille pp-esityksenä. Suomen Akatemian ja muutaman ulkomaisen rahoittajan aineistonhallintasuunnitelman ohjeita ja kysymyksiä käytettiin hyväksi laadittaessa tutkijoille [kysymyssettiä](#) alkuhaastattelua varten (liite 2).

Kirjaston projektisuunnittelijat perehtyivät pilottitutkimukseen syvällisemmin osallistumalla yhden päivän ajan maastomittauksiin Pyhäjärvellä tutkijoiden kanssa. Vesinäytteiden otto, näytteistä kertyvä mittausdata ja tutkijan työ konkretisoitui oleellisesti dataprojektissa työskenteleville tietoasiantuntijoille.

Dataprojektissa pilottihankkeeseen saatiin lisää tietoa vastuututkijalta saadusta tutkimussuunnitelmasta ja vertaisarvioidusta artikkelista. Dataprojektille hankittiin ulkoinen kovalevy (2T), johon tutkijat siirsivät siihen mennessä kerätyn datan projektisuunnittelijoiden tarkastelua varten. Eri tiedostotyyppit listattiin ja tiedostojen koot summattiin datamäärän arvioimiseksi.

Dataprojektissa tehtiin selvitys datan tallentamiseen sopivista tietojärjestelmistä Helsingin yliopistossa. Tietojärjestelmäselvitys ja tutkimusryhmän uusi haastattelu tehtiin yhdessä Tietotekniikkakeskuksen tietotekniikka-asiantuntijan Petteri Hemmilän kanssa. Tutkijoita pyydettiin miettimään kriteereitä ja vaatimuksia, jotka rakennettavan järjestelmän tulisi heidän mielestään täyttää ([haastattelukysymykset](#), liite 3).

Tutkimusryhmän tutkijoita haastateltiin kolmannen kerran joulukuussa 2011. Haastattelut suoritti Joonas Kesäniemi osana Pro gradu -tutkielmaa, joka käsittelee tutkimusaineistojen provenienssitiedon mallintamista semanttisen webin teknologioiden avulla. Opinnäytteen tavoitteena on luoda formaali malli kyseisen tutkimusryhmän toimintaan liittyvästä datasta, prosesseista ja toimijoista sekä näiden välisistä suhteista. Tätä varten haastattelujen sisältöä oli laajennettava pelkistä aineistoista koko tutkimusprosessin kattavaan tehtävien, ihmisten, paikkojen, laitteiden ja ohjelmistojen käsittelyyn. Haastattelujen ja muun työryhmän keräämän materiaalin pohjalta on tarkoituksena luoda käsitelmä, jonka avulla on mahdollista kuvata esimerkiksi tietystä vesinäytteestä saadun raakadatan polku osaksi monimutkaista kolmiulotteista visualisointia. Sama onnistuu tietysti myös toisinpäin. Voidaan siis lähteä visualisoinnista liikkeelle ja kaivaa esille kaikki näytteet, joista analysoitua tai mitattua raakadataa on hyödynnetty tietyn kuvan muodostamisessa. Mallin tarkka kuvaus ja sen implementointi semanttisen webin ontologiana julkaistaan myöhemmin tämän vuoden aikana osana Joonas Kesäniemen Pro gradu -tutkielmaa. On hyvä huomata, että kyseessä on tässä vaiheessa vasta käsitteellinen malli, joka ei siis ota kantaa siihen, miten mallin määrittelemää provenienssidataa käytännössä kerätään. Malli ja varsinkin sen ontologiakuvaus voi kuitenkin toimia pohjana esimerkiksi semanttisessa wiki-ympäristössä toimivalle tallennusjärjestelmälle.

Vuoden 2012 alussa pilottitutkimusryhmä osallistui CSC:n IDA-tallennuspalvelun testikäyttöön. IDA-palvelun esittelytilaisuuteen osallistuivat tutkijoiden lisäksi Tietotekniikkakeskuksen ja kirjaston dataprojektityöryhmän asiantuntijoita.

4 Yhteistyö muiden toimijoiden kanssa

4.1 Yhteiskuntatieteellinen tietoarkisto

Helsingin yliopiston kirjaston tutkimusdatatyöryhmän jäsenet verkostoituivat heti hankkeen alkuvaiheessa tietoarkistosaajien kanssa matkustamalla Tampereelle tutustumaan Yhteiskuntatieteelliseen tietoarkistoon. Tietoarkistoa esittelivät kehittämispäällikkö Arja Kuula, informaattikko Helena Laaksonen, tietopalvelupäällikkö Hannele Keckman-Koivuniemi ja tietopalvelupäällikkö Mari Kleemola.

Yhteiskuntatieteellinen tietoarkisto (FSD) perustettiin vuonna 1999 Suomen Akatemian aloitteesta. Tietoarkisto toimii Tampereen yliopiston yhteydessä, mutta palvelut ovat valtakunnallisia. Elektronisten tutkimusaineistojen arkistointi ja käyttö ovat maksuttomia myös esimerkiksi Helsingin yliopiston tutkijoille, koska toiminnan rahoitus saadaan opetus- ja kulttuuriministeriöltä. Tietoarkistolle on tähän mennessä tullut Helsingin yliopistosta noin 1000 aineistopyyntöä. Helsingin yliopiston tutkijoiden keräämiä aineistoja on tallennettu arkistoon n. 60 kpl.

Tietoarkiston palvelut ovat Suomen Akatemialta saadun viisivuotisen Tietoarkisto 2015 -hankeapurahan myötä laajenemassa psykologiaan, historiaan, kulttuurien tutkimukseen ja terveystieteisiin. Terveystieteellisiä aineistoja arkistossa jo onkin. Aineistot ovat pääsääntöisesti haastatteluin (mm. kyselylomakkeiden avulla) ja havainnoimalla koottuja.

Datatyöryhmä ehdottaa, että kirjasto tiedottaa Helsingin yliopiston tutkijoille mahdollisuudesta tallentaa omia tutkimusaineistojaan Yhteiskuntatieteelliseen tietoarkistoon niillä tieteenaloilla, jotka soveltuvat aineistoiltaan sinne.

4.2 Jyväskylän yliopisto

Toukokuussa 2011 Tutkimusdataprojektin jäsenet tapasivat julkaisukoordinaattori Pekka Olsbon ja tietohallintokeskuksen sovellussuunnittelija Joonas Kesäniemen Jyväskylän yliopistosta. Palaverissa kuultiin Jyväskylän yliopiston rehtorin asettaman (22.10.2009) työryhmän tutkimusaineistojen hallintaan liittyvästä selvityksestä. Hankkeen yhtenä päämääränä oli laatia sopimusmalleja, joissa sovitaan tutkimusdataan liittyvästä vastuusta eri toimijoiden kesken. Aineiston jatkokäyttöä pyritään mahdollistamaan jo ennen datan keruuvaihetta solmittavilla sopimuksilla. Datatyöryhmä sai tutkittavakseen Jyväskylän yliopistossa tehtyjä sopimusmalliluonnoksia.

4.3 Tietotekniikkakeskus

Yhteistyötä tehtiin Helsingin yliopiston Tietotekniikkakeskuksen kanssa koko projektin ajan. Kuukausittain tutkimusryhmät kysyvät Tietotekniikkakeskukselta, miten he saisivat datansa verkkoon ja hallittua sen, ketkä pääsevät käsiksi dataan. Osallistuessaan Helsingin yliopiston tutkimusdataryhmään Tietotekniikkakeskus teki selvityksen tarvittavasta levytilasta ja kustannuksista tutkimusdatan säilyttämiseen. Kirjaston tutkimusdataprojektissa valmistunut selvitys datan- tai metadatan hallintaan sopivista tietojärjestelmistä toimitettiin myös tietotekniikkapäällikkö Teo Kirkiselle. Tietotekniikka-asiantuntija Petteri Hemmilä haastatteli Kimmo Koskisen kanssa pilottitutkimusryhmää tietojärjestelmäselvitystä varten.

Tietotekniikkakeskuksen teknologiapalveluissa valmistellaan usean tyyppisiä tallennusratkaisuja, joiden avulla laitos tai tutkimusryhmä voi joko hankkia tai ostaa niille sopivan kokoista ja -laatuista tallennustilaa. Tallennuspalveluja voidaan tuottaa Helsingin yliopiston omasta tai ulkoisten palvelutarjoajien infrastruktuurista. Tietotekniikkakeskuksen pilottihankkeen etenemisestä saatiin tietoa Ville Tenhuselta ja Seppo Syrjäselältä. Lisäksi Kimmo Koskinen kirjaston Verkkopalveluista osallistui Tietotekniikkakeskuksen hankkeen seurantaryhmään.

Tietotekniikkakeskuksen keskeinen rooli tutkimusdatan tallentamisen infrastruktuuripalvelujen tuottajana, jatkuvan yhteydenpidon ja yhteiskehittelyn hyöty selkiytyi tutkimusdatatyöryhmälle kirjaston dataprojektin aikana.

4.4 Tieteen tietotekniikan keskus CSC

Kirjaston tutkimusdataprojekti sai kutsun johtaja Pirjo-Leena Forsströmiltä CSC:n asiantuntijapalaveriin. Tietoja vaihdettiin puolin ja toisin. Dataprojekti sai kuulla ajankohtaista tietoa niin kansallisista kuin EU:n tasolla alkavista dataan liittyvistä hankkeista johtaja Juha Haatajan esittelemänä. Kehityspäällikkö Kimmo Koivunen CSC:stä kertoi pitkäaikaissäilytysjärjestelmäprojektistä. Säännölliset tapaamiset johtivat mm. siihen, että dataprojektin pilottitutkimusryhmä sai käyttöluvut uuden IDA-tallennuspalvelun testikäyttöä varten. Projektipäällikkö Panu Kalliokoski toimi tässä CSC:n yhteyshenkilönä. Helsingin yliopiston kirjaston edustaja osallistuu CSC:n koordinoimaan Tutkimuksen tietoaaineistot -hankkeen ohjausryhmään. Yhteistyötä kirjaston ja CSC:n välillä voisi olla myös datanhallinnan koulutus- ja opastusmateriaalin työstämisessä.

4.5 Kansalliskirjasto

Erytisiantuntija Juha Hakala Kansalliskirjastosta vieraili tutkimusdataprojektin palaverissa syyskuussa 2011. Aiheena olivat mm. PAS (pitkäaikaissäilytys) ja pysyvät tunnisteet. Juha Hakala konsultoi dataprojektia myös toisessa palaverissa ja piti projektisuunnittelijat ajan tasalla pitkäaikaissäilytykseen liittyvien ajankohtaisten tapahtumien suhteen. Kansalliskirjasto toivoo kirjastojen panosta mm. tutkimusaineistojen pitkäaikaissäilytyksen tiedostomuotojen ja formaattien suunnitteluun sekä osallistumista viestintään tulevaisuudessa.

4.6 Helsingin yliopiston Tutkimushallinnon yksikkö, tutkimuksen toimiala

Helsingin yliopiston hallintopalveluiden Tutkimuksen toimialan ja tutkimusdataprojektin palaverissa haarukoitiin data-asioihin liittyviä tehtäviä ja mietittiin toimintojen rajapintoja. Tutkimuksen tiedonhallintaan liittyvät palvelut ja niiden edistäminen kuuluvat Tutkimuksen toimialan tehtäviin. Meneillään olevan Tuhat-järjestelmän kehittämisen lisäksi tutkimusdataprojektin nähtiin liittyvän tutkijan ja laitoksen projektisalkun sekä tutkimusprojektin elinkaaren kehittämiseen. Tutkimushallinnon yksikkö vastaa myös tutkimuksen infrastruktuuriin liittyvästä valmistelusta. Tutkimusdataprojekti kuului kehityspäällikkö Maritta Kaltiokumpua, tutkimushallinnon asiantuntija Marko Peuraa ja projektipäällikkö Aija Kaiteraa.

5 Projektin tulokset

Tulokset esitellään kappaleissa:

- Havaitut ongelmat ja esteet datan hallinnassa
- Aineistonhallintasuunnitelma
- Oikeudellisia kysymyksiä
- Tietojärjestelmät ja palvelut datanhallintaan
- Tutkimusdatan metadata
- Julkaisujen linkitys tutkimusdataan
- Osaamisen kehittäminen
- Ehdotus kirjaston rooliksi ja eri toimijoiden tehtäviksi

5.1 Havaitut ongelmat ja esteet datan hallinnassa

Havainnot ongelmista ja esteistä on kerätty pilotiksi valitun tutkimusryhmän haastatteluista.

Pilottihankkeen tutkimusryhmä koki aineistohallinnassa suurimmaksi ongelmaksi ohjeiden, suunnittelun, datanhallintaresurssien ja käytännön tuen puutteen. Tutkijat toivoivat tietoa siitä, mihin paikkaan, millaista aineistoa ja kuinka paljon dataa voi tallentaa. Datarakenteen ja tiedostokuvausten osalta olisi tarvetta apuun ja tukeen jo tutkimusprojektin suunnitteluvaiheessa. Tukea tarvittaisiin myös lupa-asioiden hoitamiseen, kuten datan omistus- ja käyttöoikeuksiin mm. projektin käynnistys- ja lopetusvaiheiden kirjauksiin sopimuksista. Miten sovitaan datan jatkokäytöstä esim. yksittäisten tutkijoiden lähtiessä pois tutkimusryhmästä?

"Tutkimusryhmän keräämä raakadata ei sinänsä ole tilaa vievää, mutta käsiteltynä tiedostokoot kasvavat. Lämpökamerakuvausten datamäärä tulee olemaan hyvin suuri. Matemaattisten mallien kanssa joudutaan miettimään tallennuskapasiteetin riittävyttä. Aineistot ovat kunkin tutkijan omalla koneella. Atk-vastaavalta kysytty, on mahdollista saada tilaa, mutta kun koko menee tietyn rajan yli, ei enää olisi kätevää käyttää laitoksen verkkolevyä. Käytettävä ulkoista kovalevyä, josta otettava itse varmuuskopioita. Turun yliopistolla tutkimushankkeella oli oma ryhmälevy" (Tutkimusryhmän haastattelu 6.5.2011). Tämä ongelma on poistumassa, kun lisälevytilan hankkiminen laitoksille on edullisempaa kuin aiemmin.

