

HELSINGIN YLIOPISTON DATAPOLITIikka

Tutkimusdatatyöryhmän ehdotukset
Toukokuu 2011

Sisällysluettelo

1	JOHDANTO	1
2	DATAPOLITIIKAN TAVOITE	3
3	YLIOPISTOLLA TOTEUTETTY DATASURVEY.....	5
3.1	Johdanto.....	5
3.2	Tutkimusaineiston keräys	5
3.3	Tulokset	6
4	KUSTANNUKSET	9
5	DATAPOLITIIKAN LÄHTÖKOHTA.....	10
5.1	Kuka omistaa tutkimusaineiston?.....	11
5.2	Miten datan saatavuus järjestetään?.....	12
5.3	Miten yhteensopivuus ja yhteiskäyttö varmistetaan?	13
5.4	Miten meritoitumista tuetaan?	14
5.5	Miten datan säilytys ja pitkäaikaissäilytys turvataan?	14
5.6	Integroitu informaatioympäristö	15
6	VÄHIMMÄISTAVOITE HELSINGIN YLIOPISTON DATAPOLITIIKALLE ..	16
6.1	Yleiset tavoitteet.....	16
6.2	ESFRI infrastruktuureiden datapolitiikka.....	17
7	DATATYÖRYHMÄN EHDOTUKSET	19

1 JOHDANTO

Helsingin yliopiston vararehtori Hannele Niemi nimitti vuonna 2008 tutkimusdatatyöryhmän, jonka puheenjohtajaksi kutsuttiin tutkimusjohtaja Marjut Salokannel ja jäseniksi seuraavat henkilöt:

1. Ryhmänjohtaja Petri Auvinen
2. Vararehtori, professori Johanna Björkroth
3. Kirjastonhoitaja Maria Forsman
4. Johtaja Pirjo–Leena Forsström
5. Dosentti Tuula Gordon
6. Yliopistonlehtori Kalle Korhonen
7. Professori Markku Koskenvuo
8. Tutkimusprofessori Helena Kääriäinen
9. Lehtori, dosentti Minna Laakso
10. Professori Markku Löytönen
11. Professori Terttu Nevalainen
12. Kampuskirjaston johtaja Annikki Roos
13. Professori Katri Räikkönen
14. Tietohallintojohtaja Ilkka Siissalo
15. Professori Esko Ukkonen

Työryhmän puheenjohtaja kutsui työryhmälle sihteerin, joka oli vuoden 2009 loppuun Jenni Laitinen ja vuoden 2010 alusta lähtien Mette Ranta Helsingin yliopiston tutkijakollegiumista. Työryhmässä on kuultu asiantuntijoina terveyden tutkimuksen yksikön johtajaa Mikael Fogelholmia Suomen Akatemiasta, tietotekniikkapäällikkö Minna Harjunniemeä Helsingin yliopiston tietotekniikkakeskuksesta ja arkistopäällikköä Juha Hannikaista Helsingin yliopiston keskusarkistosta. Työryhmän puheenjohtaja oli mukana myös kansallisessa ”Tutkimuksen tietoaineistot” – ohjausryhmässä.

Tutkimusdatatyöryhmän tehtävänä oli analysoida miten Helsingin yliopiston sisällä eri tieteenaloilla voitaisiin saada tutkimusdata paremmin tutkimusyhteisön käyttöön; ensisijaisesti tutkimustulosten verifioimista varten ja alasta riippuen, myös laajemmin jatkotutkimuksen hyödynnettäväksi. Tutkimusdata on keskeinen

tutkimuksen perusinfrastruktuuri ja datan saatavuus ja käytettävyys on tutkimusinfrastruktuurin oleellinen osa.

Opetus- ja kulttuuriministeriö on tukenut tutkimusdatatyöryhmän työtä 50 000 euron suuruisella rahoituksella, jonka avulla Marjut Salokannel on laatinut tutkimusdatan saatavuuteen liittyviä oikeudellisia kysymyksiä koskevan raportin (liitteenä). Koska raportissa on laajasti käyty läpi yliopistojen tuottaman tutkimusdatan saatavuuteen liittyviä kysymyksiä, näitä asioita ei ole erikseen sisällytetty työryhmän raporttiin vaan se käsittää kiteytetysti ehdotuksen Helsingin yliopiston datapolitiikaksi ja sen toteuttamisen edellyttämiksi toimenpiteiksi.

Tutkimusdatatyöryhmä tiedostaa, että tutkimusdata muodostaa yhden keskeisen tieteellisen tutkimuksen perusinfrastruktuurin, joka rakentuu datan avoimelle saatavuudelle ja käytettävyydelle. Koska datainfrastruktuurin luominen edellyttää yliopistolta johdonmukaista toimintalinjausta, työryhmä asetti työnsä tavoitteeksi yhteisen datapolitiikan laatimisen Helsingin yliopistolla syntyvän tutkimusdatan hallintaa ja käyttöä varten. Työryhmä tiedostaa ja korostaa, että tutkimusdata ja sen käyttö on hyvin erilaista eri tieteen aloilla. Joillain aloilla voi olla pitkälle vakiintuneet kansainväliset käytännöt ja olemassa olevat kansainväliset tietokannat, joihin tutkimusdata toimitetaan tulosten julkaisun yhteydessä tai tapauksesta riippuen jo aikaisemmin. Näin on esimerkiksi geeni- ja genomitutkimuksen alalla. Toisena ääripäänä voi olla esimerkiksi oikeustiede, jossa tutkimusdata useimmiten on julkaistua virallisaineistoa (säädöskokoelmat, oikeustapaukset) ja tutkimuksessa käytetty data dokumentoidaan pääasiassa lähdeviitteiden avulla. Tästä huolimatta työryhmä pitää mahdollisena ja tärkeänä tiettyjen yleisten datan keräämistä ja käyttöä koskevien periaatteiden lukkoon lyömistä yliopiston tasolla. Tämä on erityisen tärkeää jo siitä syystä, että kansainväliset ja useat kansalliset tutkimusrahoittajat edellyttävät datan hallintasuunnitelmaa ja datan avointa saatavuutta kaikilla tieteen aloilla.

2 DATAPOLITIIKAN TAVOITE

Helsingin yliopiston datapolitiikan tavoitteena on varmistaa, että tutkimusprosessin aikana kerättävä data ja aineistot säilytetään asianmukaisesti ja ne ovat tutkijayhteisön saatavissa ja käytettävissä mahdollisimman laajasti ja tehokkaasti kuitenkin niin, ettei tämä vaaranna tutkimuksen kohteiden tai muuten siinä esiintyvien henkilöiden yksityisyyden suojaa tai mahdollisten kolmansien tahojen immateriaalioikeuksia. Tämä edellyttää myös, että datan omistajuus on yksiselitteisesti määritelty jo datan keräämisvaiheessa.

