



YHTEISKUNTATIETEELLINEN  
TIETOARKISTO



SAMI BORG JA ARJA KUULA

## Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari

VALMISTELURAPORTTI OECD:N DATASUOSITUKSEN  
TOIMEENPANOMAHDOLLISUUKSISTA SUOMESSA

YHTEISKUNTATIETEELLISEN TIETOARKISTON JULKAISUJA 6, 2007

Yhteiskuntatieteellisen tietoarkiston julkaisuja 6

Sami Borg ja Arja Kuula (2007).

Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari.

Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa.

Julkaisija: Yhteiskuntatieteellinen tietoarkisto (FSD)  
Tampereen yliopisto

Käyntiosoite: Åkerlundinkatu 2 A, 5. krs  
Tampere

Postiosoite: Yhteiskuntatieteellinen tietoarkisto  
33014 Tampereen yliopisto

Puhelin: (03) 3551 8519

Faksi: (03) 3551 8520

Sähköposti: fsd@uta.fi

Kotisivu: <http://www.fsd.uta.fi>

Myynti: Tiedekirjakauppa TAJU  
PL 617, 33014 Tampereen yliopisto

Puhelin: (03) 3551 6055

Faksi: (03) 3551 7685

[taju@uta.fi](mailto:taju@uta.fi)

<http://granum.uta.fi>

ISSN 1459-8906

ISBN 978-951-44-6999-2

Graafinen suunnittelu  
Vinjetti Ky

Taitto  
Marita Alanko

Tampere 2007

## Esipuhe

Raportissa selvitetään hiljattain valmistuneen OECD-datasuosituksen sisältöä sekä pohjustetaan sen kansallista toimeenpanoa. Julkisrahoitteisen tutkimusdatan avoimuutta edistävää suositusta koskeva selvitystyö ja sitä varten kootut tutkimusaineistot kustannettiin hankemäärärahasta, jonka opetusministeriö myönsi vuonna 2006 Yhteiskuntatieteelliselle tietoarkistolle (FSD).

Molemmat kirjoittajat ovat työskennelleet tietoarkistossa sen toiminnan alusta, vuodesta 1999 lähtien. Sami Borg on valtakunnallisen palveluyksikön johtaja ja arkistonhoitaja Arja Kuula vastaa tietoarkiston tutkimuseettisestä tietopalvelusta ja on nykyisen Tutkimuseettisen neuvottelukunnan jäsen.

Sähköisten tutkimusaineistojen elinkaarta ja jatkokäyttöä koskevat kysymykset ovat Suomessa melko uusia ja alan suomenkielinen termistö on osin vakiintumatonta. Tietoarkistotyössä olemme pyrkineet edistämään alan suomenkielisen termistön kehittymistä.

Selvityksessä hahmotetaan tutkimusaineistojen käytön ja avoimuuteen suhtautumisen nykytilaa erityisesti ihmistieteissä. Esimerkkejä kuvataan myös muilta tieteenaloilta. Raportin päätavoitteena on jäsentää alan laajaa toimintakenttää ja avata keskustelua avoimuutta tukevista toimista.

Verkkokyselynä koottu selvityksen pääaineisto on avoimesti saatavilla tietoarkistosta tieteiliseen tutkimus- ja opetuskäyttöön.

Kiitämme raportin kuvioiden laatimisesta Pentti Kiljusta Yhdyskuntatutkimus Oy:stä, tekstin tarkistamisesta Laura Peuteretta ja Sanni Haverista sekä taitosta Marita Alankoa.

Tampereella 15.5.2007

Sami Borg ja Arja Kuula

## SISÄLLYS

1	Access to Research Data from Public Funding: OECD-prosessi ja hankkeen tausta	7
1.1	Pariisin julistukseen liittyvän datasuosituksen pääsisältö ja -tavoitteet	8
1.2	Mitä aineistoja suositus koskee ja mitä se ei koske?	9
1.3	Raportin päätavoitteet	10
2	Tutkimusdatan elinkaari ja datasuosituksen periaatteet	12
2.1	Tutkimussuunnitelma	13
2.2	Eettinen ennakoarviointi	14
2.3	Rahoitushakemus, tutkimusetiikka ja tutkimusaineistot	15
2.4	Rahoittajien datapolitiikka ja rahoituspäätökset	16
2.5	Datan keruu	17
2.6	Ensikäyttö	19
2.7	Datan arkistointi ja julkaiseminen	20
2.8	Jatkokäytön tuki	21
2.9	Jatkokäyttö	22
2.10	Julkaisu- ja data-tietokannat	23
3.	Tutkimusdatan säilytys ja avoimuuden esteet	24
3.1	Aineistojen säilyttämistä määrittävä lainsäädäntö ja yliopistojen käytännöt	24
3.1.1	Lainsäädäntö ja tekijänoikeudet	24
3.1.2	Tutkijoiden suhtautuminen professorikyselyn perusteella	26
3.2	Keskeisimmät avoimuuden esteet ja haitat	29
3.2.1	Taitamattomuus aineiston käytössä ja aineistojen soveltumattomuus uusiin tutkimusongelmiin	30
3.2.2	Epäselyydet aineiston omistajuudesta	31
3.2.3	Akateeminen kilpailu	33
3.2.4	Aineistojen tietotekniset ja dokumentoinnin puutteet	34
3.2.5	Tutkimuseettiset tietosuojakysymykset	35
4.	Aineistojen jatkokäytön nykytilanne	38
4.1	Aineistojen käyttö ja käytettävyys	38
4.2	Tutkijoiden suhtautuminen avoimuuden hyötyihin	40
4.3	Tutkijoiden suhtautuminen jatkokäyttömahdollisuuksien lisäämiseen	42
4.4	Esimerkkejä avoimuuden lisäämisestä ja jatkokäytön edistämisestä	44
4.4.1	Kouluterveystutkimus	45
4.4.2	Kansallinen terveystutkimus	45
4.4.3	European Social Survey	46
4.4.4	Luxembourg Income Study	47