"Datan lähettäminen Suomen ja Kanadan välillä on ollut ongelmallista, sillä mitään toimivaa yhteistä data repositoryä ei ole käytössä, ja dataa onkin lähetetty postissa rompuilla kun tiedostot ovat liian isoja sähköpostitse lähetettäviksi." - On olemassa ohjelmia (esim. FileSender, Dropbox), joiden avulla suuriakin tiedostoja voi lähettää. Näitä ei heikon tietoturvan takia suositella, mutta niitä käytetään silti yleisesti. CSC:n tuottamassa IDA-tallennuspalvelussa testataan, miten tutkimusryhmän yhteiseen tiedostohakemistoon päästään muualta Suomesta ja myös pilottitutkimuksen tapauksessa Kanadasta.

"Datarakenteen ja tiedostokuvausten osalta olisi tarvetta apuun ja tukeen jo tutkimusprojektin suunnitteluvaiheessa: malleja helpottamaan ja yhtenäistämään datan tallentamista." Tiedostojen informatiivinen ja yhtenäinen nimeäminen tutkimusryhmän sisällä koettiin ongelmalliseksi. Tiedostolistauksen tekemiseen ja tiedostojen sisällön tallentamiseen toivottiin neuvoja. Tiedostojen eri versioidenkin dokumentointi voisi auttaa tutkijaa löytämään aiemmin tehdyn työn. Pilottitutkimuksen perusteella voidaan sanoa, että tutkijat tarvitsevat tukea erityisesti metadatatallennuksen suunnitteluun, esimerkiksi muuttujien kuvaamiseen ja sopivien metadatatandardien valitsemiseen. Myös toimivien toimintaprosessien kehittämiseen tarvitaan tukea, esimerkiksi ohjeistuksen osalta.

Tutkimusryhmältä kysyttiin, haluttaisiinko hankkeen keräämä data tallentaa avoimeen kotimaiseen tai kansainväliseen data-arkistoon: *"Lämpökamerakuvat kiinnostanevat ihmisiä yleisestikin, mutta mieluummin kuvina kuin vaikeasti hahmotettavina videoina. Lämpökameravideon katsominen vaatii lämpökamerahjelman." - Tutkimusaineistojen jakaminen jo tutkimusryhmän jäsenten kesken saattaa olla ongelmallista, sillä laitteiden omilla ohjelmistoilla tuotettu materiaali vaatii ohjelman asennuksen kunkin omalle koneelle. Tässä tapauksessa kirjasto löysi avoimesti käytettävän kuvankatseluohjelmiston, joka soveltui tutkimusryhmän lämpökamerakuvien katseluun.*

Pitkäaikaissäilytystarpeitakin tuli esille tutkimuspilotin haastattelusta: *"Tarve pitkäaikaissäilytykselle on erittäin suuri. Tehtävät mittaustulokset ovat täysin relevantteja useammankin vuosikymmenen kuluttua (esim. ilmastonmuutostutkimus)."* *"Maaperäkairausaineistot ovat paperilla, isoja (ainakin suurempia kuin A3), pitäisi skannata, sillä aineistot ovat vain lainassa."* - Helsingin yliopistolla ei ole digitointipalveluja tarjolla, eikä Kansalliskirjastollakaan ole resursseja tehdä pieniä projekteja. Keskustakampanjalla digitointia ja tutkimusdataa selvittävä hanke (Tieteellisten aineistojen verkkopalveluhanke) esitti, että Helsingin yliopiston tulisi kilpailuttaa digitointipalvelut, jotta tutkimusryhmät voisivat ostaa palvelut tarvittaessa.

Haastattelussa tuli ilmi, että tutkimusdatat tulivat vastuututkijan mukana Turun yliopistosta ja ovat nyt Helsingin yliopiston omistuksessa. Jatkossa omistajuus- ja vastuukysymykset pitäisi tutkijoiden mukaan tarkastella selkeämmin. Aineistoista ei ole tehty omistussopimuksia, mutta graduntekijöiden kanssa on sovittu suullisesti, että heidän keräämänsä data jää tutkimushankkeelle.

Hankkeessa on jouduttu miettimään myös yksityisyydensuojaan liittyviä kysymyksiä mm. lämpökamerakuvausten suunnittelussa ja yksityisten henkilöiden vesikaivojen tutkimustietojen yhteydessä.

5.1.1 Aineistohallintasuunnitelma

Useat tutkimusrahoittajat, mm. Suomen Akatemia, edellyttävät [aineistohallintasuunnitelmaa](#) osana tutkimussuunnitelmaa ([Helsingin yliopiston datapolitiikka 2011](#)). Aineistohallinnan suunnittelun tarkoituksena on varmistaa, että tutkimusaineistojen suhteen noudatetaan [hyvää tieteellistä käytäntöä](#), että tutkimusaineisto ei vaarannu missään vaiheessa ja että informoitu jatkokäyttö on mahdollista. ([Tutkimusaineistojen tiedonhallinnan käsikirja](#)).

Aineistohallintasuunnitelmaa varten Suomen Akatemia ohjeistaa apurahan hakijoita vain muutamalla, mutta hyvin keskeisellä kysymyksellä. Akatemia edellyttää selvitystä siitä, miten hankkeen tutkimusaineistot tullaan hankkimaan, miten niitä käytetään, säilytetään ja suojataan ja miten niiden myöhempi käyttö mahdollistetaan. Selvityksestä tulee lisäksi käydä ilmi, mitkä ovat hankkeen käyttämien ja synnyttämien aineistojen omistus- ja käyttöoikeudet. Tämä aineistohallintasuunnitelma esitetään tutkimussuunnitelmassa. Akatemia suosittelee, että hankkeet luovuttavat kokoamansa yhteiskuntatieteellisen aineiston [Yhteiskuntatieteellisen tietoaarkiston](#) käyttöön. Samoin Akatemia suosittelee, että hankkeissa luodut kieliaineistot saatetaan muiden tutkijoiden käyttöön [FIN-CLARIN](#) -järjestelmän kautta.” ([Suomen Akatemian hakuohjeet](#))

Pilottihankkeessa ei ollut sellaisia rahoittajia tai muuta tahoa, jotka olisivat vaatineet aineistohallintasuunnitelmaa. Mutta jo alkuhaastatteluvaiheessa tutkijat toivat esille suunnitelman tarpeellisuuden ja siitä saatavan hyödyn. Tutkimusryhmässä todettiin hankkeen aineiston hallinnan olevan hyvin haasteellista. Mitään käytännön ohjeistusta suunnitelman laatimiseen ei ollut saatavilla ja siihen kaivattiin tukea.

Helsingin yliopiston kirjaston Solmu-projektissa (2009-2011) suunniteltiin kirjaston palvelua, joka tarjoaisi tutkijoille aineistohallintasuunnitelman tekoon tukea. Ajatus oli sinänsä hyvä, mutta vielä tällä hetkellä Helsingin yliopiston tutkijoiden on vaikea vastata aineistohallintasuunnitelman edellytyksiin, eikä kirjastollakaan ole antaa niihin vastauksia tai neuvoja. Jo hankkeen käyttämien ja synnyttämien aineistojen omistus- ja käyttöoikeuksista voi olla vaikea sopia yleisten periaatteiden puuttuessa. Aineistojen tallennus tutkimusvaiheessakin koko tutkimusryhmän käyttöön on ollut vaikeaa yhteisen levytilan puuttuessa, eikä aineistojen jakaminenkaan ole ollut mahdollista. Dataprojektin aikana moni asia edistyi, esim. CSC:n tarjoama [IDA-tallennuspalvelu](#) saatiin jo testikäyttöön. Myös Tietotekniikkakeskuksen omassa hankkeessa datan tekniseen säilyttämiseen ollaan kehittämässä useita ratkaisumalleja.

[Solmu-projektissa](#) kirjaston tietoasiantuntijat seurasivat yhden tutkijan apurahahakemuksen aineistohallintasuunnitelman tekoa ja antoivat linkkejä, mistä tiedostojen ja aineistojen hallintaan löytää neuvoja (mm. [Ohjeita aineistohallintasuunnitelmaan](#), [Tutkimusaineistojen tiedonhallinnan käsikirja](#)). Samassa yhteydessä tutkija myös kertoi hankesuunnitelmastaan, olemassa olevista ja kerättävästä tutkimusdatasta ja suunnitelluista säilytyspaikoista. Tutkija piirsi myös kaavion aineistokulusta. Prosessin aikana tutkija paneutui juuri niihin kysymyksiin, joita Akatemia esittää apurahaohjeissaan. Tuloksena syntyi esseemuotoinen suunnitelma, jota havainnollisti datakaavio. Solmu-hankkeesta saadun kokemuksen perusteella kirjaston henkilökunnan tuki voisi olla neuvova ja kysymyksiä esittävä ns. fasilitioija, joka kuuntelee tutkijan esityksiä ja pohdintoja. Lopullisen suunnitelman voi tehdä vain tutkija itse. Kirjaston rooli aineistohallintasuunnitelman osalta voisi olla myös sellainen, että tietoasiantuntijat seuraavat kansainvälistä ja kansallista kehitystä, esittävät erilaisia malleja ja työkaluja suunnitelman apuvälineiksi ja tuovat niitä tiedoksi tutkijoille, tutkimusryhmille ja tutkimuspalvelujen suunnittelijoille.

Esimerkkejä aineistohallintasuunnitelmaa edellyttävistä rahoittajista Isossa Britanniassa: <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>

5.2 Oikeudellisia kysymyksiä

Pilottiryhmän työssä nousi esille useita datan keräämiseen ja käyttöön liittyviä oikeudellisia kysymyksiä. Niistä keskeisin oli dataan liittyvät oikeudet, ennen kaikkea kysymys siitä, kuka määrää datan käytöstä. Jos tekijänoikeudet dataa sisältävään tietokantaan ovat vastuullisella tutkijalla, voiko hän ottaa datan mukaansa vaihtaessa työpaikkaa. Vielä komplisoidummaksi asia tulee, jos uusi työpaikka sijaitsee EU:n ulkopuolella, jonne henkilötietojen siirtoa on rajoitettu. Seuraavassa vastataan yleisellä tasolla näihin kysymyksiin, mutta viime kädessä käytännöt on luotava kunkin tieteenalan erityispiirteet huomioon ottaen.

5.2.1 Primäärin tutkimusdatan tekijänoikeudellinen status

Primääri tutkimusdata on tutkimuksessa käytettävää raakadataa digitaalisessa, kirjallisessa, visuaalisessa tai audiovisuaalisessa muodossa. Kyseessä voi olla havainto- ja mittaustuloksia, simuloituja tuloksia, haastatteluja, kyselyaineistoja, nauhoituksia, kuvia ja rekisteritietoja tai näiden yhdistelmiä. Arvioitaessa tekijänoikeudelliselta kannalta tutkimusdatan statusta, on kysyttävä, onko kyseessä tutkijan omaperäisen ilmaisun tai tutkijan näkemyksen leimaama teos. Näin voi olla esimerkiksi audiovisuaalisen datan ja haastatteluaineistojen kohdalla, jolloin voi syntyä tekijänoikeuslain mukainen elokuvateos tai kirjallinen teos.

Sen sijaan, jos kyseessä on puhdas mittausdata tai vastaava automaattisesti kerätty data, tähän dataan sellaisenaan ei kenelläkään ole tekijänoikeutta. Sen sijaan sitä tietokantaa, mihin data on kerätty voidaan tietyn edellytyksin suojata tekijänoikeuslain avulla joko kirjallisena teoksena, tekijänoikeuslaissa annetun erityissuojan perusteella, ns. sui generis tietokantasuoja, tai luettelosuojan perusteella.

5.2.2 Tutkimusaineiston omistajuus

Tarkasteltaessa aineiston omistajuuteen liittyviä kysymyksiä on ensiksi tehtävä ero **fyysiseen materiaaliin** liittyvän omistajuuden ja **digitaalisessa muodossa olevan tiedon** mahdollisen omistajuuden välillä. Tosin fyysiseenkin materiaaliin, kuten julkaisuun, voi liittyä tekijänoikeuksia. Yleensä fyysisten tutkimustyökalujen ja muiden välineiden omistajuuteen ei yliopistolla liity ongelmia vaan ne määräytyvät yleisten ohjeiden mukaisesti. Sen sijaan tutkimusdatan ja muun aineiston immateriaalioikeudellinen suoja on osoittautunut monimutkaisemmaksi asiaksi.

Tekijänoikeus suojaa tekijän henkisen luomistyön tulosta, sitä muotoa minkä tekijä ideoille antaa. Tekijänoikeus ei suojaa tutkimusideoita tai tosiseikkoja. Kukaan ei voi omistaa ideoita tai luonnossa esiintyviä tosiseikkoja, vaan **tekijänoikeudellisen suojan synty edellyttää aktiivista tekijän intellektuaalista panosta**. Käytännössä tämä niin sanottu tekijänoikeudellinen teostaso on ollut hyvin matala.

5.2.3 Tekijänoikeudellisen suojan monitasoisuus: tutkimusdataa sisältävien tietokantojen suoja

1) kirjallisena teoksena suojattava tietokanta

- jos tietokannan sisällön valikointi ja järjestely osoittaa sen **valmistajan omaperäistä henkistä luomistyötä**, tietokantaa voidaan suojata tekijänoikeudellisena teoksena.
- tekijällä on lain mukaiset yksinoikeudet määrätä tietokannan käytöstä.

2) sui generis tietokantasuoja

- jos tietokannan sisällön keräämiseen, valmistamiseen tai esittämiseen on käytetty huomattava työmäärä, sitä voidaan suojata sui generis tietokantasuojan avulla.
- ei edellytä minkäänlaista tekijän omaperäistä panosta
- ei sovellu automaattisen tietojenkeruun tuloksena syntyneeseen esim. raakadataa sisältävään tietokantaan
- valmistajalla on yksinomainen oikeus määrätä tietokannan laadullisesti tai määrällisesti arvioiden olennaisesta osasta valmistamalla siitä kappaleita ja saattamalla se yleisön saataviin.