Tutkimusdatan keskiössä on aina viime kädessä yksittäinen tutkija tai tutkimusryhmä, vaikka tämän lisäksi lukuisat muut toimijat – kuten rahoittaja, muut tukijat, kustantajat – tutkimusprosessin eri vaiheissa tavalla tai toisella ovat kiinnostuneita kerättävästä datasta. Tavoitteena on, että tässä datapolitiikassa määritellyt toimintatavat selkiyttävät tutkimusdataan liittyviä periaatteita ja käytäntöjä ja siten helpottavat jokaisen tutkijan työtä HY:ssä.

Datan saatavuuden ja käytettävyyden varmistamisessa on ainakin seuraavat kolme näkökulmaa:

1. Tutkimustulosten verifiointi

Tieteellisen tutkimuksen peruseriaate on, että jotta tutkimustulosta voitaisiin pitää tieteellisten metodologian mukaan syntyneenä, se on voitava verifioida tausta-aineiston avulla. Luonnontieteissä tämä voi tarkoittaa tutkimustuloksen toistamista datan avulla. Useat kansainväliset tieteelliset julkaisut edellyttävätkin, että julkaisun taustadata talletetaan tietopankkiin, jossa se on lukijoiden saatavissa.

Tutkimustulokset pitäisivät olla verifioitavissa kautta yliopiston jo nyt, ja jos näin ei ole, tiedekunnissa on pidettävä huoli, että julkaistavien tutkimusten tulosten taustadata on saatavissa verifiointia varten joko yliopistossa tai asianomaisessa kansainvälisessä tietokannassa.

2. Tutkimusdatan jatkokäyttö

Kansainväliset rahoittajat edellyttävät enenevässä määrin, että tutkimusjulkaisujen lisäksi myös tutkimusdata täytyy olla avoimesti saatavissa ja hyödynnettävissä edellyttäen, että henkilötietojen suojaan liittyvät kysymykset eivät tätä rajoita. Suomessa opetus- ja kulttuuriministeriön nimittämä poikkihallinnollinen tutkimuksen tietoaineistot -ohjausryhmä esittää loppuraportissaan, että kaikki julkisin varoin tuotettu tieto pitäisi olla saatavissa tutkimuksen käyttöön. Valtiovarainministeriö ja liikenneministeriö esittävät niin ikään, että julkisin varoin tuotettu tieto pitää olla lähtökohtaisesti koko yhteiskunnan hyödynnettävissä.

Koska ulkopuoliset paineet joka tapauksessa tulevat edellyttämään yhä laajempaa tutkimusdatan saatavuutta myös datan jatko- ja hyödyntämistä varten, Helsingin yliopiston pitää luoda mekanismit tämän mahdollistamiseksi. Tämä tarkoittaa teknisiä valmiuksia laitteiden, metadatan ja yhteensopivien rajapintojen muodossa, juridisten kysymysten ratkaisuja ja selkeiden datan hallintakäytänteiden luomista.

3. Tutkimusaineistojen pitkäaikaissäilytys

Helsingin yliopistolla on paljon erittäin arvokkaita ja kansainvälisesti merkittäviä tutkimusaineistoja, joiden säilyttäminen jälkipolville olisi erittäin tärkeää. Kansallisella tasolla on suunnitteilla tutkimusaineistojen pitkäaikaissäilytystä koskevan hankkeen aloittaminen, minkä vuoksi tämä aihekokonaisuus on rajattu yliopiston tutkimusdatatyöryhmän työn ulkopuolelle, vaikka työryhmä toteaaakin, että se liittyy kiinteästi aineistojen käyttöön.

Tässä raportissa esitetään jäljempänä mitä Helsingin yliopisto voisi tehdä kahden ensimmäisen tavoitteen toteuttamiseksi yliopistolla. Pitkäaikaissäilytystä koskeva työ on järkevä tehdä yhteistyössä muiden toimijoiden kanssa kansallisella ja

kansainvälisellä tasolla. Ensiksi tarkastellaan käytännön tutkijoiden tarpeita tutkimusdatan tuottamisen, keräämisen, käytön ja säilytyksen suhteen. Tämän jälkeen tarkastellaan yleisemmällä tasolla, miten tutkimusdatan saatavuus ja hallinta voitaisiin toteuttaa Helsingin yliopistossa.

3 YLIOPISTOLLA TOTEUTETTY DATASURVEY

Tutkimusdatatyöryhmä teki työnsä pohjaksi vuonna 2009 laajan koko yliopiston kattavan kyselytutkimuksen, jossa kartoitettiin eri laitosten kantoja tutkimusdatan teknisen infrastruktuurin, datan määrän, siihen liittyvien oikeudellisten kysymysten ja rahoitus- ja muiden käyttötarpeiden suhteen. Seuraavassa esitetään tiivistetysti kyselyn tuloksia.

3.1 Johdanto

Tutkimusdatatyöryhmä toteutti kyselyn Helsingin yliopiston laitosten tutkimusaineiston säilytyksestä keväällä 2009. Kyselyn tavoitteena oli kartoittaa nykyinen tutkimusaineiston kertymä, käyttö ja tulevaisuuden tarpeet.¹ Kysely koski ainoastaan digitaalisessa muodossa olevaa primaaritutkimusaineistoa, tosin vastauksia saatiin myös muun aineiston osalta. Vastaajia pyydettiin vastaamaan tutkimuslomakkeen kysymyksiin instituuttinsa tai laitoksensa edustajana niiltä osin, jotka ovat relevantteja hänen edustamallaan alalla ja sen mukaan kuinka tärkeänä hän pitää aineiston saatavuutta laajemmalle tutkimusyhteisölle.

3.2 Tutkimusaineiston keräys

Kyselyitä lähetettiin helmikuussa 2009 kaikkiaan 210 kappaletta. Kyselyyn vastasi 49 henkilöä edustamansa instituution tai

¹ Jenni Laitinen Helsingin yliopiston tutkijakollegiumista oli vastuussa kyselyn toteuttamisesta ja aineistonkeruusta.

laitoksen puolesta yhteensä yhdeksästä tiedekunnasta. Kolme laitosta ei kuulunut mihinkään tiedekuntaan. Vaikka kyselyaineisto on raportin kannalta tärkeä ja sisältää paljon olennaista tietoa, on pidettävä mielessä että otoskoko on melko pieni yliopiston kokoon nähden ja vastausprosentti jäi valitettavan pieneksi. Kaikista vastauksista ei ilmennyt esimerkiksi teknisiä reunaehtoja tai muita tarvittavia sovelluksia. Tämän vuoksi sovellusten kustannuksia ei ole arvioitu lukuun ottamatta varmuuskopiointiin ja laitteistojen hallintaan liittyviä lisenssimaksuja.