5.	Tutkimusaineistojen avoimuuden ja datapolitiikan vaihtoehtoja	48
5.1	Aineiston jatkokäyttö tutkijan tai tutkimusyksikön hallinnoimana	48
5.2	Data-arkistot	49
5.2.1	Käyttöluvat data-arkistoissa	51
5.2.2	Käyttöehdot data-arkistoissa	52
5.2.3	Tietosuoja data-arkistoissa	52
5.3	Datakeskukset (data enclaves)	53
5.4	Kohti periaatteellista avoimuutta: tieteellisten julkaisujen ja rahoittajien datapolitiikat	54
5.4.1	Tieteellisten julkaisujen datapolitiikat	55
5.4.2	Tutkimusrahoittajien datapolitiikat	57
5.4.2.1	National Institutes of Health (NIH)	57
5.4.2.2	National Science Foundation	58
5.4.2.3	Medical Research Council/UK	59
5.4.2.4	Economic and Social Research Council/UK	60
6.	Suhtautuminen OECD:n datasuositukseseen ja sen kansallisen toimeenpanon toteuttamiseen	61
6.1	Tutkijoiden yleisiä näkemyksiä datasuosituksesta	61
6.2	Yliopistojen, tutkimusrahoittajien ja valtion tutkimuslaitosten mielipiteitä OECD:n datajulistuksesta ja -suosituksesta	62
6.3	OECD:n datasuosituksen toimeenpano: mielipiteitä keinoista ja ensisijaisista vastuutahoista	64
7	Näkökulmia OECD:n datasuosituksen toimeenpanoon	67
7.1	Yhteenvedo tutkimusdatan elinkaaren kannalta keskeisistä tekijöistä	67
7.2	Yhteenvedo raporttia varten kootuista kyselyistä	70
7.2.1	Kulttuurin ja yhteiskunnan tutkijoiden suhtautuminen tutkimusdatan avoimuuden esteisiin ja haittoihin	70
7.2.2	Tutkimusdatan jatkokäytön nykytilanne yliopistoissa	70
7.2.3	Kulttuurin ja yhteiskunnan tutkijoiden suhtautuminen tutkimusdatan jatkokäyttöön ja avoimuuden hyötyihin	71
7.2.4	Kooste avoimuuden järjestämistä ja toimintavastuita koskevista mielipiteistä	71
7.3	Yhteenvedo aineistojen säilyttämiskäytännöistä yliopistoissa ja tutkimusorganisaatioissa	72
7.4	OECD:n datasuosituksen periaatelista	73
7.5	Miten tästä eteenpäin?	76
7.5.1	Toimeenpanon hyötyjä	76
7.5.2	Open Access: julkaisut verrattuna tutkimusdataan	77
7.5.3	Laaja-alainen yhteistyö datasuosituksen toimeenpanon tueksi	78

Liite 1.	OECD:n Pariisin julistus 2004 ja Datasuositus 2007 (OECD Principles and Guidelines for Access to Research Data from Public Funding)	80
Liite 2.	Kuvaus selvityshankkeen aineistonkeruusta vuonna 2006	83
Liite 3.	Tietoja valtion tutkimuslaitosten alustavasta suhtautumisesta OECD:n Pariisin julistukseen ja tutkimusaineistojen avoimesta saatavuudesta 2005	86
Liite 4.	Aineistonkuvailun tietoelementit	99
	Kirjallisuus ja verkkolähteet	101

# 1 Access to Research Data from Public Funding: OECD-prosessi ja hankkeen tausta

OECD:n tiede- ja teknologiaministerit hyväksyivät tammikuussa 2004 Pariisissa pidetyssä ministerikokouksessa julistuksen julkisrahoitteisten sähköisten tutkimusaineistojen käytön avoimuudesta. Julkilausuma oli osa loppukommunikointia, jossa järjestön tiede- ja teknologiapolitiikan komitea (Committee for Scientific and Technological Policy, CSTP) otti kantaa jäsenmaiden tiede- ja innovaatiopolitiikan keskeisiin haasteisiin. Julistus tarkoitti poliittista sitoutumista julkisrahoitteisten sähköisten tutkimusaineistojen avoimen saatavuuden kehittämiseen.

*Declaration on Access to Research Data from Public Funding* (liite 1) on merkittävä askel pitkään jatkuneessa prosessissa, jolla OECD on edistänyt tieteen avoimuutta ja tutkimusaineistojen tehokasta käyttöä. OECD:n aktiivisuus asiassa on ollut ymmärrettävää myös siksi, että järjestö hyödyntää itse omassa toiminnassaan runsaasti erilaisia tilasto- ja tutkimustietoja. Eri maita ja aihealueita koskevien tietojen vaikea saatavuus sekä niiden heikko vertailukelpoisuus lienevät yksi selitys OECD:n toimille asian hyväksi. Prosessin yleisenä lähtökohtana on ollut järjestön jäsenmaiden taloudellisen kehityksen tukeminen dynaamisella tiede- ja innovaatiopolitiikalla.