3) luettelosuojaan suojaamat tietokannat

- jos tietokantaan kerätty suuri määrä tietoja, sen valmistajalla on yksinomainen oikeus määrätä tietokannan laadullisesti tai määrällisesti arvioiden olennaisesta osasta valmistamalla siitä kappaleita ja saattamalla se yleisön saataviin.
- luettelosuoja ei ole voimassa Pohjoismaiden ulkopuolella.

5.2.4 Työsuhteessa syntyneet tietokannat

Oikeudet työsuhteessa syntyneisiin tietokoneohjelmiin ja tietokantoihin siirtyvät suoraan lain nojalla **työnantajalle**. Tämä ei kuitenkaan koske korkeakoulun opetus- tai tutkimustyössä itsenäisen tutkimustyön tuloksena syntyneitä tietokoneohjelmaa tai tietokantaa. (TekijäL 40b§)

- yliopistolla on tieteenalakohtaisesti keskusteltava, siirtyvätkö oikeudet tietokantoihin lain perusteella yliopistolle vai katsotaanko niiden syntyneen pääsääntöisesti itsenäisen tutkimustyön tuloksena?
- kuvailutiedot täytyy aina olla julkisesti saatavilla. Kuvailutietojen laatiminen on lähtökohtaisesti tutkijan tehtävä.

Datan omistajuuden toteaminen tärkeää, jotta tiedetään millä taholla on oikeus määrätä sen käytöstä ja turvataan jatkokäytön oikeusvarmuus. Datan julkaiseminen voi myös olla meriitti tutkijanuralla.

Mitä tutkimusdataalle tapahtuu työsuhteen/tutkimuksen päättyessä?

- oikeudet dataan on määriteltävä kunkin tutkimushankkeen ja tieteenalan käytäntöjen mukaisesti
- on sovittava, voiko tutkija viedä datan mukanaan esimerkiksi omaan yritykseen tai toiseen tutkimuslaitokseen? Työpaikan vaihdon yhteydessä on otettava huomioon myös kilpailuoikeudelliset näkökohdat.
- datan jatkokäytöstä sovittava erikseen ellei ole olemassa laitoskohtaisia sääntöjä tai käytäntöjä.

5.2.5 Yksityisyyden suoja

Aina kun tutkimuksessa käytetään aineistoa, johon sisältyy tietoja, joiden perusteella tietty henkilö voidaan tunnistaa joko suoraan tai välillisesti, henkilötietojen käsittelyyn sovelletaan henkilötietolakia. Kerätessä tällaista tietoa tutkimusta varten tutkimukseen osallistuvilta henkilöiltä on pyydettävä suostumus heidän tietojensa käyttämiseen tutkimuksessa. Tämän suostumuksen on sisällytettävä kaikki mahdolliset tavat millä henkilötietoja on tarkoitus käyttää. Henkilötietojen käyttö jatkotutkimuksessa tai niiden yhdistäminen muihin, esim. rekisteritietoihin, edellyttää myös suostumusta. Lääketieteelliseen tutkimukseen vaaditaan nimenomainen, kirjallinen lupa. Muuhunkin tutkimukseen suositellaan, että tutkimuskohteiltaan pyydetty suostumus olisi kirjallinen. Suostumusta pyydetessä on tutkimuskohteelle selkeästi ilmoitettava tietojenkäsittelyn tarkoitus, henkilötietojen mahdollinen säilytys ja hävittäminen sekä mahdollinen aikomus luovuttaa tietoja kolmannelle taholle, joka edellyttää nimenomaista suostumusta. Jos henkilötietoja on tarkoitus käyttää muuhun tarkoitukseen tai toista tutkimushanketta varten, tähän on myös saatava suostumus. Mikäli henkilötietoja aiotaan arkistoida, tähän on myös saatava tutkimuskohteiden suostumus.

EU:n komissio julkisti tammikuussa 2012 ehdotuksen tietosuojasetukseksi, jolla kumotaan olemassa olevat henkilötietojen suoja koskevat direktiivit, ja yhtenäistetään jäsenvaltioiden henkilötietolainsäädäntö. Sen jälkeen kun asetus on hyväksytty, se tulee voimaan 18 kuukauden kuluttua ja on tällöin suoraan sovellettavaa oikeutta kaikissa jäsenvaltioissa. Asetus tulee muuttamaan myös Suomen henkilötietolakia.

5.2.6 Patentit

Mahdollisesti tutkimustuloksille haettavat patentit on otettava huomioon, kun harkitaan tutkimusdatan ja muiden aineistojen asettamista laajemmin saataville. Tutkimusaineistoja ei pidä julkistaa ennen kuin patenttihakemus on tehty ja pidettävä muutenkin huoli siitä, ettei datan julkistaminen vaaranna patentin myöntämistä ja hyödyntämistä.

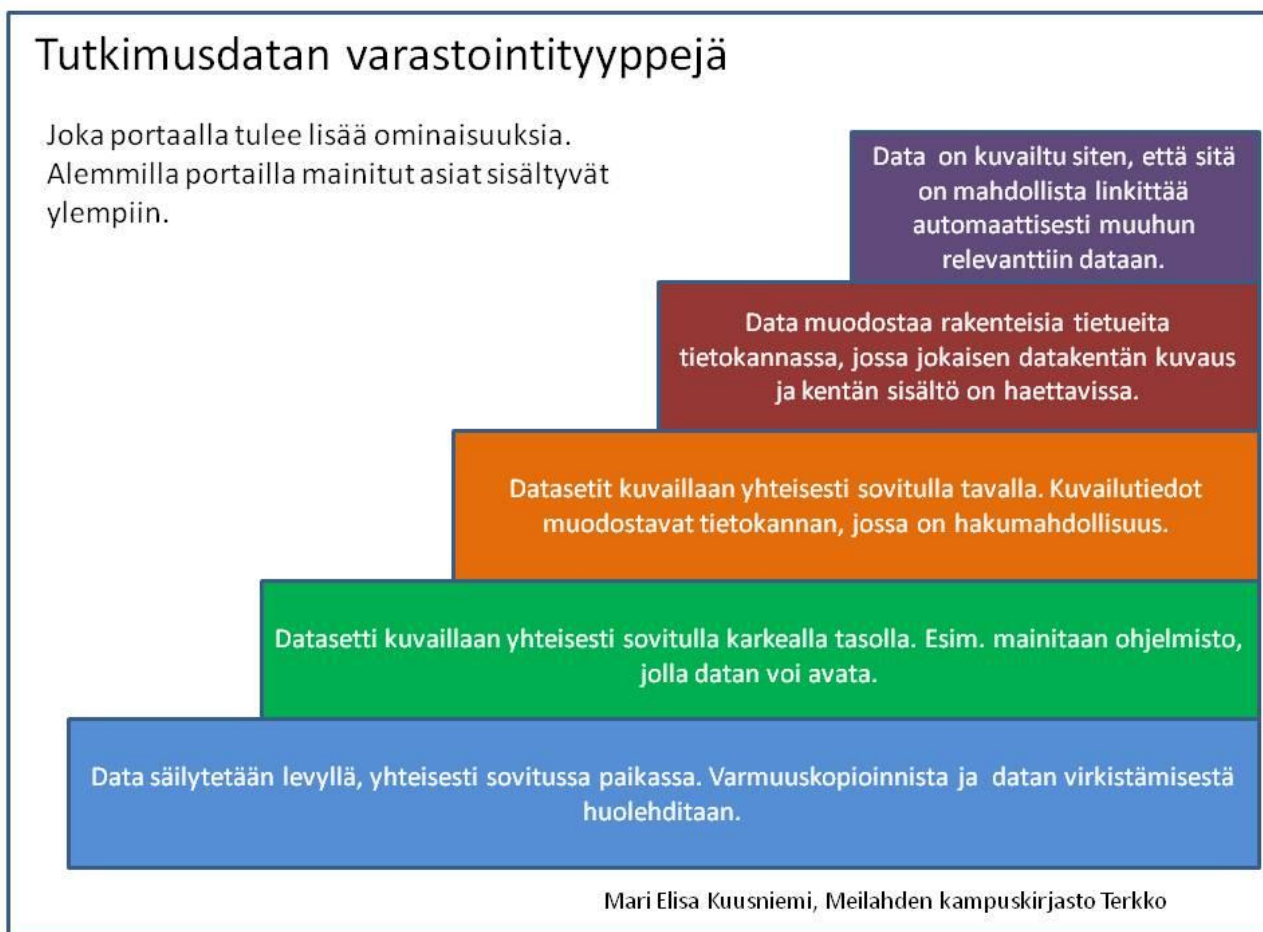
5.3 Tietojärjestelmät ja palvelut datanhallintaan

Tutkimusdatan hallinnointiin on kehitetty monia erilaisia tietojärjestelmiä. Kaikkien tietojärjestelmien tavoitteena on tarjota mahdollisuus tiedon tallentamiseen ja säilyttämiseen. Useimmat järjestelmät mahdollistavat myös tiedonhaun.

Tutkimusdatan analysointiin on kehitetty erilaisia tietojärjestelmiä ja ohjelmistoja, mutta näihin datatyöryhmä ei ota kantaa, vaan keskittyy datan säilyttämiseen liittyviin kysymyksiin. Säilyttämistavat voidaan jakaa karkeasti kahteen tyyppiin: välivarastointiin ja pitkäaikaissäilytykseen. Pilotin mukaan tutkimuksen aikana data haluttiin tallettaa välivarastoon esim. tutkimusryhmän yhteiskäyttöön jaetulle levyille tai varmuuskopiointia varten.

5.3.1 Tutkimusdatan varastointityyppejä

Datan tallettamista varten tarvitaan tutkimusdata-arkistoja. Tiettyjen alojen valtavat datamassat ovat saaneet aikaan kansainvälistä yhteistyötä ja sen tuloksena syntyneitä data-arkistoja. Myös Helsingin yliopistossa talletetaan rutiinisti esim. geenidataa kansainvälisiin geenipankkeihin, jotka ovat eräänlaisia data-arkistoja. Tutkimusdatalle on luotu myös kansallisesti ja paikallisesti erilaisia pienempiä data-arkistoja. Tällaisten data-arkistojen vaatimat formaatit, metadatatallit ja annotointikäytännöt vaihtelevat. Tietokannat voidaan jaotella tietojärjestelmäriippumattomasti erilaisiin varastointityyppeihin esimerkiksi alla olevan kuvan mukaisesti.



Kaikelle yliopistossa syntyvälle tutkimusdatalle ei tällä hetkellä löydy sopivaa pitkäaikaissäilytyksestä huolehtivaa data-arkistoa. Data voidaan säilyttää asianmukaisesti myös ilman data-arkistoa, mutta se vaatii asiaan paneutumista. Tällainen tapaus on kuvattu portaiden alimmalla tasolla, joka kuvaa pitkäaikaissäilytyksen minimivaatimuksia. Ylemmät portaat voidaan laskea eräänlaisiksi data-arkistoiksi. Pitkälle kehitetyt, vakiintuneet, kansainväliset data-arkistot ovat usein portaiden yläosassa kuvatun mallin mukaisia. Dataa linkitetään mm. julkaisuihin, tutkimushankkeiden kuvauksiin tai koostetaan erilaisiksi avoimen datan sovelluksiksi.

Data-arkistot voidaan jakaa myös niihin, jossa aineisto on kuratoitua sekä niihin, joista kuratointiprosessit puuttuvat. Kuratoinnissa aineisto käy läpi tarkistusprosessin ja siihen lisätään annotointia, jolloin aineiston uudelleenkäytettävyys huomattavasti paranee. Kuratointiprosessi vaatii runsaasti resursseja ja sen vuoksi panostus on kannattavaa vain niihin data-aineistoihin, joiden uudelleenkäyttö on merkittävää.

5.3.2 Tietojärjestelmäkartoitus

Osana Helsingin yliopiston kirjaston tutkimusdataprojektia tehtiin kartoitus tietojärjestelmistä, jotka soveltuvat tutkimusdatan hallintaan tutkimushankkeen aikana, tutkimusdatan jakamiseen työryhmän kesken tai julkaisemiseen avoimen datan muodossa. Valintakriteereinä tarkasteltavien ohjelmistojen osalta oli seuraavat lähtökohdat:

- Skaalautuvuus suurillekin datamäärille
- Ohjelmisto saatavilla avoimena lähdekoodina
- Ohjelmistoa käytetään tutkimusdatan hallintaan tai sen ominaisuudet soveltuvat tähän käyttöön

Kartoituksessa koottiin lista mahdollisista tutkimusdatalle soveltuvista ohjelmistoalustoista osoitteeseen: <https://sites.google.com/site/tutkimusdata/>. Nämä jaoteltiin tietokantoihin, hakuliittymiin, paikkatietojärjestelmiin, pilvipalveluihin ja yleisesityksiin tästä aihepiiristä. Yleishavaintona voidaan todeta, että yliopistoissa ja tutkimuslaitoksissa on paljon aktiviteettia ja projekteja näihin asioihin liittyen, mutta useimmat hankkeet ovat vielä prototyyppi- tai pilottivaiheessa ja laajassa käytössä olevia järjestelmiä on vain kourallinen. Australia on hyvä esimerkki maasta, jossa tutkimusdatan hallintaan ja tallennukseen liittyviä hankkeita on jo toteutettu laajasti yliopistoissa.