Tutkimuskysely koostui kahdesta osasta. Ensimmäisessä osassa selvitettiin laitoksessa käytettävän aineiston nykytilaa ja toinen osa oli teknisempi ja tulevaisuuden tarpeisiin keskittyvä. Kyselyn saatekirje ja itse kyselylomake ovat raportin liitteinä (1. ja 2.).

3.3 Tulokset

Nykyinen tilanne

Suurimmassa osassa laitoksista käytetään hyvin erilaisia tutkimusaineistoja. 40 % vastaajista antoi useita tutkimusaineistotyyppisiä, kuten potilastietoja, haastatteluita, tekstiä ja niin edelleen. Erityisesti lääketieteellisen tiedekunnalla on erilaisia aineistoja käytössä, samoin kuin matemaattisluonnontieteellisen tiedekunnan laitoksilla.

Taulukkoon 1. on jaoteltu vastaukset sekä lukumäärän että tarvittavan kapasiteetin mukaan. Huomattavalla enemmistöllä tarpeet ovat varsin pieniä; 80 % vastanneista uskoi selviävänsä enintään kymmenen teratavun kapasiteetilla. Kustannusarvioita laskettaessa todella suuret tarpeet on jätetty arvioinnista pois. Näin siksi, että kyseessä oleva tarve tulisi analysoida tarkemmin avoimeksi jäävien kysymysten osalta: oheiskuluina saattaa tulla huomattavia tietoliikennekuluja, tallennus- ja palvelinverkko saattaa tarvita huomattavasti nykyistä nopeampaa ja tehokkaampaa teknologiaa.

kokoluokka	Tt yhteensä	vastausten lkm
alle 1 Tt	2	23
1 – 10 Tt	15	11
10 – 50 Tt	90	3
100 Tt	200	2
1000 Tt (1 Pt)	3 000	3
16000 (16 Pt)	16 000	1

Taulukko 1. Vastausten jakauma

Selvä enemmistö (79 %) vastaajista ilmoitti aineiston säilytysaikatarpeen olevan vähintään 10 vuotta. Lisäksi on huomioitava, että aineistotyyppin ja tarpeiden erilaisuuksista johtuen laitoksissa voidaan säilyttää joitain aineistoja lyhyempiä jaksoja (vastausvaihtoehtoina kaksi vuotta tai kahdesta kymmeneen vuoteen) ja toisia pidempien jaksojen ajan.

Kyselyn mukaan suurin osa tutkimusaineistoista säilytetään omalla laitoksella. Myös ulkomailla ja yhteistyökumppanien kanssa (esimerkiksi CSC, SKS, Kotus ja FSD) säilytettiin aineistoa. Tässäkin useat vastaajat, kuten puhetieteiden laitos, Suomen molekyyliääketieteen instituutti (FIMM), kemian laitos, tietojenkäsittelytieteen laitos ja Biotekniikan instituutti valitsivat useita vastausvaihtoehtoja. Oman laitoksen lisäksi ulkomailla aineistojaan säilyttivät soveltavan kemian ja mikrobiologian laitos, metsäekologian laitos ja Englannin kielen vaihtelun, ja kontaktien ja muutoksen tutkimusyksikkö (VARIENG). Omassa laitoksessa ja yhteistyökumppanin kanssa aineistoaan säilyttivät Suomen kielen ja kotimaisen kirjallisuuden laitos, Genomibiologian tutkimusohjelma, Suomen molekyyliääketieteen instituutti (FIMM), Hyytiälän metsäasema, klassillisen filologian laitos, biolääketieteen laitos, Materiaalifysiikan osasto ja maantieteen laitos. Arvioidun tallennustilan koon mukaisessa suuruusjärjestyksessä suurimmilla yksiköillä on kansainväliset mahdollisuudet tallennukseen.

Tutkimusaineistojen käyttöoikeudet

Oikeudet käyttää Helsingin yliopiston laitoksissa kerättyä tutkimusaineistoa ovat pääosin laitoksilla tai laajemmin yliopiston sisällä. Yli puolen vastaajista (57 %) mukaan aineisto on käytettävissä vain tutkimusprojektissa työskentelevien tutkijoiden keskuudessa tai laitoksen sisällä. Tähän ryhmään kuului laitoksia lähes kaikista kyselyyn vastanneista tiedekunnista, kuten farmasian, humanistisen, käyttäytymistieteellisen, lääketieteellisen, matemaattisluonnontieteellisen ja valtiotieteellisen tiedekunnista. Matematiikan ja tilastotieteen laitos ilmoitti sen sijaan kaiken aineiston olevan julkista.

Tutkimusaineiston käyttö on täysin rajoitettua käyttäytymistieteellisen tiedekunnan laitoksilla. Myös klinisen tuotantoeläinlääketieteen laitos, kehitysmaatutkimuksen laitos, farmakologian ja toksikologian osasto ja metsäekonomian laitos vastasivat, että heidän aineistonsa käyttö on rajoitettua. Rajoitettua, luvanvaraista käyttöä puolsivat myös farmaseuttisen kemian osasto, musiikkitieteen osasto, Fysiikan tutkimuslaitoksen CMS-ohjelma, yleisen valtio-opin laitos ja Svenska social- och kommunalhögskolan. Aineistojen globaalia verkkopohjaista saatavuutta kannattivat muun muassa biolääketieteen laitos, biotekniikan instituutti, maantieteen laitos, yleisen valtio-opin laitos, soveltavan kemian ja mikrobiologian laitos, biofarmasian ja farmakokinetiikan osasto, klassillisen filologian laitos ja tietojenkäsittelytieteen laitos.

Säilytettävien aineistojen koko

Vastaajilta tiedusteltiin, säilyttävätkö he aineistonsa pieninä (alle 100 kB/tiedosto), keskikokoisina (100 kB–1 GB/tiedosto) vai suurina (suurempi kuin 1 GB/tiedosto) tiedostoina. Pelkästään suurina tiedostoina aineistoaan säilyttävät biofarmasian ja farmakokinetiikan osasto, Ruralia-instituutti, Fysiikan tutkimuslaitoksen CMS-ohjelma, yleisen valtio-opin laitos ja kotitalous- ja käsityötieteiden laitos. Kyselyaineiston perusteella

aineistoa säilytetään Helsingin yliopistolla ainoastaan pieninä tiedostoina kemian laitoksella, matematiikan ja tilastotieteen laitoksella, musiikkitieteen osastolla, Svenska social- och kommunalhögskolanissa, psykologian laitoksen potilastietotutkimuksissa ja metsäekonomian laitoksella. Lopuilla laitoksilla aineisto on säilytetty useissa erikokoisissa tiedostoissa.

Datan säilytystarve

Keskitettyä tutkimusaineiston säilytystä toivoivat 40 % vastaajista, jotka edustivat tasaisesti eri tiedekuntia: biotieteellistä, farmasiaa, humanistista, käyttäytymistieteellistä, lääketieteellistä, maatalousmetsätieteellistä, matemaattisluonnontieteellistä, valtiotieteellistä ja erillislaitoksia. Matemaattisluonnontieteellisen tiedekunnan laitokset kuitenkin pääosin olivat keskittämistä vastaan, samoin kuin käyttäytymistieteellisessä tiedekunnassa psykologian laitos.