Pariisin julistusta edelsi vuosia kestänyt pohjustustyö, jonka aikana syntyi useita valmisteluasiakirjoja ja julkaisuja (esim. Arzberger et al. 2004a ja 2004b). Suomesta hankkeen työryhmäasiantuntijana toimi tuolloin Leif Laaksonen CSC:stä. Julistuksen jälkeen tiede- ja teknologiapolitiikan komitea asetti hankkeen jatkotoimille työryhmän. Sille annettiin tehtäväksi työstää suositus julistuksen periaatteiden soveltamiseksi OECD:n jäsenmaissa. Jatkotyössä Suomesta työryhmän jäsenenä toimi Sami Borg.

OECD toteutti prosessin loppuvaiheessa kaksi kansallista kuulemiskierrosta suosituksen valmistelemiseksi. Niissä jäsenmaiden työryhmäedustajat kokosivat oman maansa keskeisten tutkimusorganisaatioiden alustavia mielipiteitä Pariisin julistuksesta ja siihen liittyvän suositusluonnoksen sovellettavuudesta. Suomessa ensimmäinen mielipidetiedustelu lähetettiin keväällä 2005 valtion tutkimuslaitoksille.

Lisäksi vuoden 2006 keväällä vapaaehtoisiksi ilmoittautuneissa jäsenmaissa kartoitettiin mielipiteitä julistuksesta ja silloisesta suositusluonnoksesta. Suomesta kuulemiskierrokselle osallistuivat Valtion teknillinen tutkimuskeskus (VTT), Kansanterveyslaitos (KTL) sekä Yhteiskuntatieteellinen tietoarkisto (FSD).

Kolmannen kerran mielipiteitä kerättiin Suomessa nyt käsillä olevaa raporttia varten loppuvuodesta 2006. Kartoituksella selvitettiin kulttuurin ja yhteiskunnan tutkimusta (KY-alat) edustavien tutkijoiden sekä yliopisto-organisaatioiden ja tutkimusrahoittajien näkemyksiä. Tutkijakyselyn tulosten esittely muodostaa keskeisen osan tätä raporttia. Kolmannen keruuaallon aineistojen sisältö ja keruutavat selostetaan tarkemmin liitteissä 2 ja 3.

Pariisin julistuksen toimeenpanoa ohjeistava lopullinen datasuositus julkaistiin OECD:n verkkosivuilla huhtikuussa 2006. Suositus on raportin liitteenä 1.

## 1.1 Pariisin julistukseen liittyvän datasuosituksen pääsisältö ja -tavoitteet

OECD:n datasuosituksessa korostetaan julkisrahoitteisen sähköisen tutkimusdatan avoimen saatavuuden ja erityisesti uusien, Internet-pohjaisten käyttöympäristöjen etuja. Suosituksen mukaan tutkimusdatan avoin saatavuus (*accessibility*) on olennaista julkisrahoitteisen tutkimuksen tehokkaalle ja taloudelliselle toteuttamiselle, vahvojen innovaatioketjujen luonnille sekä tulokselliselle kansainväliselle yhteistyölle.

Suositus tähtää sähköisten tutkimusaineistojen jatkokäyttömahdollisuuksien kohentamiseen sen jälkeen, kun aineistojen alkuperäiset kokoajat eivät enää itse hyödynnä niitä aktiivisesti. Suuret yhteisrahoitteiset ja usein kansainväliset tutkimusaineistoinfrastruktuurit voivat tosin mahdollistaa avoimen käytön jo heti aineistojen keruun jälkeen.

Suosituksen mukaan tutkimusaineistojen avoin saatavuus ja tutkimusaineistojen jakaminen avoimeen käyttöön edistää merkittävästi

- tieteen avoimuutta
- vaihtoehtoisia tutkimusasetelmia
- tiedonkeruun ja tutkimusmenetelmien tutkimusta
- tutkijakoulutusta
- uusien, aineistojen kerääjiltä tutkimatta jääneiden aiheiden tutkimusta
- erilaisten aineistojen ja tietojen liittämistä toisiinsa.

Datasuositus tukee periaatteellisella tasolla

- tutkimusaineistojen avoimuutta tukevia toimintakulttuureja
- tutkimusaineistojen avoimuutta edistäviä hyviä käytäntöjä
- tutkimusaineistojen avoimuuden hyötyjen ja haittojen avointa tiedottamista
- jäsenmaiden tiedepoliittisia toimenpiteitä tutkimusaineistojen avoimuuden edistämiseksi
- suosituksia, joilla tuetaan tutkimusaineistojen avoimuutta kansainvälisessä toimintaympäristössä.

Avoimuuden hyödyt liittyvät monin tavoin myös tieteellisen tutkimuksen laadunvarmistukseen. Tutkimusdatan avoimuus tukee tieteen avoimuuden sekä tutkimustulosten toistettavuuden ja kontrolloitavuuden periaatteita.

Tutkijoiden ja tutkimustulosten hyödyntäjien tietotarpeet ovat muuttuneet ajan mittaan. Kansainvälistyminen on lisännyt tieteelliseen näyttöön perustuvan vertailutiedon tarvetta ja samalla vaatimuksia tutkimusdatan avoimuudelle. Etenkin monilla teknisillä ja luonnontieteellisillä tieteenaloilla haasteeseen on pyritty vastaamaan luomalla toimintamalleja, joilla on tuettu datan avointa yhteiskäyttöä.

Tutkimusdatan avoimuutta on tuettu myös luomalla tutkimusdataa säilyttäviä, dokumentoivia ja välittäviä erityisorganisaatioita, jotka toimivat luotettuna kolmantena osapuolena aineiston kokoajien ja jatkokäyttäjien välissä. Suomessa tällainen on Yhteiskuntatieteellinen tietoarkisto (FSD). Se on toiminut valtakunnallisena tutkimusinfrastruktuurina Tampereen yliopiston yhteydessä vuodesta 1999 alkaen.