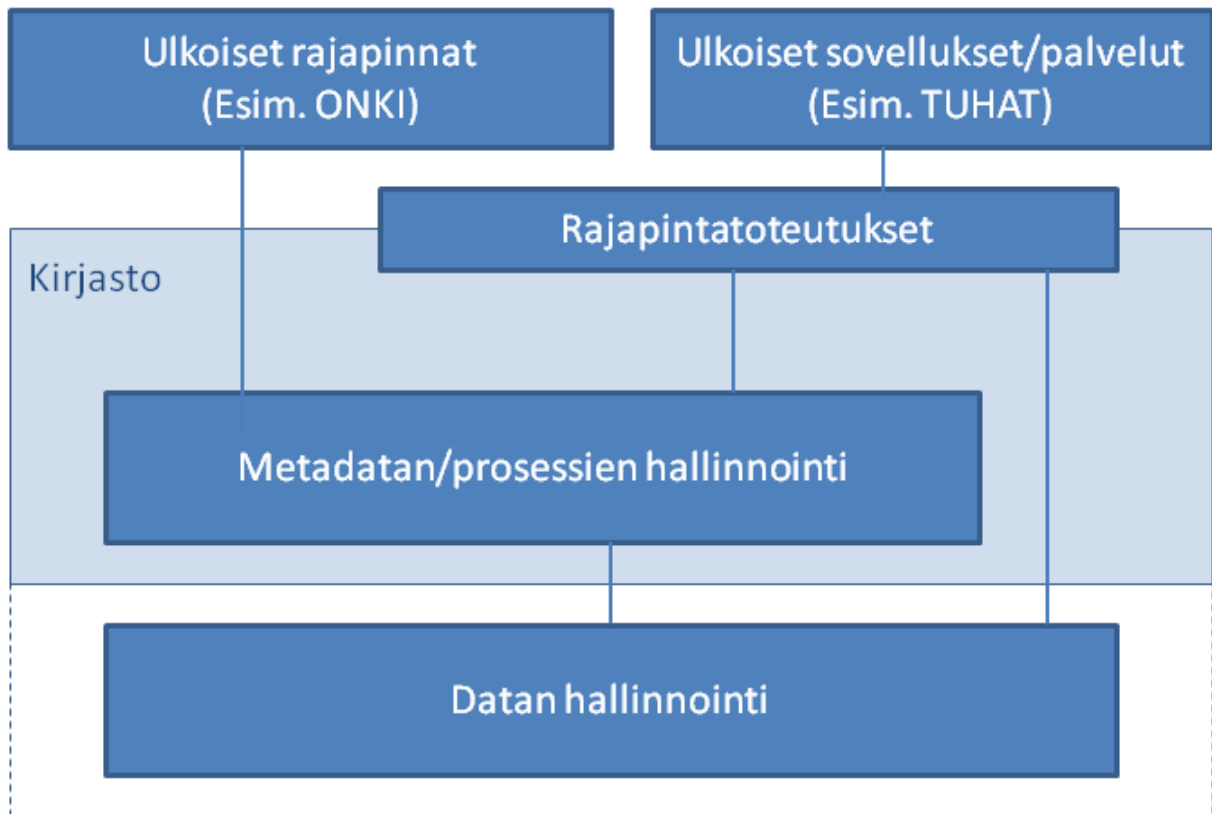
Tähän raporttiin on valittu lähempää tarkastelua varten muutamia ohjelmistoja. Valinnan kriteerinä on lähinnä arvio järjestelmän tarjoamista mahdollisuuksista ja sen taustalla toimivan kehittäjäyhteisön aktiivisuudesta. Tarkasteltavat ohjelmistot ovat:

- SciDB - Open Source Data Management and Analytics Software for Scientific Research
- KiWi - Open Source development platform for building Semantic Social Media Applications
- Semanttinen Mediawiki
- Terrastore
- OrientDB
- MongoDB
- SenseiDB
- CKAN - the open source data portal software
- DSpace
- hydra
- Archivematica
- iRODS

Näiden ohjelmistojen tarkempi tarkastelu löytyy liitteestä [Tietojärjestelmäkartoitus](#).

5.3.3 Tietojärjestelmien integraatioskenaarioita

Tutkimusaineistot ja niiden hyödyntämiseen liittyvät vaatimukset ovat monimuotoisia ja onkin hyvin epätodennäköistä, että yksi järjestelmä tai palvelu voisi tyydyttää kaikkien alojen tieteenekijöiden tarpeet. Todennäköisempi ratkaisu on integroitu järjestelmäympäristö, josta löytyy vaihtoehtoisia komponentteja esimerkiksi jättimäisten datamassojen, monimutkaisen metadatan sekä julkisen ja rajatun datan hallinnointiin.



Joonas Kesäniemi, Helsingin yliopiston kirjasto, verkkopalvelut

Kuvassa on hahmoteltu yksi mahdollinen korkean tason integraatiokuvaus tutkimusdatan hallinnointiin liittyvän toiminnallisuuden järjestämiseksi. Alimpana kerroksena on datan hallinnointi, jolla tarkoitetaan tässä tapauksessa varsinaiseen tutkimusdataan liittyen bittien tallentamista, muokkaamista ja jakelua, on kyse sitten eri tyyppistä tiedostoista tai tietokannoista. Esimerkkejä datan hallinnointiin liittyvistä konkreettisista toteutuksista ovat esimerkiksi CSC:n kehittämä Tutkimus-IDA, Amazonin pilvipalveluina tarjoamat S3 tiedostopalvelu ja SimpleDB-tietokanta sekä yliopiston tarjoama verkkolevy. Palvelut tarjoavat yleensä mahdollisuuden myös metadatan käsittelyyn, mutta pääpaino on geneerisen datan turvallisessa ja tehokkaassa hallinnoinnissa.

Jotta data olisi löydettävissä ja uudelleenkäytettävissä, tarvitaan sen päälle metadatakerros, jonka toteuttaminen voidaan eriyttää omaksi komponentiksi. Kyseisen komponentin on ajateltu tarjoavan pääasiallisen käyttöliittymän tutkimusdataan. Tämän takia mukana on mahdollinen prosessien tai työnkulkujen hallintaan liittyvä toiminnallisuus. Metadatahallinnointiin liittyvä palvelu voidaan toteuttaa esimerkiksi tietojärjestelmäkartoituksessa mainitun Semantic Media Wiki -ohjelmiston avulla, missä yhdistyy formaalin skeemapohjaisen tiedon ja vapaamman wikitekstin hallinnointi.

Pääsy tutkimusdataan ja sitä kuvaavaan metatietoon pitää pystyä määrittelemään tilanteesta riippuen julkiseksi tai rajoitetuksi. Esimerkiksi tutkimusprojektin aikana sekä data että metadata saatetaan haluta pitää vain tiettyjen käyttäjien saatavilla. Projektin päätyttyä dataan voidaan avata pääsy kaikille, ja aineisto voidaan julkaista metatiedon rajapintojen kautta haravoitavaksi esimerkiksi OpenAIREplus -tyyppisiin keskitettyihin hakupalveluihin. Riippumatta siitä, käytetäänkö tutkimusdatan hallinnointiin yhtä tai useampaa järjestelmää/palvelua, on tärkeää, että palvelussa voidaan hyödyntää sekä ulkopuolisten rajapintojen tarjoamaa dataa että omaa dataa standardien rajapintojen kautta ulkoisten sovellusten käyttöön.

Kirjaston roolin on tässä kuvauksessa katsottu keskittyvän nimenomaan metatiedon hallinnoinnin ympärille. Käytännössä tämä voi tarkoittaa esimerkiksi ontologioihin ja skeemoihin liittyviä asiantuntijatehtäviä, koulutusta tai opastusta aina palvelun tekniseen ylläpitoon ja kehitykseen.

5.3.4 Tietotekniikkakeskuksen hanke Tutkimusdatan säilyttämisen tekninen infrastruktuuri

Tietotekniikkakeskus toteutti marraskuusta 2011 maaliskuulle 2012 kestäneen hankkeen Tutkimusdatan säilyttämisen tekninen infrastruktuuri. Hankkeeseen perustettiin seurantaryhmä, jossa oli edustus myös Tutkimuksen toimialalta, Biotekniikan instituutista ja Helsingin yliopiston kirjastosta. Hankkeen tavoitteena oli selvittää tutkimusryhmien tarpeita datan tallennukseen ja valmistella tarvittavan tallennusinfrastruktuurin hankintaan liittyvää tarjouskilpailua. Hankkeessa todettiin, että tutkimusryhmien tarpeet vaihtelevat suuresti, eikä yksi malli riitä kaikkiin tarkoituksiin. Jatkossa tarvitaan halvempaa tallennuskapasiteettia, ja tallennustilan hankinta tulee tehdä entistä helpommaksi. Tallennukseen tarvitaan myös yhtenäisempää arkkitehtuuria ja parempia työvälineitä metadatan hallintaan. Valmisteltava puitesopimus voi sisältää myös mahdollisuuden pilvipalveluiden käyttöön datan tallennukseen. Pilvipalveluja yliopistolla on toistaiseksi käytetty lähinnä Microsoftin Azure-palvelun muodossa. Seuraavassa vaiheessa Tietotekniikkakeskus kilpailuttaa levykapasiteetin hankinnan puitesopimuksen, ja hanketta jatketaan vuoden loppuun projektissa nimeltä Tutkimusdatan talteenotto ja säilytys.

5.4 Tutkimusdatan metadata

Tutkimusdatan metadata on koottua, kuvailevaa tietoa tutkimuksen aineistosta, josta ilmenevät oleelliset tiedot aineiston ominaisuuksiin ja käyttöön liittyen. Asianmukainen metadata mahdollistaa aineiston löydettävyyden ja käytettävyyden tutkimuksen aikana, sekä sen jälkeen.

Tutkimusaineistoon liittyvä metadatan rakenne ja sisältö riippuvat aineistosta. Eri tieteenaloilla, ja tieteenalojen sisälläkin, käytetään hyvin erilaisia aineistoja, joten tarvittava metadata vaihtelee. [NISO](#) jakaa metadatan kolmeen tyyppiin: kuvaileva, rakenteellinen ja hallinnollinen metadata. Kuvailevaa metadataa käytetään kun etsitään ja paikannetaan kohdetta, kuten nimi, tekijä, aiheet, avainsanat, kustantaja. Rakenteellinen metatieto kuvaa miten kohteen komponentit järjestetään. Hallintaan liittyvällä metatiedolla tarkoitetaan teknisiä tietoja, mukaan lukien tiedostotyyppi. Hallintaan liittyvällä metatiedolla on vielä kaksi alatyyppeä: oikeuksien ja säilyttämisen metatiedot.

Metadata voi sisältää esimerkiksi seuraavia tietoja:

- aineiston koontimenetelmä
- aineiston formaatti
- ohjelmisto, jolla aineisto on luettavissa
- tekijät ja/tai tekijöiden pysyvät tunnisteet
- omistajat
- oikeudelliset asiat, kuten käyttöoikeudet
- tutkimuksen rahoitus
- kuvaus muuttujista, kuten käytetyistä ikäryhmistä
- datan pysyvät tunnisteet
- julkaisut, jotka on tuotettu aineiston pohjalta
- aineistoa kuvaavat asiasanat

5.4.1 Standardit ja ohjeet

Metadatan tuottamisessa käytetään olemassa olevia ja yleisesti hyväksyttäviä standardeja, jotta aineistojen siirto sekä yhteensopivuus muiden aineistojen kanssa onnistuu tarkoituksen mukaisesti.

Metadastandardeja on olemassa useita satoja. Jenn Riley'n (Indiana University Library) tekemässä koosteessa on esitelty tärkeimmät julkaisu-, kirjasto-, arkisto- ja museotalalla käytettävistä metadastandardit posterissa, jossa standardit jaotellaan käyttökohteiden mukaan. (Riley, 2009) Osa tieteellisten tutkimusaineistojen metadastandardeista on samoja kuin julkaisu-, kirjasto- ja arkisto- ja museotalalla, mutta tämän lisäksi on käytössä erityisiä tutkimusaineistoille tarkoitettuja metadastandardeja. Erityisesti kansainvälisissä tutkimusdata-arkistoissa käytetään standardien mukaista metadataa. Euroopassa OpenAIREplus projekti pyrkii määrittelemään aineistoille metadatomallin, jota EU:n puiteohjelmien (FP7) rahastoista saaneet hankkeet noudattavat.

Kansallinen digitaalinen kirjasto (KDK) on hanke, jonka tarkoituksena on varmistaa kulttuurin ja tieteen sähköisten aineistojen laadukas hallinta, palvelujen esteetön saatavuus ja luotettava pitkäaikaissäilytys. Hankkeessa on suunniteltu metadatumallia, jota edellytetään pitkäaikaissäilytykseen (PAS) otettavalle aineistolle. PAS-metadatumallia on hyvin yleistä. Tutkimusdatan uudelleen käytettävyys vaatii aineistoista riippuen aiheen mukaista erityistä metadatumallia. PAS-järjestelmä taas tulee edellyttämään arkistoinnin kannalta tarpeellista yhdenmukaista metadatumallia.

Tutkimusaineistojen laadukas metadatumallia takaa varsinaisen datan ymmärrettävyyden, käytettävyyden, ylläpidettävyyden tutkimuksen aikana ja osaltaan mahdollistaa tutkimuksen toistettavuuden tutkimuksen jälkeen. Monissa tutkimushankkeissa on ongelmana hallita suuria tutkimusdatamääriä. Erityisesti metadatumallia ei aina ole riittävästi. Pilottitutkimuksen perusteella voidaan sanoa, että tutkijat tarvitsevat tukea erityisesti metadatumallin suunnitteluun, esimerkiksi muuttujien kuvaamiseen ja sopivien metadatumallien valitsemiseen. Myös toimivien toimintaprosessien kehittämiseen tarvitaan tukea, esimerkiksi ohjeistuksen osalta. Tiedostojen informatiivinen ja yhtenäinen nimeäminen tutkimusryhmän sisällä koettiin ongelmalliseksi. Tiedostolistauksen tekemiseen ja tiedostojen sisällön tallentamiseen toivottiin neuvoja. Tiedostojen eri versioidenkin dokumentointi voisi auttaa tutkijaa löytämään aiemmin tehdyn työn.