4 KUSTANNUKSET²

Yhteenveto kustannuksista

Taulukkoon 2. on koottu eri ratkaisuvaihtoehtojen kustannukset nyt käytössä olevalla teknologialla. Kilpailutilanne saattaa alentaa kustannuksia vielä tästä merkittävästikin, ja toisaalta nopeamman ja hitaamman levytilan suhdetta säätelemällä saadaan kustannuksia myös alaspäin. Tallennustilan osalta ei ole erikseen kuvattu tietotekniikkakeskuksen työntekijäkuluja.

² Tämän luvun on laatinut tietotekniikkapäällikkö Minna Harjunniemi Helsingin yliopiston tietotekniikkakeskuksesta. Tarkempi kuvaus kustannuksista liitteessä 2.

Tarve	Ratkaisuehdotus	Hintataso (2010)	Elinkaari
50–100 Tt (yhteensä runsaat 200 Tt)	SAN	325 000€	5v
alle 50 Tt (yhteensä noin 90Tt, mutta huomioitava pieni vastausprosentti – yliopistolla tarve todennäköisesti kaksinkertainen)	NAS	355 000€	5v
Lähitallennustila	Pieni levypalvelin	13 000€ / kpl	3v
Varmuuskopiointi	Varmuuskopiointi	315 000 € – 550 000€	7v

Taulukko 2. Eri ratkaisujen kustannukset

Muuta huomioitavaa

Varsinaisen fyysisen säilytyksen lisäksi tarvitaan usein myös joku sovellus tai muu rajapinta, jolla dataan pääsee käsiksi. Nämä sovellukset saattavat olla tieteenalakohtaisia, jolloin yleiskäyttöistä haku- ja metadatan kuvailusovellusta ei voida käyttää.

Kapasiteetin lisäksi tutkijat tarvitsevat datan hallintaa koskevaa ohjausta. Ohjauksen pitäisi kattaa datan säilyttämiseen liittyvät kysymykset, metadatan laatimisen ja datan käyttöön liittyvä oikeudelliset kysymykset. Jos aineistojen joukossa on erittäin korkealla turvaluokalla olevaa dataa, ei aiemmin kuvattu laitteistoympäristö riitä.

5 DATAPOLITIIKAN LÄHTÖKOHTA

Helsingin yliopisto on sitoutunut edistämään tutkimustiedon avointa saatavuutta. Tätä edesauttaa konkreettisesti rehtorin tutkimusjulkaisuja koskeva rinnakkaistallentamispäätös.

Tutkimustulosten verifioiminen edellyttää tutkimusdatan ja muiden tutkimuksen tausta-aineistojen avointa saatavuutta. Yhdessä tutkimustulokset ja niiden taustadata muodostavat yliopiston keskeisen tutkimusinfrastruktuurin, jonka merkitys kasvaa sitä mukaan, kun infrastruktuuri verkottuu yhteiskäyttöisten kansainvälisten ratkaisujen kanssa. Tämä antaa myös yliopiston tutkijoille joustavan mahdollisuuden ja välineet siirtyä sujuvasti kansainvälisiin tutkimusprojekteihin ja verkostoihin. Vastaavasti yli tieteenalojen rajojen tehtävä tutkimus saa uuden ulottuvuuden.

Datan saatavuudessa on kuitenkin otettava huomioon tietosuojakysymykset. Datan omistajuudesta ja tekijänoikeuksista voidaan ja täytyy sopia. Datan saatavuuden järjestäminen ja hyödyntäminen muissa yhteyksissä edellyttää selkeiden sopimusten ja yliopiston toimintapolitiikan olemassaoloa. Tätä tukee teknisen infrastruktuurin luominen. Yliopiston datapolitiikan laatiminen edellyttää ainakin seuraaviin kysymyksiin vastaamista:

5.1 Kuka omistaa tutkimusaineiston?

a) *yliopiston rahoittamassa tutkimuksessa*

Tekijänoikeuslain mukaan oikeudet tutkimustuloksiin ja dataan ovat lähtökohtaisesti tutkijalla tai tutkimusryhmällä. Tästä olisi nimenomaisesti sovittava jo ennen datan keräämistä. Datan omistajuus tulisi käydä ilmi myös tutkimuksen tiedonhallintasuunnitelmasta. Tutkimusdatan omistajuutta käsitellään laajasti Marjut Salokanteleen tutkimusdatan oikeudellisia reunaehtoja koskevassa raportissa.

Olisi pohdittava, haluaako yliopisto, että data olisi laajemmin/yleisesti saatavilla tietyn ajankohdan jälkeen?

b) *muussa julkisella rahoituksella tehtävässä tutkimuksessa*

Jos rahoittaja edellyttää tutkimustulosten ja/tai datan saatavuutta, yliopistolla täytyy olla mekanismit toteuttaa tämä. Tämä on jo käytäntö muun muassa kansainvälisessä lääketieteellisen tutkimuksen rahoituksessa, European Research Councilin rahoituksessa ja yhä enenevässä määrin myös EU:n muussa tutkimusrahoituksessa. Myös lukuisat kansalliset tutkimusrahoittajat edellyttävät tätä rahoituksen edellytyksenä (mm. Ison-Britannian Research Council'it). Tällöin on oltava yksiselitteinen taho, jolla on valta määrätä datan saatavuudesta.

c) *osittain yksityisellä rahoituksella tehtävässä tutkimuksessa*

Oikeudet hyödyntää tutkimustuloksia ja dataa jatkotutkimuksessa pitäisi säilyä yliopistolla.

d) *kokonaan yksityisellä rahoituksella tehtävässä tutkimuksessa*

Vaikka tutkimus tapahtuisi kokonaan yksityisellä rahoituksella ja omistus- ja hyödyntämisoikeudet tuloksiin olisivat rahoittajalla, yliopisto voi edellyttää, että ainakin tutkijan nimi mainitaan tulosten ja datan yhteydessä (moraaliset oikeudet) ja tapauksesta riippuen yliopiston pitäisi pidättää hyödyntämisoikeudet tieteelliseen tutkimukseen yliopistolla.

5.2 Miten datan saatavuus järjestetään?

Helsingin yliopistossa on luotava päätöksenprosessi seuraaville datan käyttöön ja säilyttämiseen liittyville asioille:

- Kuka päättää, mitä tallennetaan ja minne?
- Kenellä on vastuu data laadusta ja oikeuksien hankinnasta?

Yliopiston tietopääoman säilyttämisen ja hyödyntämisen kannalta on oleellista, että arvokkaat tutkimusaineistot saadaan tehokkaasti hyödynnettyä ja turvallisesti säilytettyä.