## 1.2 Mitä aineistoja suositus koskee ja mitä se ei koske?

OECD:n datasuositus koskee erityisesti *julkisrahoitteisesti koottuja sähköisiä data-aineistoja, jotka on jo alun perin koottu tutkimuskäyttöön*. Sähköinen tutkimusaineisto eli (tutkimus)data tarkoittaa suosituksessa ja tässä raportissa *digitaaliseen muotoon tallennettua tietokokoelmaa, joka voi sisältää numeroita, tekstejä, kuvia, ääntä tai näihin liittyviä kuvailutietoja*.

Määrittely rajaa suosituksen ulkopuolelle suuren joukon julkisen sektorin tutkimustoi-minnassa syntyviä sähköisiä aineistoja. Esimerkiksi sähköiset julkaisut tai useimmat Internetin aineistot eivät kuulu suosituksen piiriin. Myöskään tutkimuksissa syntyvät alustavat tulostiedot, tutkimusten suunnitteluun tai arviointiin liittyvät asiakirjat, laboratorionäytteet tms. ei-digitaaliset havaintoaineistot eivät kuulu suosituksen tarkoittamaan dataan.

Suosituksen piiriin eivät myöskään kuulu yksityisten yhtiöiden ja yritysten omistamat aineistot. Lisäksi kansalliseen turvallisuuteen, yksityisyyden suojaan, immateriaalioikeuksiin tai tietojen muuhun luottamuksellisuuteen liittyvät syyt saattavat olla perusteena sille, että tutkimusaineisto jää avoimen käytön ulkopuolelle. Suosituksen periaatteet ja tavoitteet voivat silti soveltua myös yksityisrahoitteisiin tutkimusaineistoihin.

Tiivistetysti datajulistuksen ja -suosituksen sisältö on seuraava:

### *Mitä tutkimusdataa suositus koskee ja miten?*

- Ensisijaisesti julkisrahoitteisia, *sähköiseen* muotoon tallennettuja, tutkimustulosten pohjaksi ja validoimiseksi koottuja tutkimusaineistoja.
- Sähköinen tutkimusdata voi koostua numero-, teksti-, kuva- tai äänitallenteista.
- Tutkimusdata on usein tutkimuksen ns. pohja- tai perusaineisto.
- Suosituksen piiriin kuuluu suuri osa yliopistoissa, valtion ja muun julkisen sektorin tutkimuslaitoksissa sekä muiden julkisen sektorin tiedontuottajien piirissä kerättävistä tai näiden tahojen hallinnoimista tutkimusaineistoista.
- Suosituksen periaatteet ovat useissa tapauksissa sovellettavissa muihinkin kuin julkisrahoitteisesti koottuihin sähköisiin tutkimusaineistoihin.

### *Mitä suositus EI koske?*

- Ei koske julkaisuja tai käsikirjoituksia.
- Ei koske yksityissektorin kokonaan tai osittain kustantamia tutkimusdataa tai tutkimusta, jossa yksityinen taho saa tutkimustulokset omistukseensa.
- Ei koske aineistoja, joiden avoimuutta täytyy rajoittaa luottamuksellisuuden tai salassapidon, kaupallisten intressien, tekijänoikeuksien, kansallisen turvallisuuden tai muun erityisen synn vuoksi.
- Ei koske monia tutkimusprosessin aikana kertyviä muistiinpanoja tai alustavia analyysejä eikä fyysisiä tutkimusnäytteitä tai -kohteita (esim. laboratorionäytteet tai koe-eläimet).

### *Mitä avoin saatavuus (Open Access) tarkoittaa sähköisten tutkimusaineistojen osalta?*

- Keskeisten julkisrahoitteisten tutkimusdatajien tulisi lähtökohtaisesti olla avoimessa käytössä.
- Mahdollisimman pienin kokonaiskustannuksin turvataan mahdollisimman suurelle käyttäjäkunnalle esteetön ja tasavertainen tutkimusdatan käyttömahdollisuus.

- Tyypillisesti kyse on tutkimusdatan kerääjän ensikäytön tai muun aktiivikäytön jälkeisestä jatkokäytöstä.
- Suurissa aineistoinfrastruktuureissa kyse voi olla myös ensikäyttöön verrattavasta käyttömahdollisuudesta, joka perustuu usein laajaan yhteisrahoitukseen ja -käyttöön.
- Tutkimusdatojen Open Access ei aina tarkoita täysin avointa ja maksutonta tutkimusdatan (jatko)käyttöä.
- Avoimen saatavuuden lisäksi suosituksessa kiinnitetään paljon huomiota tutkimusdatan informoidun jatkokäytön turvaamiseen (asianmukaiseen ja riittävään tietoon perustuvaan käyttöön).
- Tutkimusdatan avoimen saatavuuden lisääminen edellyttää usein tutkimusaineistojen dokumentointia ja muokkausta, joilla aineisto saatetaan käyttökuntoon jatkokäyttöä varten.
- Jatkokäyttö voi sisältää käyttörajoituksia ja avoimenakin edellyttää erilaisia käyttöluhia, käyttöehtositoumuksia ja käyttäjätunnistusta.
- Aineistojen saatavuutta voidaan edistää tehokkaasti tietoverkkojen avulla.