Pilottihankkeen aineistoille pyrittiin löytämään sopivia metadatumallia tutkittujen hankkeiden aineistoille. Tämä vaatii aiheeseen paneutumista ja kiinteää yhteistyötä tutkijoiden kanssa. Metadatumallin tuottaminen manuaalisesti saattaa olla työlästä, joten sitä on pystyttävä myös hyödyntämään heti tutkimuksen aikana. Monet tutkimuslaitteistot ja -ohjelmat tuottavat automaattisesti korkeatasoista metadatumallia ja tämän täysipainoista käyttöä olisikin syytä edistää.

5.4.2 Tutkimusdatan pysyvät tunnisteet

Yksi tärkeä osa metadatumallia ovat datan liittyvät pysyvät tunnisteet. Juha Hakalan mm. Tietolinjalla julkaistussa [katsauksessa esitellään käytössä olevia pysyviä tunnistejärjestelmiä](#) ja Esa-Pekka Keskitalo esittelee artikkelissaan ["PersID - eurooppalaista yhteistyötä pysyvien tunnusten parissa"](#) aihetta tutkimusdatan pitkäaikaissäilytyksen näkökulmasta. Pysyvät tunnisteet auttavat myös tutkimusdatalle tyypillisessä kohtu nopeassa muutostahdissa, auttaen versioinnin hallinnassa (Buneman, et al. 2004).

Pysyvistä tunnisteista on hyötyä ([ANDS:n](#) mukaan):

- Edistää tutkimusdatan hyväksymistä aineistoksi, johon voidaan tieteessä viitata ja sitä kautta meritoitua.
- Tukee arkistointia, mikä mahdollistaa sen, että tuloksia voidaan tarvittaessa korjata ja käyttää uudelleen tulevaisuuden tutkimuksessa.
- Mahdollistaa aineiston viittausanalyysien tuottamisen ja aineiston käytön seuraamisen.

Seuraavassa kappaleessa on käsitelty tarkemmin julkaisujen linkitys tutkimusdataan.

5.5 Julkaisujen linkitys tutkimusdataan

Tiede rakentaa nykyisen ja tulevan tutkimustyön aiemmin julkaistun tutkimuksen varaan. Tutkijan on voitava seurata todistusketjuja, jotka ulottuvat hyvinkin kauas. Tutkimuksen tulokset ja johtopäätökset julkaistaan tieteellisinä artikkeleina, monografioina, tutkimusraportteina jne. Nämä indeksoidaan kirjastojen toimesta viitetietokantoihin, joista tieteenalojen tai organisaatioiden tärkeimmät tutkimusjulkaisut ovat yleensä helposti paikannettavissa. Sen sijaan tutkimuksessa käytetyille datalle ei välttämättä ole luontevia hakukanavia. Eräs sellainen voi olla linkitys tutkimusjulkaisuihin.

Vertaisarvioituista julkaisusta tulisi käydä ilmi mitä dataa on käytetty, uutta tai jo aiemmin koottua, niin että tutkimustulokset voidaan muiden toimesta arvioida tai jopa toisintaa. Harvat tutkijat kuitenkin liittävätkin artikkeleihinsa suoria datalinkkejä. Data ei ehkä ole lainkaan julkisesti saatavilla tai linkitettävissä, tai siihen ei voi muodostaa pysyvää linkkiä (DOI, PURL, URI, tms.) Vakiintuneita viittauskäytäntöjä on toistaiseksi vain joillakin aloilla. Esimerkiksi Yhteiskuntatieteellinen tietoaarkisto (FSD) ohjeistaa arkistoon talletettuun dataan viittaamisesta: <http://www.fsd.uta.fi/aineistot/jatkokaytto/viittaaminen.html> Lisäksi FSD:n aineistoluettelossa on mainittu ne tutkimukset, joissa kyseistä aineistoa on käytetty. Aineistot on myös luetteloitu LINDA-tietokantaan, niin että tutkija lähdeaineistoa hakiessaan löytäisi myös aiheeseen liittyvää dataa.

Esimerkkinä luonnontieteiden käytännöstä on tähtitiede, jossa data yleensä sijaitsee kansainvälisissä palvelimilla. Artikkeleissa käytetään erilaisia objektien ja havainto-ohjelmien tunnisteita, joita voidaan skannata ohjelmallisesti (esim.

DJIN) artikkeleiden kokoteksteistä. Osumien perusteella datakeskus voi lisätä artikkeliviitetietokantoihin suoria datalinkkejä. NASA:n ylläpitämässä [Astrophysics Data System](#) -tietokannassa haku voidaan rajata viitteisiin, joissa on ko. datalinkitys. Tämän havainnollistamiseksi tehtiin vuosina 2005-2010 ilmestyneille julkaisuille affiliaatiohaku **helsinki**, jonka tuloksena löytyneissä 1217 viitteestä kaikkiaan 112 viitteessä oli linkitys artikkelissa käytettyyn dataan. Monet tieteelliset kustantajat tallentavat palvelimilleen artikkeleissa käytettyä dataa, mutta eivät toistaiseksi tarjoa mahdollisuutta etsiä dataa omien hakuliittymiensä kautta.

Helsingin yliopiston tutkijat tallentavat julkaisutietonsa vuonna 2010 käyttöön otettuun TUHAT-tutkimustietokantaan, jossa ei kuitenkaan ole mahdollista tallentaa erikseen tunnistettavissa olevia datalinkkejä muualle kuin vapaalinkkikenttään. Datan tallentamisesta ei myöskään ole toistaiseksi saanut mitään meriittejä esim. arvioinneissa. Mikäli tutkimusdatan ilmoittamisesta meritoituisi, myös datalinkitykset julkaisuihin ja datalinkkien kirjaaminen tutkimustietokantoihin lisääntyisivät.

- **TUHATissa tulisi olla nykyistä näkyvämpi mahdollisuus ilmoittaa datan sijainti, sillä tämä nostaisi datan selkeämmin näkyviin ja tutkijoiden tietoisuuteen.**

5.6 Pikakysely ilmatieteen tutkijoille

Fysiikan laitoksen ilmakehätieteiden osasto tuottaa runsaasti mm. pienhiukkasiin liittyvää tutkimusdataa. Osaston tutkijoille tehtiin marraskuussa 2011 osaston postituslistan kautta suppea kysely jossa heiltä tiedusteltiin: "Onko tutkimusdatasi linkitetty julkaisuihisi? Missä linkittämäsi data sijaitsee?" Kyselyyn tuli vajaa kymmenkunta vastausta. Vain yksi tutkija ilmoitti julkaisustaan löytyvästä datalinkistä: "Kyllä, laitoksen palvelimella jonne linkki julkaisussa: <http://www.atm.helsinki.fi/eusaar>". Muut vastaukset jakaantuivat seuraavan tyylisiin: "ei ole", "ei toistaiseksi", "ei, mutta suunnitelmassa on ollut".

5.7 Osaamisen kehittäminen

5.7.1 Kirjastot tarvitsevat datanhallinnan osaajia

Helsingin yliopiston strategian mukaisesti yliopiston kirjaston tarjoamia tutkimuksen tukipalveluita / tutkimuspalveluita vahvistetaan. Mikäli niihin sisällytetään tutkimusdatan hallinnan tukea tai siihen liittyviä asioita, on erityisesti tietoasiantuntijoiden osaamista näissä tehtävissä kehitettävä.

Tällä hetkellä datanhallintaan liittyvää täydennyskoulutusta kirjastoammattilaisille ei ole Suomessa tarjolla. Helsingin yliopiston kirjaston henkilökunta on osallistunut aihepiiriin liittyviin seminaareihin, joita on ollut runsaasti tarjolla. Kirjastossa on käynnistetty dataopintopiiri, jossa perehdytään monipuolisesti tutkimusdatan hallintaan: sen tallentamiseen, säilyttämiseen, linkittämiseen, kuvailuun ja jakamiseen.

Esitys: Kirjaston henkilöstön osaamista kehitetään niin, että kullakin kampuksella on vähintään yksi tietoasiantuntija, joka pystyy tukemaan tutkijoita tutkimusdatan hallinnassa.

Esitys: Tavoitteena on, että kirjastolla on vähintään yksi syvällisesti tutkimusdatan hallintaan perehtynyt, tutkimusmenetelmiä ja dataan liittyviä oikeudellisia tekijöitä tunteva tietoasiantuntija, joka on lisäksi tietoteknisesti suuntautunut (data librarian).

Esitys: Tutkimusdatan hallintaa koskeva yleinen osaaminen ja tietämys kuuluvat jokaisen tutkimuksen tukipalveluissa työskentelevän osaamisalueelle. Asiakaspalvelussa tunnetaan dataan liittyvät tiedonlähteet ja tahot, joihin tiedon tarvitsija voi ottaa yhteyttä.

5.7.2 Datanhallinnan taidot opiskelijalle ja tutkijalle

Tutkimusdataprojektissa pohdittiin, missä vaiheessa opiskelijan ja tutkijan olisi perehdyttävä tutkimusdatan käsittelyyn ja hallintaan. Kysyttiin myös, mille taholle tutkimusdataan liittyvä opetus kuuluu - yliopiston opettajille, tutkimusryhmien vastuututkijoille, Tietotekniikkakeskukselle vai mahdollisesti kirjaston tiedonhallinnan opettajille.

Tutkijankoulutuksessa opetetaan tutkimusjulkaisujen laatimiseen liittyviä taitoja, mutta tutkimusdatan julkaisemisen opetus jää siinä yhteydessä paljolti käsittelemättä. Edes väitöskirjaohjeissa ei kerrota, mitä datalle tulisi tehdä. Yhteiskuntatieteellisestä tietoarkistosta tosin löytyy gradujenkin dataa. Onkin yleistä, ettei väitöskirjojen dataa löydy eikä siihen linkitetä. [Bennett \(2009\)](#) esitti vertailun genetiikan, solubiologian, psykologian, biokemian ja sosiologian väitöskirjojen tutkimusdatasta. Hänen selvityksensä mukaan tutkimusdataa oli useimmissa tapauksissa mahdotonta löytää väitöskirjassa esitetyn tiedon perusteella, tai viittaukset tutkimusdataan olivat epäselviä. Helsingin yliopistossa kannattaisi harkita tutkimusdatan käsittelyn ja linkittämisen opettamista tutkijankoulutuksessa. Esimerkkinä [Cambresyn ym. \(2011\)](#) artikkeli, jossa kerrotaan, miten nuorille tutkijoille opetetaan datan linkittämistä tekstiin.

Tutkimusmenetelmät, tutkimusaineiston analysointi ja käsittely kuuluvat laitoksella opetettaviin asioihin. Osa tutkimusdatan hallintaan liittyvistä asioista on sellaisia, että ne voidaan sisällyttää jatko-opinto- ja maisterivaiheen tiedonhankinnan kursseihin. Kirjaston kurssien sisällöt pyritään suunnittelemaan yhteistyössä laitosten opetushenkilökunnan kanssa, jolloin voidaan tarkentaa vastuualueet. Kirjaston tarjoaman opetuksen laajuus riippuu siitä, kuinka paljon henkilökuntaa on käytettävissä opetukseen ja miten hyvin tiedonhankinnan opettajat ovat pätevöityneet tutkimusdatan hallintaan.

Esitys: Dataosaamisen tulisi kertyä viimeistään jatko-opintovaiheessa.

Esitys: Tutkimusdataan liittyvää opetusta suunniteltaessa yhteistyötahoina olisivat Tietotekniikkakeskus, verkko-opetuksen tukihenkilöt, Opetusteknologiakeskus, tutkimusryhmien vastuulliset tutkijat ja tutkimusmenetelmäkursseiden opettajat. Myös CSC:n rooli opetuksen antajana tulisi selvittää, sekä muut tutkimusdatapalveluja tarjoavat tahot, kuten Biocenter Finland.

5.7.3 Opetusmateriaalin tuottaminen

Kirjaston Solmu-hankkeessa tuotettuun ja tutkimusdataprojektissa ylläpidettyyn sivustoon ([Opas tutkimusaineistojen tiedonhallintaan](#)) on koottu datanhallintaa koskevaa tietoa ja tutkimusdataan liittyviä tiedonlähteitä. Wiki-sivujen ylläpito on työlästä ja vaatii jatkuvaa huomiota. On todennäköistä, että tutkijat käyttävät erillisiä sivustoja vähän. Niiden linkittäminen tai vielä mieluummin sivuston sisällön liittäminen esim. Helsingin yliopiston hallinnon ylläpitämään [Yleiseen tutkimusprojektien oppaaseen](#) voisi olla järkevää. Toisaalta avoimesti verkossa olevia oppaat ovat tärkeitä. Tietotekniikkakeskuksen avoimesti saatavaa oppimateriaalia voivat niin opiskelijat kuin tutkijatkin hyödyntää (esim. [TVT-ajokortti oppimateriaali](#)). Yhteiskuntatieteellisen tietoarkiston tuottama [Tutkimusaineistojen tiedonhallinnan käsikirja](#) on kattava tietopaketti, joka on verkossa avoimesti käytettävissä. Opas keskittyy lähinnä yhteiskuntatieteellisiin tutkimusaineistoihin.

Verkkoon tuotetun neuvonta- ja opetusmateriaalin tulisi olla nopeasti hyödynnettävää - tutkijoiden aika ei riitä pitkiin ja syväisiin artikkeleihin perehtymiseen. Datanhallintapalvelut tulee muodostaa siten, että tieto on sen tarvitsijoiden saatavilla nopeasti ja täsmällisesti. Muualla maailmalla tuotettua, olemassa olevaa opetusmateriaalia kannattaa käyttää myös hyödyksi. Jos sivut rakennettaisiin semanttisen webin mallin mukaan, koko sisältö olisi ontologisesti järjestettyä (vrt. [Järviwiki](#)) ja tieto löytyisi loogisten linkityksien avulla vaivattomasti.

Neuvontaa voidaan visualisoida ja havainnollistaa myös sivuille sijoitettujen videoklippien avulla. Näitä löytyy joidenkin data-arkistojen yhteydestä esim. Iso-Britanniassa.

Elektronisen oppimateriaalin lisäksi voidaan tarvita painettuja esitteitä. Niistäkin löytyy kansainvälisiä esimerkkejä.