Ehdotamme, että Helsingin yliopistossa datan käyttöoikeuksista ja mahdollisesta pitkäaikaissäilytyksestä sovittaisiin seuraavasti. Lähtökohdaksi otetaan, että merkittävää dataa syntyy vasta jatkotutkimusvaiheessa, aikaisintaan väitöskirjan materiaalia kerätessä. Tällöin datan säilyttämisestä ja käytöstä neuvotellaan väitöskirjan ohjaajan kanssa, jolle ei kuitenkaan pelkän ohjauksen perusteella synny oikeuksia dataan. Tutkimushankkeissa noudatetaan seuraavaa menettelyä.

- a) tutkimusryhmätasolla sovitaan kenellä on mahdolliset tekijänoikeudet dataan. Yleensä dataan sinänsä ei synny tekijänoikeuksia vaan datatietokantoja suojataan tapauksesta riippuen joko tekijänoikeuslain mukaisen luettelosuojan tai ns. tietokantaoikeuden avulla (TekijäL 49§) sillä perusteella, että datan kerääminen on edellyttänyt suurta työpanosta.
- b) tiedekunta- tai laitostasolla luodaan päätöksentekomekanismi sille, mitä dataa ehdotetaan pitkäaikaissäilytettäväksi.
- c) Yliopiston tasolle perustetaan paneeli, joka muodostuu esimerkiksi tieteellisen neuvoston jäsenistä. Tämä paneeli päättää tiedekuntien ehdotuksesta, mikä data säilytetään ja kuinka kauan.

5.3 Miten yhteensopivuus ja yhteiskäyttö varmistetaan?

Tutkimustoiminnassa kerättävän datan on oltava mahdollisimman korkealaatuista. Tämä edellyttää datankeruun ja -hallinnan huolellista suunnittelua sekä soveltuvien kansainvälisten ja kansallisten standardien noudattamista kaikissa tutkimuksen ja datan elinkaaren vaiheissa.

Digitaalisessa muodossa olevan tutkimusdatan tulisi olla helposti löydettävissä ja vähimmillään siihen liittyvän metadatan tulisi olla saatavissa mahdollisuuksien mukaan yhteiskäyttöön muiden kansallisten ja kansainvälisten järjestelmien kanssa. Tämä edellyttää mm. kansallisesti ja kansainvälisesti yhteensopivien metadatastandardien noudattamista.

5.4 Miten meritoitumista tuetaan?

Datan julkaisuun liittyvän meritoitumisjärjestelmä edistäisi konkreettisesti tutkimusdatan käyttöä ja tämän mukaisia ehdotuksia on tehty sekä Suomessa että Suomen ulkopuolella. Myös Suomen Akatemia pitää datan käyttöön liittyvän meritointijärjestelmän toteuttamista suositeltavana.

Tekijänoikeuslaki edellyttää, että tekijän nimi on aina mainittava datan yhteydessä. Datan löytyminen ja lähteen tunnistaminen nimitietoineen edellyttää, että datassa on pysyvät tunnisteet. Pysyvissä tunnisteissa noudatetaan kansallisia, alakohtaisia tai kansainvälisiä tunnisteita.

5.5 Miten datan säilytys ja pitkäaikaissäilytys turvataan?

Arvokas tutkimusdata on säilytettävä siten, että sen käyttö tulevaisuudessa on turvattu. Tutkimusdatan säilytykselle tutkimuskäyttöön on luotava yliopistolla järkevät tekniset ja hallinnolliset ratkaisut. Tutkimustulosten verifioiminen edellyttää, että taustadata on saatavilla alasta riippuen noin 10–20 vuotta.³ Jälkipolville suuntautuvat pitkäaikaissäilytysratkaisut on koordinoitava kansallisten ja kansainvälisten pitkäaikaissäilytysratkaisujen kanssa.

³ UK:n research council'eiden soveltama aika.

5.6 Integroitu informaatioympäristö

Tutkimusdatan saatavuuden ja käytön edistämistä tukee sellaisen infrastruktuurin rakentaminen, missä tietoaineistot tyypistä riippumatta (data, metadata, informaatio, tieto jne.) muodostavat kokonaisuuden, jossa käyttäjä voi liikkua tietotarpeidensa mukaan joustavasti aineistotyypistä toiseen. Tällainen integroitu informaatioympäristö voi syntyä vain useiden toimijoiden yhteistyönä, ja se edellyttää standardointia, erilaisten järjestelmien ja välineiden yhteensopivuutta sekä yhteisiä rajapintoja. Erinomainen esimerkki tällaisesta toteutuksesta ovat NCBI:n (National Center for Biotechnology Information) rakentamat palvelut, tietokannat ja työvälineet.

Useat keskeiset tiedekustantajat (esim. Nature ja Science) edellyttävät, että artikkeliin liittyvä data asetetaan avoimesti saataville tutkimustulosten julkaisemisen yhteydessä. Tiedelehdissä julkaistuista artikkeleista löytyy jo linkkejä dataan, johon tutkimuksen tulokset perustuvat. Tutkimusdatan arvo korostuu entisestään linkittämällä siihen (tai sen metadataan) kaikki ne julkaisut, jotka kerätyn aineiston pohjalta on tuotettu.

Integroidun informaatioympäristön rakentaminen on tiedeyhteisön etujen mukaista. Integroinnin tarve ja koettu kiireellisyys voivat vaihdella tieteenaloittain. Vaikka eri toimijoiden kansallinen ja kansainvälinen yhteistyö on välttämätöntä, tarvitaan myös paikallista, yliopistotason vastuunjakoja ja organisointia. Asiassa ehdotetaan edettävän kahdella toisiaan tukevalla tavalla:

- 1) käynnistetään pilotti, jonka avulla mallinnetaan tarvittavaa kokonaisuutta; ja

- 2) yliopisto asettaa toimielimen, jossa kootaan yhteen keskeiset tutkimusdatan hallinnointiin, käytön edistämiseen ja säilyttämiseen liittyvät toimijat. Toimielimen tavoitteeksi tulisi asettaa sellaisten toimenpiteiden valmistelu ja toteutus, joilla HY

varautuu integroidun informaatioympäristön haasteisiin.
Toimielimen vetovastuuta ehdotetaan kirjastolle.

6 VÄHIMMÄISTAVOITE HELSINGIN YLIOPISTON DATAPOLITIIKALLE

6.1 Yleiset tavoitteet

Yliopiston tulisi sisällyttää datapolitiikan ja integroidun informaatioympäristön luomisen vuosien 2013–2016 strategisiin tavoitteisiin osana välttämätöntä ja toimivaa tieteen infrastruktuuria.

Ehdotamme, että Helsingin yliopistossa toteutetaan vähintään ESFRI –hankkeiden edellyttämä yhteinen datapolitiikka. Helsingin yliopisto on mukana useassa ESFRI –hankkeessa. Yliopiston on pystyttävä tarjoamaan näiden hankkeiden edellyttämät palvelut koko yliopiston tasolla.