#### *Data access regime*

- Suositus koskee kaikkia tutkimus- ja tieteenaloja sekä hyvin erilaisia tutkimusorganisaatioita. Se jättää tilaa kansalliselle harkinnalle periaatteiden soveltamisessa.
- Käsitteellä data access regime (DAR) viitataan laajasti tutkimusdatan elinkaaren hallintaa tukeviin toimiin, joiden avulla julkinen sektori voi edistää monipuolisesti rahoittamansa tutkimusdatan optimaalista käyttöä.
- DAR-keinovalikoima koostuu säännöistä sekä toimintaohjeista tai suosituksista, joilla tutkimusorganisaatiot, tutkimusrahoittajat sekä muut asianomistajat määrittelevät julkisrahoitteisen tutkimusdatan elinkaarta sekä käytön ja saatavuuden ehtoja.
- Tutkimusorganisaatioille tyypillisiä DAR-toimintoja ovat virallisten datapolitiikkojen laatiminen (data policy) sekä tutkimusdatan avointa saatavuutta tukevat erityissuositukset ja -ohjeet.
- Tutkimusrahoitukseen voidaan liittää suosituksia ja ehtoja, jotka tukevat tutkimusdatan jatkokäyttöä ja avointa saatavuutta.

### 1.3 Raportin päätavoitteet

Datasuositus jättää paljon tilaa kansalliselle keskustelulle ja periaatteiden ja tavoitteiden toimeenpanoa on hyvä pohtia riittävän laaja-alaisesti.

Raportissa käsitellään datasuosituksen keskeisiä aihepiirejä, mutta suosituksen rakenteessa tai yksityiskohdissa ei pitäydytä tarkasti. Julistus ja siihen liittyvä OECD-suositus ovat verraten yleisluonteisia. Tutkimusdatojen avoimuutta on tässä raportissa haluttu kuitenkin käsitellä sekä laajemmin että täsmällisemmin. Se on ollut mahdollista rajautumalla ihmistieteisiin, lähinnä kulttuurin ja yhteiskunnan tutkimuksen aloille.

Raportti on laadittu ennen kaikkea palvelemaan aihetta koskevaa tiedepoliittista keskustelua. Tätä silmällä pitäen raportin alkuosassa on käsitelty avoimuuteen ja saatavuuteen liittyviä kysymyksiä sähköisten tutkimusaineistojen koko elinkaaren näkökulmasta Suomessa.

Datasuosituksen aihepiiri ja peruskäsitteet eivät ole Suomessa laajasti tunnettuja. Siksi raportissa kuvataan koti- ja ulkomaisia esimerkkejä sähköisten tutkimusaineistojen elinkaaren hallinnan toimintamalleista. Lisäksi työ esittelee tutkimustuloksia tutkijoiden, yliopistojen, muiden tutkimusorganisaatioiden ja tutkimusrahoittajien suhtautumisesta suosituksessa esiin nostettuihin kysymyksiin.

Raportti on siis nykyisiä käytäntöjä, toimintoja ja vallitsevaa suhtautumista esittelevä *yleiskuvaus ja keskustelunavaus*. Joissakin osuuksissa painottuu akateeminen yliopistotutkimus, koska hankkeen toimeksiantajana on opetusministeriö. OECD-suosituksen kansallinen toimenpano ei kuitenkaan ole yksin opetusministeriön vastuulla, vaan asia kuuluu myös OECD-toiminnassa keskeisesti mukana oleville muille viranomaisille ja toimialoille.

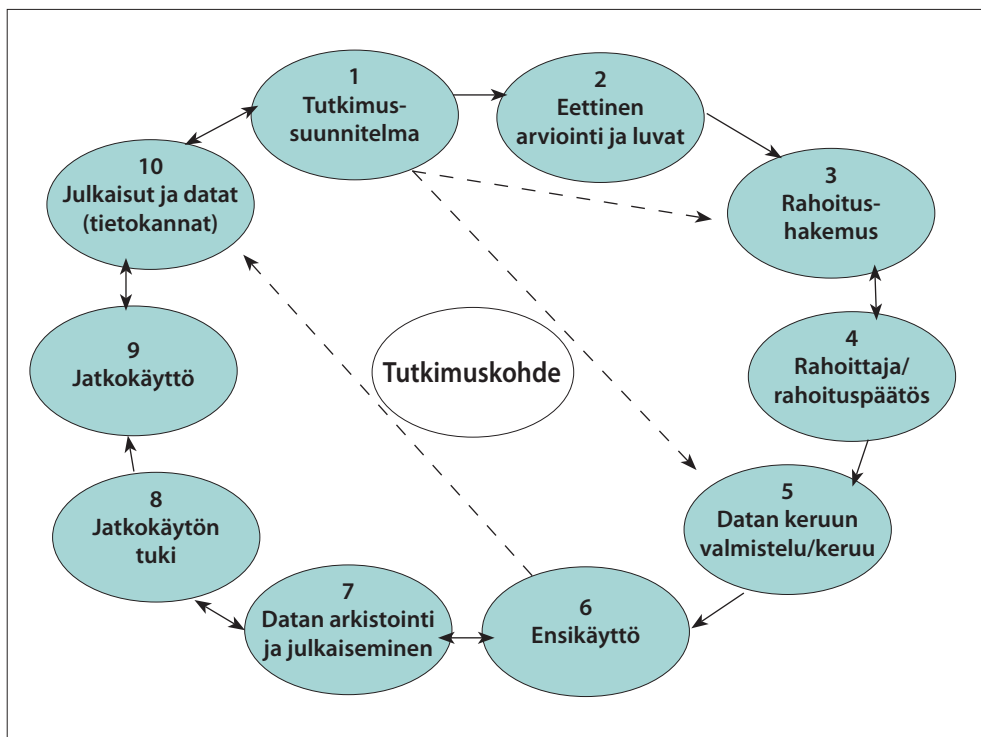
Datasuositus koskee kaikkia julkisen sektorin tiedontuottajia ja tutkimusorganisaatioita. Siksi raportissa on malleja ja esimerkkejä eri tutkimusaloilta ja -organisaatioista. Kirjoittajien tietoarkistotausta ja yhteiskuntatieteellinen koulutus näkyvät eri tavoin aihevalinnoissa ja näkökulmissa. Toivottavasti raportti tarjoaa kuitenkin mahdollisimman monille tieteenaloille, tutkimusorganisaatiolle ja tiedontuottajille aineksia oman alan datakeskusteluun.