Esitys: Tutkijoiden saatavilla pitäisi olla tietoa, opastusta ja koulutusmateriaalia tutkimusdatan käsittelyyn ja hallintaan. Näiden neuvontapalveluiden tulisi olla samassa paikassa kuin muutkin tutkimuksen tukipalvelut tai tutkijapalvelut. Tutkijapalveluiden järjestäminen helposti löydettäväksi ja lähestyttäväksi vaatii paljon resursseja, laajaa osaamisen kehittämistä ja eri toimintojen koordinoitua palveluntarjoajan puolella. Tähän panostaminen olisi kuitenkin kannattavaa, jotta tutkijat saisivat tiedon tehokkaasti käyttöönsä. Tiedon kokoamiseen ja esittämiseen tarvitaan helposti ylläpidettävä julkaisualusta.

Esimerkkejä opetusmateriaalista muualla:

- [Managing and sharing data](#) (UK Data Archive)

- [Datatrain](#) (Cambridge University Library)

5.8 Ehdotus kirjaston rooliksi ja eri toimijoiden tehtäviksi

Vaikka tutkimuskäytännöt eri tieteenaloilla vaihtelevatkin suuresti, tutkimuksen tietoympäristö on globaali. Työryhmä on halunnut listata oheiseen taulukkoon laajasti eri tahoja, jotka eri tavoin liittyvät datan hallintaan. Toimijaluettelo ei ole kattava, eikä toimijoiden tehtäviä ole yritetty listata täydellisesti. Kansallinen tehtäväjako on selkiytymässä ja tulee siten lähitulevaisuudessa vaikuttamaan osaltaan pilotissa havaittujen ongelmien ratkaisemiseen. Tietotekniikkakeskuksen aktiivinen rooli teknisen infrastruktuurin monien eri vaihtoehtojen rakentajana tulee käsityksemme mukaan olemaan keskeinen Helsingin yliopistossa. Kirjaston olisi hyödyllistä toimia Tietotekniikkakeskuksen pysyvänä yhteistyökumppanina tutkimusdata-asioissa. Yliopiston hallinnon, tutkimuksen toimialan ja kirjaston keskinäinen työnjaon läpikäynti vaatii jatkotoimenpiteitä.

5.8.1 Datatyöryhmän ehdotus eri toimijoiden rooleista tutkimusdatan ja -aineistojen hallintaan liittyvissä tehtävissä

		Tehtävät tai roolit
Kansainvälinen toimija		
	EU (komissio, European Research Council)	EU:n datapolitiikka
	Tutkimuksen rahoittajat	Voivat asettaa vaatimuksia datan hallinnoinnille tai saatavuudelle
	Kustantajat	Voivat tarjota omia alustoja tutkimusdatalle, voivat linkittää julkaisuihin datan
	Tieteenalakohtaiset datapankit ja -arkistot	Kokoavat, järjestävät ja tarjoavat käyttöön oman aihealueensa datan
	Monialaiset avoimet arkistot	Tarjoavat datanhallinta- ja säilytyspalveluja
	Yritykset	Tarjoavat infrastruktuuripalveluja ja datan hallinnointiin liittyvää konsultointia
Kansallinen toimija		
	Opetus- ja kulttuuriministeriö	Kansallinen datapolitiikka yhdessä muiden ministeriöiden kanssa.
	CSC Tieteen tietotekniikan keskus	Kansallinen koordinointi, datakatalogi, IDA-tallennuspalvelu, koulutus, datan pitkäaikaistallennus Kansalliskirjaston kanssa, kansainvälinen yhteistyö (mm. EUDAT)
	Yhteiskuntatieteellinen tietoarkisto FSD	Yhteiskuntatieteellisten ja humanististen tieteiden tutkimusaineistojen koordinointi, arkistointi ja välitys, kansainvälinen yhteistyö
	UNIFI ja yliopistot yhteisesti	Yliopistojen yhteiset kannanotot ja hyvien käytäntöjen jakaminen

	Tutkimuslaitokset ja virastot	Tietovarantojen avaaminen ja tietojärjestelmien rajapinnat
	Kansalliskirjasto	Standardisalkku, Kansallinen digitaalinen kirjasto ja PAS (pitkäaikaissäilytys)
Helsingin yliopisto		
	Yliopiston johto Tutkimuksen toimiala	Politiikat ja strategiat yliopiston tasolla Hyvän datapolitiikan luominen ja ylläpitäminen Sopimukset, rahoitusmekanismit, projektinhallinta, tutkimushallinnon portaali
	Tietotekniikkakeskus	Tallennuksen tekninen infrastruktuuri ja sen kehittäminen
	Opetusteknologiakeskus	Koulutus datanhallinnan työkalujen käyttöön
	Helsingin yliopiston kirjasto	Datan hallintapalvelujen integrointi, koordinointi ja tuki. Metadatan ja järjestelmien käyttöliittymät (esim. haettavuus, linkitykset), tietty osa koulutuksesta, viestintä, neuvonta, opetusaineisto
	Tiedekunnat	Tiedekunnan linjaukset, resurssointi, päätökset pitkäaikaissäilytettävistä aineistoista
	Laitokset	Datanhallinnan organisointivastuu, pitkäaikaissäilytettävien aineistojen valinta
	Opettajat	Datanhallintaan liittyvä koulutus osaksi muuta jatko-opiskelijoiden ja tohtorikoulutettavien koulutusta
	Tutkimusryhmät	Hyvän datanhallinnan käytäntöjen noudattaminen, päätös datan käyttöoikeuksista projektin aikana ja sen päätyttyä (koskee tiettyjä tieteenaloja)
	Vastuututkijat	Vastuu hyvän datanhallinnan käytäntöjen noudattamisesta, tutkimusryhmien jäsenten koulutus
	Tutkijat	Hyvän datanhallinnan käytäntöjen noudattaminen ja datan metadatan tuottaminen, päätös datan käyttöoikeuksista (koskee lähinnä humanistisia ja yhteiskuntatieteitä)

5.8.2 Kirjaston rooli tutkijan tukena tutkimusdatan hallinnassa

Euroopan tutkimuskirjastojen liitto LIBER keskittyy kirjastoalan järjestönä nimenomaan kirjastoammattilaisten asemaan. Katsauksessaan (Report on Integration of Data & Publications, 2011) todetaan, että kirjastolla on tärkeä rooli tiedonhallinnan kokonaisuuden hahmottamisessa. <http://www.libereurope.eu/news/report-on-integration-of-data-publications>

Eri puolilla maailmaa on viime aikoina etsitty voimakkaasti malleja, joissa kirjastotyöntekijät muodostavat uudenlaisia kumppanuuksia tutkimusryhmien kanssa vaativien tiedonhallintahaasteiden hallitsemiseksi. Tiedonhallinnan uudet haasteet jäävät helposti tutkijoilta ratkomatta, sillä tutkijoiden omat voimavarat eivät useinkaan riitä niiden pohtimiseen. Uusiksi tiedonhallintahaasteiksi määritettiin Helsingin yliopiston kirjastossa toteutetussa Solmu-projektissa (2009-2011)

tutkimusaineistojen ja tutkimuksen työaineistojen tallentaminen, järjestäminen, hallinta ja jatkokäyttöön saattaminen, tieteellisen julkaisemisen uudelleen hahmottaminen, avoin julkaisutoiminta ja aineistonhallintasuunnitelma.

Kirjaston tärkeä tehtävä tutkimusaineistojen tiedonhallinnassa on sama kuin julkaisujen kohdalla: edistää tutkimustiedon (tässä siis tutkimusdatan) löydettävyyttä, saatavuutta, käytettävyyttä ja vaikuttavuutta. Kirjasto on neutraali tieteen toimija kampuksella ja siten tämä tehtävä ja rooli sopii kirjastolle hyvin.

5.8.3 Datatyöryhmän ehdotus Helsingin yliopiston kirjaston roolista

- yhteistyön ja eri toimijoiden välisen kanssakäymisen aktiivinen edistäminen
- järjestelmien ja käyttöliittymien kehittäminen yhdessä muiden toimijoiden kanssa
- datanhallinnan tuen tarjoaminen
- osallistuminen erilaisiin työryhmiin kehittäjänä ja asiantuntijana
- jatko-opiskelijoiden ja tutkijoiden datanhallinnan koulutuksen järjestäminen ja edistäminen
- datanhallinnan ja julkaisemisen kehityksen aktiivinen seuraaminen ja viestintä
- kirjaston oman osaamisen kasvattaminen (datanhallinta, metadata, avoin julkaiseminen)

6 Lähdeluettelo

- Australian National Data Service, www-sivut http://ands.org.au/cite-data/resources.html#Data_Citation_Benefits (Viitattu 22.2.2012)
- Bennett, T., Herndon, J. Nicholson, S. & O'Reilly, R. 2009. Welcome to C4: Data Sharing Across the Disciplines. pp-esitys. http://www.fsd.uta.fi/iassist2009/presentations/C4_Nicholson_Bennett.ppt
- Borg, S. ja Kuula, A. 2007. Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa. Yhteiskuntatieteellisen tietoarkiston julkaisuja 6. http://www.fsd.uta.fi/julkaisut/julkaisusarja/FSDjs06_OECD.pdf
- Buneman, P., Khanna, S., Tajima, K. & Tan, W. 2004. Archiving scientific data, ACM Transactions on Database Systems 29(1), s. 2-42.
- Cambresy, L. et al (2011). From your paper to VizieR and SIMBAD. Teoksessa C. Sterken (ed.): Scientific writing for young astronomers. Part I. EAS Publication Series 49 (2011). s. 135-157. [DOI:10.1051/eas/1149008](https://doi.org/10.1051/eas/1149008)
- Committee on Science, E., National Academy of Sciences (U.S.), National Academy of Engineering., & Institute of Medicine (U.S.). (2009). Ensuring the integrity, accessibility, and stewardship of research data in the digital age. 162 s.
- DataTrain. [aineistonhallinta:verkkajulkaisu]. Cambridge University Library, Cambridge University [aineistonhallinta:ylläpitäjä ja tuottaja]. <<http://www.lib.cam.ac.uk/dataman/datatrain/datatrainintro.html>> (Viitattu 09.02.2012.)
- Engeström, Y., Laitio, A. ja Mylly, H. 2010. Helsingin yliopiston kirjaston Solmu-hanke. Väliraportti. http://www.helsinki.fi/kirjasto/julkaisut/Solmu_Valiraportti_Viikki.pdf
- Engeström, Y., Kaatrakoski, H., Laitinen, A. ja Rantavuori, J. 2011. Solmutyöskentely kirjastossa. Raportti Helsingin yliopiston keskustakampuksen kirjaston ja pilottitutkimusryhmien kehittämishankkeesta. Toiminnan kehityksen ja oppimisen tutkimusyksikkö, Helsingin yliopisto. Työpapereita 8/2011. http://www.helsinki.fi/kirjasto/julkaisut/Solmuty%C3%B6skentely_kirjastossa%20_julkaisu.pdf
- Forsman, M. ja Karhula, P. 2010. Tutkimusaineistot haasteena yliopiston kirjastolle. s. 129-137. Teoksessa: Rajapinnassa: uusi Helsingin yliopiston kirjasto. Sinikara, K., Forsman, M., Karppinen, I. ja Lammi, P. (toim.). <http://hdl.handle.net/10138/17102>
- Hakala, J. 2010. Yleiskatsaus pysyviin tunnisteisiin. Tietolinja 2/2010. <http://urn.fi/URN:NBN:fi-fe201012143109>
- Hallinnolliset ja rakenteelliset metatiedot ja aineiston paketointi v 1.0 28.10.2011. Kansallinen digitaalinen kirjasto. http://www.kdk.fi/images/stories/tiedostot/kdk_metatiedot_ja_aineiston_paketointi.pdf
- Helsingin yliopiston datapolitiikka. Tutkimusdatatyöryhmän ehdotukset. Toukokuu 2011. <http://hdl.handle.net/10138/28457>.
- Hyvä tieteellinen käytäntö ja sen loukkausten käsitteleminen. 2002. Tutkimuseettinen neuvottelukunta. http://www.tenk.fi/hyva_tieteellinen_kaytanto/htkfi.pdf
- Keskitalo, E-P. 2011. PersID - eurooppalaista yhteistyötä pysyvien tunnusten parissa Tietolinja 1/2011. <http://urn.fi/URN:NBN:fi-fe201106081712>
- Koponen, T. Lehto, A., Luukkanen, Nevalainen E. ja Sassali J. 2010. Miten kirjastot voivat tukea tutkimuksen tiedonhallintaa? Signum 1:2010. s. 23-27.
- Library Roles in University Research Assessment. 2010. Liber quarterly, Volume 20(2). <http://liber.library.uu.nl/publish/issues/2011-1/index.html?000522>