Tämän lisäksi yliopiston on luotava edellytykset sille, että yksittäinen tutkija tai tutkimusryhmä voi ottaa kansainvälistä rahoitusta vastaan myös sellaisilta rahoittajilta, jotka edellyttävät, että tutkimustulosten pohjana oleva data on avoimesti saatavilla. Euroopassa tällaisia rahoittajia ovat muun muassa ERC, Wellcome Trust ja CERN. Yhdysvalloissa tärkein on the National Institutes of Health (NIH), mutta myös National Science Foundation on edellyttänyt vuoden 2011 alusta, että hankehakemukset sisältävät yksityiskohtaisen datanhallintasuunnitelman. Hankkeen viimeisen rahoitusosuuden maksu on tehty riippuvaiseksi tämän suunnitelman toteutumisesta. Nämä rahoittajat osoittavat sen teknisen alustan (tietokannan), mihin datan voi tallentaa. Sen sijaan yliopiston on tutkijoiden puolesta pidettävä huoli siitä, että datan saatavuuteen liittyvät juridiset kysymykset, ennen kaikkea henkilötietolainsäädäntöön ja datan omistajuuteen liittyvät seikat, on ratkaistu.

Suomessa hyväksyttiin joulukuussa opetus- ja kulttuuriministeriön

nimittämän tutkimuksen sähköisten tietoaineistojen hyödyntämisen edistämiseksi.”⁴ Maaliskuussa 2011 annettiin valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä.⁵ Lähtökohtana näissä ehdotuksissa on, että julkisin varoin tuotettu tieto tulee olla jatkotutkimuksen hyödynnettävissä. Yliopiston tulisi olla aktiivisesti mukana kansallisen tietopolitiikan toteuttamisessa myös jatkossa.

Datan avoin saatavuus luo perustan e-tieteelle, joka tulevaisuudessa on jo nyt mukana useassa eurooppalaisessa infrastruktuurihankkeessa, jotka perustuvat yhteisen sähköisen tietoympäristön luomiselle. Seuraavassa esitellään ESFRI –infrastruktuureille yhteinen datapolitiikka, jota soveltaen voimme laatia Helsingin yliopiston tarpeisiin sopivan datapolitiikan.

6.2 ESFRI infrastruktuureiden datapolitiikka

Seuraavassa esitellään ESFRI –infrastruktuurihankkeiden yhteisesti sopimat datapolitiikan periaatteet.

Datan saatavuus

Jokaisella ESFRI –projektilla on datan ja metadatan saatavuutta koskeva datapolitiikka, joka on sovittu yhdessä käyttäjäyhteisön kanssa. Tutkimusinfrastruktuurit ovat velvollisia soveltamaan ja hallinnoimaan asianomaisen tutkimuksen piiriin kuuluvaa datan saatavuutta raakadatasta harmaaseen kirjallisuuteen.

Laatu

Jokaisella ESFRI infrastruktuurilla on käyttäjäyhteisön kanssa sovittu ja toteutettu datan laadunvarmistuksen takaava politiikka, jonka pitäisi kattaa myös se, mitä lisäinformaatiota

⁴ Loppuraportti on saatavissa osoitteesta <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2011/liitteet/okm04.pdf?lang=fi>.

⁵ www.valtioneuvosto.fi/toiminta/periaatepaatokset/periaatepaatos/fi.jsp?oid=322887

kerätään ja säilytetään, jotta loppukäyttäjä voisi arvioida datan laadun sekä sen oikeellisuuden ja soveltuvuuden aiottuun käyttöön.

Jokaisen ESFRI infrastruktuurin täytyy saattaa julkiseksi datan laatua koskeva tieto laadunvarmistukseen riittävän metadatan avulla.

Pysyvyys

Jokaisella ESFRI infrastruktuurilla on käyttäjäyhteisön kanssa sovittu ja toteutettu politiikka, joka kattaa datan säilytyksen, ylläpidon ja kuraation.

Tieteellisen datan pysyvyys on repositorioiden käytön välttämätön edellytys. Datan pysyvän säilytyksen takaaminen on osa tutkimusinfrastruktuurien digitaalisten repositorioiden tavanmukaista toimintaa.

Käyttöoikeudet (tekijänoikeudet)

Jokaisella ESFRI infrastruktuurilla on käyttäjäyhteisön kanssa yhteisesti sovittu politiikka, joka takaa datan julkisen saatavuuden, mahdollisilla asianomaiselle yhteisölle ominaisilla rajoituksilla.

Jokaisen julkisin varoin rahoitetun tutkimusinfrastruktuurin täytyy myöntää yleisölle käyttöoikeus dataan ellei datan saatavuutta ole rajoitettu ennen julkaisemista tai aikaisemmalla sopimusjärjestelyllä.

Yhteiskäytettävyys

Jokaisella ESFRI infrastruktuurilla on käyttäjäyhteisön kanssa yhteisesti sovittu, avoimiin standardeihin ja niitä tukevaan implementointiin perustuva politiikka.

Digitaalisen repositorion sisällön täytyy tukeutua yhteisiin tunnisteisiin. Suositeltavaa on, että yhteiskäytettävyyttä koskeva politiikka ja mekanismit olisivat samat kaikille ESFRI infrastruktuureille.

7 DATATYÖRYHMÄN EHDOTUKSET

Helsingin yliopiston datatyöryhmä ehdottaa, että yliopistolla toteutetaan seuraavat periaatteet ja toimenpiteet tutkimusdatan käytön tehostamiseksi:

1. Datan saatavuus tutkimustulosten verifiointiksi

Tutkimusdatan pitää olla saatavilla tutkimustulosten verifiointiksi. Jos dataan ei sisälly henkilötietolainsäädännön mukaan salassa pidettävää aineistoa, data on julkisesti saatavilla.

2. Kansainväliset vaatimukset tutkimusdatan suhteen

Suuret kansainväliset tutkimusrahoittajat edellyttävät tutkimustulosten perustana olevan tutkimusdatan saatavilla oloa. Keskeiset tieteelliset aikakauskirjat edellyttävät myös sen datan saatavilla oloa, mihin julkaisuissa viitataan. (esim. Science ja Nature).

Ehdotamme, että Helsingin yliopistossa kiireisesti luodaan mekanismit, jotka pystyvät tukemaan tutkijoita heidän rahoitushakemuksissaan ja tutkimusdatan hallinnoimisen ja tallettamisen edellyttämissä toimenpiteissä (omistusoikeuksista sopiminen, metadatan laadinta, henkilötietolainsäädännön soveltaminen).