## 2 Tutkimusdatan elinkaari ja datasuosituksen periaatteet

Pariisin julistuksessa ja OECD:n datasuosituksessa (liite 1) ei tarkastella tutkimusaineistojen elinkaarikysymyksiä systemaattisesti. Molemmissa asiakirjoissa painopiste on tutkimusdatan saatavuutta edistävissä yleisissä periaatteissa. Julistuksessa ratkaisu on perusteltu, mutta sovel-lusohjeena sisällön jäsentelytapa ei ehkä toimi parhaalla mahdollisella tavalla. Kaikkiaan suosituks jää epäkonkreettiseksi, vaikka otsikkonsa mukaan sen pitäisi ohjeistaa periaatteita käytännöiksi (*OECD Principles and Guidelines for Access to Research Data from Public Funding*).

Seuraavaksi pureudutaan systemaattisemmin siihen, mitkä tahot, asiat ja toiminnot liittyvät sähköisen tutkimusdatan elinkaareen ja miten ne kytkeytyvät toisiinsa. Sen avulla aihepiiriin vihkiytymättömät voivat hahmottaa tutkimusdatan saatavuuden ja jatkokäytön kokonais-kuvaa.

Tarkastelu on pelkistävä eikä se kuvaa tyhjentävästi eri tutkimusalojen julkisrahoitteisten sähköisten tutkimusaineistojen erityispiirteitä. Varsinkin silloin kun tiedonhankinnan koh-teina ovat ihmiset, tutkimusaineistojen avoimen saatavuuden turvaamisen kysymykset muo-dostuvat monista eri aineksista. Tämä ilmenee kaaviosta 2.1, jossa tarkastellaan tutkimusai-neiston elinkaaren päävaiheita tavanomaisessa tutkimusprosessissa (vrt. datan elinkaarimalli DDI Alliance 2004).



Kaavio 2.1

Julkisrahoitteisen tutkimusdatan elinkaari tutkimussuunnitelmasta aineiston jatkokäyttöön

## 2.1 Tutkimussuunnitelma

Kaikki tutkimus tähtää uuden tiedon tuottamiseen. Tieteellisessä toiminnassa tutkimusta ohjaa tutkimussuunnitelma, jossa esitellään muun muassa tutkimuksen tavoitteet, aineistot ja menetelmät. Tieteellisessä tutkimussuunnitelmassa jäsennetään tutkimuksen kohde, teoreettiset pyrkimykset sekä käytettävät aineistot ja tutkimusmenetelmät nimenomaan uuden tiedon tuottamisen näkökulmasta.

Uuden tiedon tuottaminen on mahdollista ainoastaan perehtymällä uuden tutkimusongelman kannalta oleellisiin tutkimusjulkaisuihin ja –aineistoihin. Yleensä nämä löytyvät erilaisten julkaisu-, hanke- ja datatietokantojen avulla. Tieteellisiin julkaisutietokantoihin kuvailut julkaisut ovat tavallisesti käyneet läpi tieteellisen arviointiprosessin, minkä vuoksi julkaisut ovat ensisijaisia tieteellisen tiedon lähteitä.

Pelkästään tutkimusjulkaisuihin perehtyminen ei kuitenkaan tuo kattavasti esille kaikkea uuden tutkimuksen kannalta merkityksellistä ja hyödynnettävissä olevaa tietoa. Sekä tiedot aihepiiristä käynnissä olevista tutkimushankkeista että aiheesta saatavilla olevista, relevanteista tutkimusaineistoista ovat tärkeitä uudelle tutkimukselle. Mitä paremmin tällaiset tiedot ja jo tuotetut resurssit ovat tutkijoiden käytettävissä, sitä varmemmin ja tehokkaammin tieteellinen tutkimus onnistuu päätehtävässään.

Tutkimusaineistojen elinkaari voidaan usein hahmotella melko pitkälle jo tutkimussuunnitelman laatimisen aikaan. Käytännöt tosin vaihtelevat siinä, kuinka yksityiskohtaisesti tiedonkeruuvälineet ja tutkimuksen toteuttamiseen ehkä tarvittavat lupa- ja sopimusasiakirjat on suunniteltava tai toteutettava tutkimuksen alkuvaiheessa. Mitä keskeisemmästä tutkimusrahoittajasta ja suuremmasta tutkimushankkeesta on kyse, sitä valmiimpia kaikkien tutkimuksen edellytysten ja suunnitelmien – myös esimerkiksi aineistonkeruun instrumenttien ja lupien – tulee olla.

Julkisrahoitteisen tutkimusdatan elinkaaren kannalta tutkimussuunnitelma on erittäin keskeinen asiakirja, koska läheskään aina tutkimuksia varten ei laadita erillisiä tutkimussopimuksia, kuten sopimus- tai tilaustutkimuksissa.<sup>1</sup> Etenkin yliopistojen julkisrahoitteisessa tutkimuksessa tutkimussuunnitelma voi olla ainoa asiakirja, jossa esitetään (jos esitetään) tutkimusaineiston käyttöä koskevat perustavat tiedot ja suunnitelmat. Tällaisia keskeisiä tietoja voivat olla ainakin

- tutkimuskohteita koskevat informoinnit sekä suostumus-, lupa- ja sopimusasiakirjat
- tutkimusaineistoon liittyvät omistus-, hallinta- ja tekijänoikeudet, ainakin jos ne poikkeavat tutkijaryhmän tutkimusorganisaation normaalikäytännöistä
- suunnitelmat tutkimusaineiston ensikäyttövaiheen säilyttämisestä, käyttötavoista ja käyttöoikeuksista
- suunnitelmat tutkimusaineiston jatkokäytön mahdollistamisesta ensikäytön jälkeen.