- Opas tutkimusaineistojen tiedonhallintaan [aineistohallinta:verkkojulkaisu]. Helsingin yliopiston kirjasto [aineistohallinta:ylläpitäjä ja tuottaja]. <<https://wiki.helsinki.fi/display/aineistohallinta/Etusivu/>> (Viitattu 09.02.2012.)
- PAS-järjestelmän tekninen ja hallinnollinen metadata v. 2.5. 2010.Kansallinen digitaalinen kirjasto. http://www.kdk.fi/images/stories/PAS-jarjestelman_tekninen_ja_hallinnollinen_metadata_v2.5.pdf
- Paskin, N. 2005, Digital Object Identifiers for scientific data. Data Science Journal 4, s. 12-20.
- Peltonen, E. 2009. Tutkimusaineistojen tiedonhallinta- ja jatkokäyttökysymykset kiinnostavat tutkijoita – ja kirjastoammattilaisia. Verkkaari 7/2009.
- Pryor, Graham (ed.). 2012. Managing Research Data. 239 s.
- Rautio, A. & Korkka-Niemi, K. 2011: Characterization of groundwater -lake water interactions at Pyhäjärvi, a lake in SW Finland. Boreal Environment Research 16: 363-380. <http://www.borenv.net/BER/pdfs/ber16/ber16-363.pdf>
- Riley, J. 2009-2010, Seeing Standards: A Visualization of the Metadata Universe. <http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards.pdf>.
- Ristikartano, V. 2010.Tutkimusjulkaisujen saatavuus ja tutkimusaineistojen jatkokäyttö keskiössä Tieteellisten aineistojen verkkopalveluhankkeessa 2007-2009. Verkkaari 1/2010. <http://blogs.helsinki.fi/verkkari-lehti/tutkimusjulkaisujen-saatavuus-ja-tutkimusaineistojen-jatkokaytto-keskiossa-tieteellisten-aineistojen-verkkopalveluhankkeessa-2007-2009>
- Ristikartano, V. ja Forsman, M. 2009. Mikä voisi olla tiedekirjaston rooli tutkimusdatan arkistoinnin edistämässä? Verkkaari 5/2009. <http://blogs.helsinki.fi/verkkari-lehti/mika-voisi-olla-tiedekirjaston-rooli-tutkimusdatan-arkistoinnin-edistamisessa>
- Sassi, Raija. 2009. Helsingin yliopiston elnfrastruktuuri 2010--2012. Helsingin yliopiston hallinnon julkaisuja 63; Raportit ja selvitykset. <http://hdl.handle.net/10138/17496>
- Tieto käyttöön. Tiekartta tutkimuksen sähköisten tietoaineistojen hyödyntämiseksi. 2011. Opetus- ja kulttuuriministeriön julkaisuja 2011:4. http://www.minedu.fi/OPM/Julkaisut/2011/Tiekartta_tutkimuksen_sahkoisten_tietoaineistojen_hyodyntamiseksi.html
- Tutkimusaineistojen tiedonhallinnan käsikirja [aineistohallinta:verkkojulkaisu]. Tampere : Yhteiskuntatieteellinen tietoarkisto [aineistohallinta:ylläpitäjä ja tuottaja]. <<http://www.fsd.uta.fi/tiedonhallinta/>>. (Viitattu 09.02.2012.)
- TVT-ajokortti. [aineistohallinta:verkkojulkaisu]. Helsingin yliopisto, tietotekniikkaosasto, kirjasto [aineistohallinta:ylläpitäjä ja tuottaja]. <<http://www.helsinki.fi/tvt-ajokortti/materiaali.htm>> (Viitattu 09.02.2012.)
- Understanding Metadata. 2004. NISO. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
- Van den Eynden, V., Corti, L., Woollard, M, Bishop, L. & Horton, L. 2011. Managing and Sharing Data. UK Data Archive. Best practice for researchers. <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

7 Liite Tietojärjestelmäkartoitus

7.1 SciDB - Open Source Data Management and Analytics Software for Scientific Research

Vuonna 2008 käynnistetyssä kunnianhimoisessa SciDB-hankkeessa on tavoite paljon korkeammalla kuin pelkän tallennusalueen kehittämisessä tutkimusdatalle. Taustalla on tutkimustyötä erittäin suurten tietokantojen parissa, esimerkiksi tähtitieteellisten teleskooppien tuottamilla data-aineistoilla. Yksi hankkeen pääkehittäjistä on MIT:ssä vaikuttava Mike Stonebraker, joka on taustaltaan pitkän linjan tietokantaguru.

Perinteisen tietokantatauluihin perustavan relaatiotietokantamallin sijaan SciDB-hankkeessa pyritään rakentamaan uudenlainen matriiseihin (array) perustuva tietokanta. Tiedon tallennus perustuu versiointiin eikä päivittämiseen, ja tiedon alkuperän säilyttämistä tuetaan (proveniensi), ohjelmisto tukee erilaisia työnkuluja, ja se voi käsitellä myös epävarmaa tietoa. Ohjelmisto skaalautuu satojen petatavujen datamäärille ja tuhansien rinnakkaisten palvelinsolmujen käyttöön.

Tärkeä lähtökohta kehitykselle on, että ohjelmisto julkaistaan avoimena lähdekoodina, jolloin sen ympärille voi syntyä kehittäjäyhteisö, joka varmistaa, että datavarannot eivät jää toimittajakohtaisten suljettujen järjestelmien lukitsemiksi. Datan hallinnan lisäksi ohjelmisto sisältää monipuoliset välineet aineiston tilastolliseen analyysiin ja visualisointiin. Ohjelmiston kehitystyötä rahoittaa kaksi amerikkalaista ja kaksi venäläistä tutkimusorganisaatiota sekä kaksi yritystä. SciDB-ohjelmisto on vasta rakenteilla, mutta sen etenemistä kannattaa seurata, sillä siitä voi jatkossa muodostua tärkeä alusta tutkimusdatan hallintaan ja analysointiin.

<http://www.scidb.org/>

7.2 KiWi - Open Source development platform for building Semantic Social Media Applications

Semanttisen webin teknologiat tarjoavat aiempaa parempia mahdollisuuksia datan integrointiin ja linkittämiseen sekä ontologioiden hyödyntämiseen metadatan tuottamisessa. KiWi on itävaltalainen EU-hankkeessa kehitetty ohjelmisto, joka yhdistää piirteitä wiki-ohjelmistojen selkeästä käyttöliittymästä ontologioiden ja linkitetyn datan tukeen sekä sosiaalisen median vapaamuotoiseen aineiston kuvailuun tagien avulla. Ohjelmiston ensimmäinen versio julkaistiin loppuvuodesta 2010, jolloin KiWi-projekti päättyi. Tässä hankkeessa toteutettu ohjelmistoversio perustuu java-pohjaiseen JBoss-sovelluspalvelimeen, joka on varsin järeä ja massiivinen alusta. Tekeillä on kuitenkin KiWi-versio 2.0, joka toteutetaan modulaarisena perustuen kevyempään Jetty-sovelluspalvelimeen.

<http://www.kiwi-project.eu/>

Toinen semanttisen webin teknologiaa hyödyntävä EU-hanke on LOD2. Tämä hanke tarjoaa runsaan valikoiman avoimen lähdekoodin sovelluksia RDF-tietokannoista ontologioiden ylläpitoon ja linkitetyn datan hallintaan. Näistä komponenteista OntoWiki voisi soveltua myös tutkimusdatan hallintaan, silloin kun halutaan tarjota tukea semanttisen webin tekniikoille.

<http://lod2.eu/WikiArticle/TechnologyStack.html>

7.3 Semanttinen Mediawiki

Wikipedian tekninen alusta on Mediawiki, jonka kehittämistä koordinoi Wikimedia Foundation. Semanttinen Mediawiki on tämän ohjelmiston laajennus, joka lisää siihen siihen semanttisen webin ominaisuuksia, esimerkiksi tukea semanttiselle haulle. Tämän ohjelmiston varaan on toteutettu tutkimusdatan hallinnan kannalta kiinnostavia sovelluksia, esimerkiksi kotimainen JärviWiki, sveitsiläinen Swiss Experiment ja brittiläinen FigShare. Se on valittu alustaksi myös Suomen Akatemian rahoittamassa Tieteen termipankki -hankkeessa.

Ohjelmistoon on saatavilla erilaisia lisäpaketteja. kuten SMW+, joka parantaa edelleen sen käytettävyyttä. Semanttinen Mediawiki on lupaava kandidaatti tutkimusdatan hallinnan käyttöliittymäksi, ja sen tarjoamia mahdollisuuksia tulisi selvittää tarkemmin asentamalla ohjelmisto yliopiston virtuaalipalvelimelle.

<http://semantic-mediawiki.org/>

7.4 Terrastore

Terrastore-ohjelmistoa luonnehditaan ajanmukaiseksi dokumenttitietokannaksi, joka tarjoaa skaalautuvuutta ja elastisuutta tinkimättä tiedon eheydestä. Tämä alusta edustaa uudentyyppisiä Nosql-tietokantoja, ja näiden alatyypeistä dokumenttitietokantoja. Tämä ei ole aivan uusi tietokantatyyppi, sillä tähän kategoriaan lukeutuu myös pitkään käytössä ollut Lotus Notes. Ohjelmiston pääkehittäjä on italialainen Sergio Bosso, ja hankkeen tavoitteena on luoda erittäin vakaa ja suurille datamäärille skaalautuva hajautettu tietokantaohjelmisto. Skaalautuvuus perustuu ohjelmiston jakaantumiseen useille palvelinsolmuille, joita tarvitaan vähintään kaksi. Ohjelmisto voidaan integroida myös elasticsearch-hakuliittymään. Tietomalli on joustava, tallennettavan aineiston ei tarvitse noudattaa mitään tiettyä skeemaa. Ohjelmisto on vielä rakenteilla, uusin versio tammikuussa 2012 on versio 0.82.

Tutkimusdatan tallennusta varten tämä ohjelmisto voisi tarjota tietovaraston, jonka lisäksi tarvitaan käyttöliittymän kehittämistä tietovaraston yhteyteen. Ohjelmisto tarjoaa hyvät rajapinnat integrointiin muiden järjestelmien kanssa.

<http://code.google.com/p/terrastore/>

7.5 OrientDB

OrientDB on Nosql-tyyppinen dokumenttitietokanta, joka sisältää myös verkkomallin mukaisesti toimivan verkkotietokannan piirteitä (graph database). Tämä Java-pohjainen ohjelmisto on selkeä ja helposti asennettavissa. Ohjelmiston pääkehittäjä on Luca Garulli, ja projektin kehittäjäyhteisössä on yhteensä 14 kehittäjää. Ohjelmisto on hyvin nopea ja skaalautuu suurillekin datamäärille sekä suurille palvelinmäärille, jotka toimivat yhdessä klusterina. OrientDB-ohjelmiston versio 1.0 julkaistaan helmikuussa 2012. Ohjelmisto soveltuu käytettävissä kokonaan ilman kenttärakennetta tai skeemaa, skeeman kanssa, tai näiden vaihtoehtojen yhdistelmänä.

Ohjelmisto on varsin uusi tulokas, joten sillä ei ole toistaiseksi laajaa käyttäjäyhteisöä. Ominaisuuksuksiltaan se on kuitenkin lupaava, ja sen käyttömahdollisuuksia kannattaisi selvittää tarkemmin.

<http://www.orienttechnologies.com/>

7.6 MongoDB

Tarkasteltavista dokumenttitietokannoista MongoDB on kaikkien pisimmälle tuotteistettu, ja sillä on laaja käyttäjäkunta erityisesti yrityssovellusten alustana. Ohjelmistoa käytetään myös tutkimusdatan käsittelyyn, esimerkiksi CERN-tutkimuskeskuksen LHC-hiukkaskiihdyttimen data-aineistojen hallintaan. Myös brittiläinen National Archives käyttää tätä ohjelmistoa verkkopalvelunsa alustana. Jyväskylän yliopiston tietotekniikan laitoksella on tehty tällä ohjelmistolla opiskelijatyönä prototyyppi tutkimusaineiston hallintaan. Tässä sovelluksessa MongoDB valittiin taustalla toimivaksi tietokannaksi, ja web-käyttöliittymä toteutettiin Ruby on Rails -sovelluskehityksellä.

MondoDB-ohjelmiston kehittämistä koordinoi yritys nimeltä 10gen, joka myös tarjoaa tähän alustaan liittyviä kaupallisia tukipalveluja.

<http://www.mongodb.org/>

7.7 SenseiDB

Uusin tulokas tässä esiteltävien tietokantojen joukossa on SenseiDB, jonka versio 1.0 julkaistiin tammikuussa 2012. Ohjelmistoa kehitetään suosittuun LinkedIn-verkostoitumispalvelun tarpeisiin. Kyseessä on Java-pohjainen

dokumenttitietokanta, joka tarjoaa vikasietoisuutta, hajautusta useille palvelinsolmuille, kapasiteettia suurille datamäärille sekä skaalautuvuutta. Ohjelmisto kykenee myös nopeaan indeksointiin ja hakuun sekä reaaliaikaiseen päivitykseen. Sitä on myös mahdollista räätälöidä plugin-laajennuksilla. SenseiDB on suunniteltu haku- ja navigointiliittymäksi dokumenteille, jotka sisältävät sekä strukturoimatonta tekstiä että rakenteisessa muodossa olevaa metatietoa.

<http://senseidb.com/>

7.8 CKAN - the open source data portal software

Tiedon avoimuutta edistävä organisaatio Open Knowledge Foundation kehittää CKAN-ohjelmistoa dataportaalien alustaksi. Tätä ohjelmistoa käytetään esim. uudessa EU:n dataportalissa <http://publicdata.eu/>, joka on parhaillaan beta-vaiheessa. Ohjelmisto tarjoaa monipuolisia välineitä data-aineistojen tallentamiseen, kuvailuun, julkaisemiseen ja hakuun. Siihen kuuluu myös versionhallinta ja standardi rajapinta, jonka avulla ohjelmisto voidaan integroida muihin tietojärjestelmiin (RESTful JSON API). Ohjelmiston avulla data-aineistoa voidaan myös visualisoida ja siihen sisältyy paikkatiedon hallintaan sovelituvaa piirteitä, esim. karttanäkymä ja haku alueen perusteella.

CKAN-ohjelmistoa kehitetään aktiivisesti ja sitä käytetään laajalti avoimen datan julkaisemiseen, ja tähän käyttöön sen ominaisuudet soveltuvat parhaiten. Sen sijaan tutkimusryhmän sisäiseen käyttöön tai rajatulle käyttäjäkunnalle tarjottavaan palveluun se ei tarjoa valmiita välineitä.