3. Tutkimusdatan omistajuus

Helsingin yliopistossa vahvistetaan datan omistajuutta koskevat

määräykset seuraavien periaatteiden mukaisesti:

- a. Lähtökohtaisesti oikeudet dataan ovat tutkijoilla tai tutkimusryhmällä. Datan omistajuudesta ja mahdollisista siihen liittyvistä tekijänoikeuksista sovitaan aina tutkimushankekohtaisesti kirjallisesti. Mikäli oikeudet ovat laitoksella, tästä on vastaavasti sovittava kirjallisesti.
- b. Jos tutkimushanke tehdään kokonaan tai osittain ulkopuolisella julkisella rahoituksella, yliopisto pidättää itsellään oikeudet hankkeessa syntyvän datan käyttöön tutkimuksessa.
- c. Jos tutkimushanke tehdään kokonaan tai osittain ulkopuolisella yksityisellä rahoituksella, yliopisto pidättää ensisijaisesti itsellään oikeudet hankkeessa syntyvän datan käyttöön tutkimuksessa. Yliopisto sitoutuu tässä tapauksessa käyttämään dataa tavalla, mikä ei vaaranna tutkimustulosten mahdollista kaupallistamista. Vain poikkeustapauksissa nämä oikeudet voivat siirtyä rahoittajalle.

4. Datan käyttöoikeudet ja datan saatavuus

Julkisin varoin toteutetuissa tutkimushankkeissa tutkimusdatan tulee olla julkisesti saatavilla tietyn ajankohdan jälkeen, mikäli dataan mahdollisesti sisältyvistä kolmansien osapuolten oikeuksista (esim. henkilötietojen suoja) ei muuta johdu.

Jos dataan sisältyy olemassa olevia tekijänoikeuden suojaamia teoksia, tutkijoiden on hankittava tarpeelliset oikeudet näiden teosten käyttöön tutkimuksessa. Samassa yhteydessä voidaan sopia teosten myöhemmästä mahdollisesta tutkimuskäytöstä.

Yliopistolla selvitetään tutkijoiden keskuudessa, mistä ajankohdasta lähtien eri aloilla olisi tarkoituksenmukaista

saattaa data julkisesti saataville ja muiden käytettäväksi. Tämä aika voisi olla esimerkiksi 2 vuotta tutkimustulosten julkaisemisesta ellei alan käytännöistä muuta johdu.

5. Integroitu informaatioympäristö

Helsingin yliopistossa luodaan integroitu informaatioympäristö, jonka puitteissa luodaan yhtenäiset rajapinnat kansainvälisten järjestelmien kanssa, olemassa olevia metadastandardeja noudatetaan ja varmistetaan aineistojen yhteiskäytettävyys. Integroitu informaatioympäristö tähtää siihen, että yliopistolla kaikki tietoaineisto (data, metadata, informaatio, julkaistut tutkimustulokset), olomuodosta riippumatta muodostavat mahdollisimman yhtenäisen kokonaisuuden ja käyttäjä voi liikkua aineistotyyppistä toiseen tietotarpeensa mukaan. Julkaistut tutkimustulokset linkittyvät tällaisessa ympäristössä saumattomasti niiden tausta-aineistoon.

Työryhmä ehdottaa, että integroidun informaatioympäristön luominen annetaan kirjaston vastuulle. Tämän raportin liitteenä on kirjaston esitys tätä koskevaksi pilottiprojektiksi.

6. Tutkimusaineistonhallintasuunnitelma

Työryhmä ehdottaa, että yliopistossa toteutetaan kampuskohtaisesti tutkimusaineistonhallintasuunnitelman laatimista koskeva koulutus. Tämä tukee myös tutkijoiden ulkopuolisen rahoituksen hakemista, koska useat tutkimusrahoittajat, mm. Suomen Akatemia, edellyttävät aineistonhallintasuunnitelmaa osana tutkimussuunnitelmaa.

7. Tekninen infrastruktuuri

Kansainvälisissä tutkimuksissa ja Helsingin yliopiston omassa selvityksessä on käynyt ilmi, että tutkijoiden toiveena on, että tutkimusdata sijaitsee mahdollisimman lähellä ja on tutkijoiden määräysvallassa vähintään siihen saakka, kun sen pohjalta

saavutetut tulokset on julkaistu. Tämän jälkeen data voidaan siirtää tutkimusalakohtaiseen kansainväliseen repositorioon tai muuhun vastaavaan paikkaan, jossa se on laajasti saatavilla.

Tutkimusdatatyöryhmä ehdottaa, että Helsingin yliopistolla toteutetaan kampuskohtainen lisälevytilan hankkimisratkaisu, jolla katetaan tutkijoiden välittömät datan tallennukseen liittyvät tarpeet.

8. Tutkimusdatan pitkäaikaissäilytys

Työryhmä pitää tarkoituksenmukaisena, että tutkimusdatan käytettävyys saadaan ensin instituutiotasolla toimivaksi eri alakohtaisten kansainvälisten käytäntöjen mukaisesti. Tämän jälkeen pohditaan, minkä tyyppistä ratkaisua datan pitkäaikaissäilytys mahdollisesti edellyttäisi kansallisella tasolla. Tässä yhteydessä on pidettävä mielessä, että tutkijoiden preferenssi on pitää data mahdollisimman lähellä tai kansainvälisessä subject based repositoriossa. Työryhmä esittää, että Helsingin yliopisto on aktiivisesti mukana kansallisessa tutkimusaineistojen pitkäaikaissäilytystä koskevassa työssä.



TUTKIMUSDATAPILOTTI
Annikki Roos/Helsingin
yliopiston kirjasto/Viikin
kampuskirjasto

10.05.2011

Liite 1 Tutkimusdatapilotin käynnistäminen Helsingin yliopistossa

1 Tausta

Helsingin yliopiston datapolitiikkatyöryhmän muistiossa ehdotetaan, että yliopisto asettaisi toimielimen, jossa kootaan yhteen keskeiset tutkimusdatan hallinnointiin, käytön edistämiseen ja säilyttämiseen liittyvät toimijat ja, että kirjastolla olisi tämän toimielimen vetovastuu. Lisäksi ehdotetaan, että tutkimusdatan problematiikkaan pureuduttaisiin käynnistämällä pilottihanke, jossa mallinnettaisiin tutkimusdataan liittyvää kokonaisuutta.

Ylikirjastonhoitaja Kaisa Sinikara on asettanut kirjaston sisäisen tutkimusdatatyöryhmän, joka aloitti toimintansa vuoden 2011 alussa. Työryhmän tehtäviksi ylikirjastonhoitaja määräsi 1) valmistella integroidun informaatioympäristön toteuttamista ja 2) määrittellä Helsingin yliopiston kirjaston tehtävät ja rooli jatkokyöskentelyä varten tutkimusdatan hallinnan kokonaisuudessa.

Kirjaston työryhmässä on edustus kaikista kampuskirjastoista ja kirjaston verkkopalveluista, puheenjohtajaksi on määrätty kampuskirjastonjohtaja Annikki Roos ja sihteeriksi tietoasiantuntija Liisa Siipilehto.