<sup>1</sup> Tilaus- ja sopimustutkimukset ovat useimmiten ainakin osaksi yksityisrahoitteisia, eikä niihin siksi puututa tarkemmin tässä julkisrahoitteisiin sähköisiin tutkimusaineistoihin keskittyvässä raportissa.

Suotavinta olisi, että sähköisten tutkimusaineistojen elinkaari otettaisiin huomioon nykyistä paremmin jo tutkimussuunnitelmassa. Tämä tukee täsmällistä ja tarkoituksenmukaista uuden tiedon keruuta sekä auttaa tiedeyhteisöä hyödyntämään uusia tutkimusaineistoja mahdollisimman tehokkaasti ja pitkäaikaisesti.

## 2.2 Eettinen ennakoarviointi

Suomessa ei ole institutionalisoitua yleistä eettistä ennakoarviointia ihmistieteiden tutkimukselle. Esimerkiksi Jyväskylän yliopistossa voi kuitenkin pyytää lausuntoa eettiseltä toimikunnalta, jos tutkimuksen toteuttaminen voi aiheuttaa tutkittaville fyysisiä, psyykkisiä tai oikeusturvaan liittyviä haittoja. Jyväskylän yliopistossa käytäntönä on, että lausuntoa haetaan kokeellisiin tutkimusasetelmiin. Tämä korostuu liikuntatieteellisissä ja psykologisissa tutkimuksissa.

Laajimmin ihmistieteiden eettinen ennakoarviointi on käytössä Yhdysvalloissa, Kanadassa ja Australiassa. Näissä maissa tutkimuslaitoksen eettinen elin arvioi, onko suunnitellussa tutkimuksessa huomioitu tutkimuseettiset kysymykset ja tutkittavan oikeudet asianmukaisella tavalla. Esimerkiksi Kanadassa tutkimusaihe ei määritä ennakoarvioinnin tarvetta, vaan ainoastaan se, ovatko tutkimuksen kohteena ihmiset. Siten Kanadassa arvioitaisiin ennakkoon esimerkiksi kirjaston tietoteknisten palveluiden toimivuutta koskeva kysely.

Norjassa arvioidaan ennakkoon tutkittavien informointi, koska siellä yksityisyyden suoja on katsottu keskeisimmäksi tutkimuseettiseksi riskiksi ihmistieteissä. Ruotsissa eettinen ennakoarviointi tarvitaan, jos tutkitaan arkaluonteisia tietoja ilman tutkittavan nimenomaista suostumusta tai tutkimusmenetelmä on sellainen, että se puuttuu ihmisen fyysiseen ja psyykkiseen koskemattomuuteen. Tällä viitataan lähinnä rekisteritutkimuksiin ja psykologisiin kokeellisiin tutkimuksiin.

Suomessa tutkimussuunnitelmien eettinen ennakoarviointi on lakiin perustuva velvoite vain, jos kyseessä on ihmisiin kohdistuva lääketieteellinen tutkimus tai tieteellinen eläinkoe. Arviointi on yleensä tarkkaa ja siksi esimerkiksi lääketieteellisen tutkimuksen suunnitelmia ohjataan verraten yksityiskohtaisesti ja määrämuotoisesti. Laki lääketieteellisestä tutkimuksesta velvoittaa, että tutkimuksen hyötyjen ja haittojen vertailun on pystyttävä osoittamaan, että tutkimuksen hyöty on selvästi mahdollisia riskejä ja haittoja suurempi. Lisäksi on erityisen tärkeää, että tutkittava osallistuu tutkimukseen vapaaehtoisesti ja saa tutkimuksesta kirjallisesti tietoa. (Launis 2007, 28.)

Klaus Mäkelä (2007) on perehtynyt eri maiden ihmistieteiden eettiseen ennakoarviointiin. Hänen mukaansa ennakoarviointijärjestelmä saa aikaan tutkittaville tarkoitettuja informointitekstejä, jotka ovat äärimmäisen teknisiä ja monimutkaisia. Yleensä niihin sisällytetään hyvin tiukasti määritellyt tutkimusaineistojen käyttötarkoitukset. Se kaventaa huomattavasti tunnisteellisten aineistojen jatkokäyttömahdollisuuksia. (Mt. 23.) Henkilötietolaki itsessään ei estäisi tutkimusaineistojen käyttötarkoituksen laivempaa muotoilua (ks. Kuula 2007).

Henkilötietolain tiukan tulkinnan sisällyttäminen eettisten toimikuntien antamiin ohjeisiin on jo nyt kaventanut huomattavasti lääketieteellisten aineistojen jatkokäyttömahdollisuuksia. Suomessa ollaankin valmistelemaan biopankkilakia, jotta vanhoja näytekokoelmia voitaisiin käyttää uusin lääketieteellisiin tutkimuksiin.