<http://ckan.org/>

7.9 DSpace

Avoimeen lähdekoodiin perustuvan DSpace-ohjelmiston kehitystä koordinoi DuraSpace-organisaatio. DSpace on avaimet käteen -periaatteella kehitetty ohjelmisto, joka on helppo ja nopea laittaa käyttökuntoon. Toisaalta modulaarinen rakenne mahdollistaa helposti myös eri tasoiset kustomoinnit. DSpace mahdollistaa tarkan pääsynhallinnan, jonka avulla voidaan rajoittaa pääsyä aineistojen hyvinkin tarkasti esimerkiksi käyttäjän, ryhmän tai IP-osoitteen perusteella. Ohjelmisto tarjoaa myös monipuoliset rajapinnat tiedon hakuun, haravointiin ja hallinnointiin muun muassa OpenSearch, OAI-PMH, OAI-ORE ja SWORD protokollien avulla. Versiosta 1.7.x lähtien mukana on ollut myös mahdollisuus käyttää suosittua SOLR-hakukonetta ja sen tarjoamaa moninäkömähakua arkiston selailuun ja hakuun.

DSpace sopii hyvin tiedostotyyppisen datan ja yksinkertaisen avain-arvo-pareihin perustuvan metadatan hallinnointiin. Toisaalta yleiseksi datavarastoksi tarkoitettu ohjelmisto ei tarjoa suoraan välineitä esimerkiksi multimediatiedostojen helppoon esittämiseen suoraan www-käyttöliittymässä.

Esimerkkejä DSpace-pohjaisista data-arkistoista ovat muutaman amerikkalaisen yliopiston kehittämä Dryad <http://www.datadryad.org/> ja Edinburghin yliopiston DataShare <http://datashare.is.ed.ac.uk/>.

<http://www.dspace.org/>

7.10 hydra

Hydra-ohjelmisto on kokoelma ohjelmistokomponentteja, joiden avulla voidaan toteuttaa erilaisia käyttöliittymiä taustalla toimivan digitaalisen arkiston arkiston varaan. Hankkeen pääosapuolina on viisi tutkimusyliopistoa, joista yksi on brittiläinen ja muut amerikkalaisia, mukaanlukien Stanfordin yliopisto. Hydra-järjestelmässä datan tallennuskomponenttina on Fedora-arkisto, ja käyttöliittymän komponentti on Blacklight, joka on toteutettu Ruby on Rails -sovelluskehityksen avulla. Ohjelmisto tukee neljää metadataformaattia sekä räätälöityjä käyttäjäkohtaisia formaatteja.

Kysymyksessä on melko järeä ohjelmistopaketti, jonka käyttöönotto ei ole aivan suoraviivaista. Suomessa tämäntyyppistä ratkaisua kehittää Oulun yliopiston kirjasto julkaisuarkistolleen, jossa Fedora-pohjaisen arkiston käyttöliittymäksi on valittu Vufind-ohjelmisto.

<http://projecthydra.org/>

7.11 Archivematica

Archivematica-ohjelmistoa kehitetään digitaalisen aineiston pitkäaikaissäilytykseen, ja se toteuttaa tähän tarkoitukseen laadittua OAIS-mallia. Kyseessä on yksi harvoista avoimen lähdekoodin hankkeista pitkäaikaissäilytyksen saralla. Hankkeen taustalla on kanadalaisia arkistoalan organisaatioita, kaksi yliopistoa sekä Unesco. Ohjelmiston kehitykseen osallistuu myös kanadalainen yritys Artefactual Systems, joka tarjoaa digitaaliseen pitkäaikaissäilytykseen liittyviä tukipalveluja. Ohjelmiston nykyinen versio on 0.8 alpha, ja versio 1.0 suunnitellaan julkaistavaksi syksyllä 2012.

Tämän ohjelmiston avulla voisi olla mahdollista tuottaa jatkossa pitkäaikaissäilytyksen edellyttämiä digitaalisia objekteja. Ohjelmisto ei kuitenkaan sovellu ilman räätälöintiä monimuotoisen tutkimusdatan hallintaan, sillä se on suunniteltu etupäässä arkisto- ja julkaisuaineistoa silmälläpitäen.

<http://archivematica.org/>

7.12 iRODS

Avoimen lähdekoodin iRODS-ohjelmistoa (Integrated Rule-Oriented Data System) kehittää amerikkalainen DICE Center, joka toimii University of North Carolinan yhteydessä. Monipuolisten ominaisuuksiensa vuoksi iRODS-ohjelmistoa on hyödynnetty laajasti eri maiden datakeskuksissa. Eri iRODS-palvelut voidaan myös integroida siten, että yhden palvelun käyttäjät pystyvät jakamaan datansa toisen iRODS-palvelun käyttäjien kanssa. Ohjelmiston avulla tallennettu data voidaan hajauttaa useille palvelimille, ja se tarjoaa sekä selainkäyttöisen liittymän että komentorivipohjaisen liittymän, jonka avulla ohjelmisto voidaan integroida muihin järjestelmiin.

Suomessa iRODS-ohjelmistoa on käytetty CSC:llä vuodesta 2010 lähtien IDA-tallennuspalvelun alustana. Ohjelmistoa käytetään myös Helsingin yliopistossa Biotekniikan instituutin valomikroskopian ryhmässä, jossa sen on integroitu tämän erikoisalan ohjelmistoon.

<https://www.irods.org/>

8 Liite Haastattelukysymykset

8.1 Helsingin yliopiston kirjaston tutkimusdatatyöryhmä

Tutkimusryhmäpilotti: tutkimusdatatilanteen kartoitus

8.1.1 Haastattelurunko

Haastateltava tutkimusryhmä/tutkija:

Tiedekunta ja laitos:

Tieteenala/tieteenalat:

Päivämäärä:

Haastattelijat:

Hankkeen nimi/tutkimusryhmän nimi:

8.1.2 I TAUSTA

Hankkeen/tutkimusryhmän vaihe?

- olemassa oleva hanke
- uusi hanke
- uusi tutkimusryhmä
- hankkeen suuruusluokka

Haastateltavan työn vaihe?

- tutkimussuunnitelma
- datan keruu
- kirjoitus
- julkaiseminen
- rahoituksen hakuvaihe

Hankkeen rahoittajat?

- kotimainen
- kansainvälinen rahoittaja
- yksityinen rahoittaja

Yhteistyökumppanit? Muut tutkijat/tutkimusryhmät/hankkeet...

- Helsingin yliopistossa
- Muissa korkeakouluissa
- Tutkimuslaitoksilla
- Yritykset
- Kansainvälisiä kumppaneita

8.1.3 II AINEISTONHALLINTA

Onko olemassa aineistohallintasuunnitelma?

- aineistohallintasuunnitelman laatiminen
- onko ohjeistusta? mikä/missä?
- vastuututkijan rooli, tutkijan rooli?
- miten käsittelet? ohjelmistot? ongelmia – mitä?

Onko jollakin taholla tai toimijalla vaatimuksia aineistohallinnan suhteen?

- rahoittajat kuten Suomen Akatemia
- julkaisija
- laitoksen sisäinen ohjeistus)

8.1.4 III DATAINVENTOINTI

Miten tutkimusaineistot hankitaan?

- olemassa oleva oma data, mistä?
- hankitaan data muualta, mistä?
- kerätään oma data

Mitä omaa tutkimusaineistoa/dataa on olemassa tällä hetkellä? Miten sitä hankitaan? Ohjelmistot, formaatit? Miksi päädyitte tiettyyn ohjelmistoon/formaattiin?

- Raakadataa
- Havaintodataa
- Empiiristä dataa
- Simulaatioihin perustuva dataa
- Epäsuoraa tai koottua dataa
- Muistiinpanoja/päiväkirjoja
- Näytteitä
- Valokuvia
- Ääni- ja kuvanauhoja
- Muuta

Missä aineisto sijaitsee?

- Omalla koneella
- CD:llä/USB-tikulla/ulkoisella kovalevyllä
- Laitoksen palvelimella
- Tietotekniikkakeskuksen palvelimella
- CSC:llä
- Kansallisessa data-arkistossa
- Kansainvälisessä data-arkistossa
- Aiheenmukaisessa arkistossa
- onko alallanne kansainvälinen aiheenmukainen arkisto tutkimusdatalle?
- Rahoittajan arkistossa
- Julkaisijan arkistossa
- Muussa arkistossa

Onko metadataa, onko tiedostot kuvailtu? Datan rakenne, muuttujaluettelot

8.1.5 MUU DATA

Pystytkö arvioimaan, kuinka suurista aineistomääristä on kyse säilytyksen kannalta? Paljonko dataa syntyy vuodessa?

Käytetäänkö projektissa muiden tuottamaa dataa?

Kyllä. Mitä ja missä?

- Data-arkistot
- Muut organisaatiot
- Kollegat
- Yhteistyökumppanit
- Open data Internetissä

Ei tällä hetkellä, mutta haluaisi käyttää

- Mitä – missä - miten?
- Yhteisprojektit
- Rekisteröityminen
- Lisensoitu käyttö
- Open data
- Meta-analyysit

Onko jokin käytettävä tai haluttu data maksullista? Hintaa?

8.1.6 V Onko tulossa lisää dataa (mitä dataa tullaan keräämään)? (samat kysymykset näistä)

Tutkimusmenetelmät

8.1.7 VI DATAN KÄYTTÖ, OIKEUDELLISET ASIAT

Mitkä ovat aineiston omistus- ja käyttöoikeudet?

- Omistusoikeus (kuka omistaa datan ja kenellä on oikeus määrätä sen käytöstä? tutkija/tutkimusryhmä, joku muu taho? Kenelle oikeudet saattaisivat siirtyä?)
- Tekijänoikeudet
- Yksityisyydensuoja, anonymisointi
- Eettiset periaatteet
- Luvat aineiston jatkokäyttöön
- Sisältyykö dataan henkilötietolainsäädännön alaista aineistoa? Jos sisältyy, mihin kaikkiin käyttöihin tutkimuskohteilta on hankittu oikeudet?
- Valtion turvallisuuteen liittyvät asiat
- Luonnonsuojelulliset näkökohdat

8.1.8 VII DATAN JATKOKÄYTTÖ

- Onko ollut harkinnassa käyttää itse jatkotutkimuksiin?
- Entä missä vaiheessa datan voisi jakaa? Embargoaika?
- Kenelle voisi jakaa ja millä ehdoilla?
- Voidaanko asettaa avoimeen arkistoon/tietokantaan vapaasti saataville?
- Entä pelkkä metadata (esim. jos aineisto on vain paperimuodossa)?

- Verifiointi, kuinka kauan tulisi säilyttää? Edellyttävätkö alanne julkaisut tätä?
*Jos tutkimusdataan perustuen on julkaistu, löytyykö julkaisusta linkki johonkin osaan tutkimusdatasta?
Esim. niin että data on jossain muodossa tallennettu ulkomaiseen tietokantaan ja linkitetty joko suoraan tai taulukkotiivistelmän kautta artikkeliin.

Mitkä olisivat parhaat kannustimet hyvälle aineistohallinnalle? (meriitti, viittaukset, uskottavuus, rahoitus, prosessi?)

Mitkä ovat suurimmat ongelmat/haasteet?

Mitä tukea/toimenpiteitä toivoisitte HY:ltä datanhallinnan ja jatkokäytön suhteen?

Miten näette datan pitkäaikaissäilytystarpeen?

9 Liite Kysymyksiä pilottiryhmälle

Kysymyksiä pilottiryhmälle:

9.1 Prosessi:

- Kuvaus tehtävästä tutkimuksesta
- Tallennus-, muokkaus- ja katseluprosessit kuvattuna omin sanoin
- Kuvaus järjestelmätarpeesta omin sanoin
- Millaisia ohjelmistoja ja tiedostomuotoja on käytössä?
- Onko täydennettävää jo laadittuun listaan (ks. Excel)

9.2 Apukysymyksiä:

- Mistä dataa täytyy voida tallentaa
- Ketkä tallentavat dataa
- Millaisella laitteistolla
- Täytyykö dataa voida tallentaa automatisoidusti ulkoisista järjestelmistä / mittalaitteista?
- Mistä dataa täytyy voida muokata?
- Ketkä muokkaavat dataa
- Millaisella laitteistolla
- Millä tavoin dataa pitäisi voida muokata?
- Mistä dataa täytyy päästä tarkastelemaan / hyödyntämään
- Ketä tarkastelevat dataa
- Millaisella laitteistolla
- Täytyykö dataa voida tuoda ulkoisiin järjestelmiin (rajapinnat)?
- Tarvitaanko ym. toimintoihin käyttäjähallintaa
- Tarvitaanko esim. Web-liittymää ja jos tarvitaan, täytyykö olla esim. sisällytettyjä video-/audio-streamaus-toimintoja

9.3 Lisäominaisuuksia:

- Pitäisikö dataa pystyä analysoimaan (esim. toteuttamaan matriisioperaatioita)?
- Pitäisikö järjestelmässä olla ryhmätyöominaisuuksia (pikaviestimet, foorumit)
- Raportointiominaisuuksia?
- Millaisia hakuominaisuuksia halutaan?
- Millaisia tarpeita datan julkaisulle?
- Millaisia vaatimuksia metadatalle, onko tarvetta auditointidatalle?
- Millaisia vaatimuksia versionhallinnalle?
- Muuta?

9.4 Data:

- Kuinka paljon dataa kertyy, kuinka paljon tilaa tarvitaan?
- Millaisia tiedostotyyppisiä halutaan tallentaa?
- Tallennusrytmi
- Halutaanko dataa tallentaa eräajoina vai bittivirtana (stream)? Jos eräajoina, kuinka suurista eräajoista olisi kysymys?
- Onko vaatimuksia tallennusnopeudelle?
- Kumpi on tärkeämpää: levytinnän nopeus vai levytinnän hinta?
- Säilytyksen pitkäkestoisuus
- Kuinka kauan dataa tulee minimissään säilyttää?

- Onko tarvetta datan pitkäaikaiseen arkistointiin (yli 5v)?
- Jos on, niin kuinka paljon arkistoitavaa dataa arvioidaan olevan?
- Tietoturvan taso
- Onko vaatimuksia liittyen datan fyysiseen säilytyspaikkaan (Euroopassa, Suomessa, HY:n tiloissa?)
- Täytyykö järjestelmään pääsyä rajata tietoliikennetasolla (esim. rajoittaa pääsy vain laboratorion sisältä tms)?
- Kuinka arkaluontoista data on? Tarvitaanko esim. salattua tietoliikennettä?