2 Pilottiprojektin toteuttaminen

2.1 Projektin tavoite

Projektin tavoitteena on

- tutkia ja selvittää niitä edellytyksiä, joilla integroitua informaatioympäristöä voidaan edistää ja tehdä ehdotuksia toimenpiteiksi, jotka mahdollistavat tämän päämäärän toteuttamisen
- käynnistää pilotti, jonka avulla selvitetään tutkimusryhmien ja yksittäisten tutkijoiden ongelmia tutkimusdatan hallinnoimisessa sekä esteitä ja kannustimia datan jakamisessa
- pyrkiä yhdessä muiden toimijoiden kanssa löytämään välittömiä ratkaisuja akuutteihin tutkimusdatan käsittelyssä ja hallinnoinnissa koettuihin tai havaittuihin ongelmiin
- tehdä ehdotus kirjaston roolista ja tehtävistä tässä kokonaisuudessa.
- koota yhteen keskeiset tutkimusdatan hallinnointiin, käytön edistämiseen ja säilyttämiseen liittyvät toimijat Helsingin yliopistossa sekä käynnistää työnjaollinen keskustelu.

Projektin tuloksena syntyy mm. suosituksia, toimenpide-ehdotuksia, käytännön ohjeita sekä linjausehdotuksia.



2.2 Projektiorganisaatio

Ylikirjastonhoitajan asettama tutkimusdatatyöryhmän on tarkoitus jatkaa projektiryhmänä ja projektia tulee johtamaan projektipäällikkö. Helsingin yliopiston kirjaston johtoryhmä toimii projektin ohjausryhmänä.

Projektiryhmä raportoi säännöllisesti kirjaston johtoryhmälle ja tekee esityksiä johtoryhmälle ja johtokunnan tai ylikirjastonhoitajan päätettäväksi. Tärkeimmistä havainnoista ja projektin etenemisestä tiedotetaan kaikkien kampuskirjastojen neuvottelukunnissa ja dekaanitapaamisissa.

Tämän lisäksi eri vaiheissa tullaan käyttämään useita eri alojen asiantuntijoita, mm. liittyen metadataan ja oikeudellisiin kysymyksiin. Kirjasto tekee läheistä yhteistyötä yliopiston sisäisten toimijoiden, mm. tietotekniikkapalvelujen kanssa sekä muiden kansallisten ja kansainvälisten toimijoiden kanssa.

Kirjasto on valmis tarkistamaan projektiorganisaatiota, mikäli se osoittautuu aiheelliseksi.

2.2.1 Projektin kesto ja kustannukset

Työryhmän työ käynnistyi tammikuussa 2011. Aktiivinen pilotointivaihe tulee olemaan noin yhdeksän kuukauden mittainen (huhti-joulukuu 2011). Tulosten raportointiin ja jatkotyön valmisteluun on varattu kaksi kuukautta vuoden 2012 alusta ja työn on arvioitu päättyvän 28.2.2012.

Tutkimusdataprojektin kustannukset ovat n. 105 000 euroa. Summa koostuu pääosin palkkamenoista ja asiantuntijapalkkioista. Kustannukset katetaan kirjaston budjetista sisäisin järjestelyin.

Kirjasto odottaa, että tämän pilottiprojektin avulla saadaan paljon lisätietoa ja ymmärrystä siitä, millaisiin toimenpiteisiin yliopistossa olisi ryhdyttävä tutkimusdatan hallinnointiin, käytön edistämiseen ja säilyttämiseen liittyvien kysymysten ratkaisemiseksi, ja millaista resursointia se seuraavalla strategiakaudella tulisi vaatimaan.

Liite 2

KUSTANNUKSET

Palvelimeen kytkettävä levytila (SAN)

Osa tallennettavasta datasta on säilöttävä erilliselle levytilalle, joka on yhden tai useamman palvelimen käytettävissä. Kokoonpanon, jonka kapasiteetti on yhteensä noin 250 teratavua (kolmannes kapasiteetista nopeilla levyillä, kaksi kolmannesta hitaammilla), hinta on 325 000€. Tämä ratkaisutyyppi sopinee parhaiten taulukon 1 kategoriassa 50–200 teratavua oleviin tarpeisiin.

Kilpailutilanne saattaa laskea kustannuksia merkittävästikin, ja mitä isommissa erissä hankinta toteutetaan, sen alhaisemmaksi gigatavun hinta muodostuu. Hinnat ovat siis arvioita, jotka perustuvat tuoreisiin hankintoihin.

Palvelimen vuotuiset kustannukset (keskisuuri laite, ylläpitotyö) ovat runsaat 5 000e. Laitteiston tyypillinen elinkaari on 5 vuotta.

Palvelimeton levytila (NAS)

Pientarpeisiin paras ratkaisu lienee tälläkin hetkellä käytössä oleva ryhmähakemistotyyppinen ratkaisu, jossa ei tarvita erillistä palvelinta ylläpidettäväksi. Tiedostot näkyvät työasemalle ns. levyjakojen kautta, ja käyttöoikeuksia voidaan säätää HY:n omille käyttäjätunnuksille.

Kokoonpano, jonka kapasiteetti on noin 196 teratavua, maksaisi noin 355 000€. Kustannukset ovat jonkin verran korkeammat kuin SAN-levytilassa, mutta säästö tulee vastaavasti työpanoksen säästymisenä sekä palvelinlaitteista.

Mittalaitteisiin kytkettävät tallennuslaitteet

Verkon suorituskyvystä tai muista syistä johtuen aktiivikäytössä oleva data saattaa olla järkevintä olla kytkettynä yksikön käyttämään mittalaitteeseen tms. Noin 14 teran kokoisen kapasiteetin tarjoava palvelin ja levypakka hitaampaa levyä maksaa noin 13 000€ / 3v. Datan aktiivikäytön jälkeen se voidaan siirtää keskitetyille palveluille.

Varmuuskopiointi

Datan turvaamiseksi se on myös varmuuskopioitava. Tässä arvioitu laitteistokokoonpano voi huolehtia noin 700 teratavun datan varmuuskopioinnista, eli riittää varsin pitkälle. Laitteiston tyypillinen elinkaari on selvästi yli viisi vuotta, eli pidempi kuin levyjärjestelmien.

Kokoonpano, jossa on noin 650 nauhapaikkaa ja useita LTO5-nauhureita, maksaa laitteet, nauhat ja lisenssikulut huomioiden noin 315 000€ (tuki viidelle vuodelle). Tämä sisältää kuitenkin varmistuksen vain yhteen kiinteistöön: mikäli halutaan vähentää riskejä, on hankittava toiseen konesaliin toinen nauhakirjasto, jonka kustannukset ovat hieman vähäisemmät kuin ensimmäisen, arvio on 236 000€.