Mikäli Suomeen suunnitellaan myös ihmistieteiden eettinen ennakoarviointijärjestelmä, yllä mainitut ongelmat on syytä ottaa huomioon. Järjestelmä on hyvä suunnitella niin, että esimerkiksi biotieteissä ilmenneet ennakoarvioinnin ongelmat onnistuttaisiin ihmistieteissä ennalta välttämään. Järjestelmä ei saisi kaventaa tutkimusaineistojen avoimuutta ja tutkimustulosten kontrolloitavuuden periaatetta.

### 2.3 Rahoitushakemus, tutkimusetiikka ja tutkimusaineistot

Käytännössä tutkimusrahoittajat edellyttävät tutkimushankkeilta hyvän tutkimusetiikan noudattamista. Suomessa periaatteet on kirjattu tutkimuseettisen neuvottelukunnan eli TENK:n laatimiin ohjeisiin (Hyvä tieteellinen käytäntö ja sen loukkausten käsitteleminen 2002). Esimerkiksi Suomen Akatemia edellyttää, että sen rahoittamissa tutkimushankkeissa noudatetaan tutkimuseettisen neuvottelukunnan ohjeita.

Tutkimuseettisen neuvottelukunnan ohjeissa sivutaan tutkimusaineistoja suoraan tai välillisesti useissa kohdissa:

*”Yksi tieteellisen hyväksyttävyyden sekä sen luotettavuuden ja tulosten uskottavuuden edellytys on, että tutkimus on suoritettu hyvän tieteellisen käytännön (good scientific practice) edellyttämällä tavalla. Hyvään tieteelliseen käytäntöön kuuluu mm. että tutkijat ja tieteelliset asiantuntijat*

- noudattavat tiedeyhteisön tunnustamia toimintatapoja: rehellisyyttä, yleistä huolellisuutta ja tarkkuutta tutkimustyössä, tulosten tallentamisessa ja esittämisessä sekä tutkimusten ja niiden tulosten arvioinnissa;*
- soveltavat tieteellisen tutkimuksen kriteerien mukaisia ja eettisesti kestäviä tiedonhankinta-, tutkimus- ja arviointimenetelmiä sekä toteuttavat tieteellisen tiedon luonteen kuuluvaa avoimuutta tutkimuksensa tuloksia julkaistessaan;*
- ottavat huomioon muiden tutkijoiden työn ja saavutukset asianmukaisella tavalla niin, että he kunnioittavat näiden työtä ja antavat heidän saavutuksilleen niille kuuluvan arvon ja merkityksen omassa tutkimuksessaan ja sen tuloksia julkaistessaan;*

*Hyvän tieteellisen käytännön mukaista on edelleen, että:*

- tutkimus on suunniteltu, toteutettu ja raportoitu yksityiskohtaisesti ja tieteelliselle tiedolle asetettujen vaatimusten edellyttämällä tavalla;*
- tutkimusryhmän asema, oikeudet, osuus tekijyydestä, vastuu ja velvollisuudet sekä tutkimustulosten omistajuutta ja aineistojen säilyttämistä koskevat kysymykset on määriteltä ja kirjattu kaikkien osapuolten hyväksymällä tavalla;*
- rahoituslähteet ja tutkimuksen suorittamisen kannalta merkitykselliset muut sidonnaisuudet ilmoitetaan tutkimukseen osallistuville ja raportoidaan tutkimuksen tuloksia julkaistaessa; sekä*
- noudatetaan hyvää hallintokäytäntöä ja henkilöstö- ja taloushallintoa.”*  
*(Hyvä tieteellinen käytäntö ja sen loukkausten käsitteleminen 2002, 3.)*

Uusien tutkimusaineistojen eettisesti kestävä hankinnan periaatteisiin kuuluu, että tutkijat ottavat keruussa huomioon sekä lainsäädännön että tutkimuseettiset kysymykset. Aineistojen tarkoituksenmukaista hankintaa edesauttaa paitsi tutkimusjulkaisujen, myös olemassa olevien sähköisten tutkimusaineistojen saatavuus jatkokäyttöön.

Aineistojen elinkaaren suunnittelu ensivaiheesta jatkokäyttöön on keskeinen osa huolellista ja tarkkaa tulosten tallentamista. Se on samalla olennainen osa tieteellisen tiedon luonteen kuuluvaa avoimuutta; tutkimusaineistoja ei pidä rajata avoimuuden ulkopuolelle ilman perusteltuja syitä. Yleiset tutkimuseettiset näkökohdat puoltavat eettisesti kestävä tapaa kerätä ja käyttää tutkimusaineistoja. Julkisrahoitteisten sähköisten tutkimusaineistojen osalta tämä voi tarkoittaa, että yhä suurempi osa tutkimushankkeiden keräämistä keskeisistä aineistoista kiertää edellä esitetystä kaaviossa 2.1 kuvatun tutkimusdatan elinkaaren.

## 2.4 Rahoittajien datapolitiikka ja rahoituspäätökset

Tutkimussuunnitelma ja tutkimuksen toteuttamisen muut asiakirjat eivät vielä takaa tutkimusaineistojen avoimuutta. Suomen Akatemialla sen paremmin kuin muillakaan tutkimusrahoittajilla ei ole vahvoja välineitä tutkimusaineistoja koskevien periaatteiden toteutumisen valvontaan, eikä se välttämättä edes ole rahoittajan tehtävä. Hyvien tieteellisten käytäntöjen noudattamisen ensisijainen vastuu on tutkijoilla itsellään ja heidän tutkimusorganisaatioiltaan.

Käytännön kokemukset osoittavat, että pelkästään tutkijoiden oman toimeliaisuuden varaan jättäytyminen tukee vain harvoin riittävästi tutkimusaineistojen pitkää elinkaarta. Tämä pätee varsinkin pieniin ja keskisuuriin hankkeisiin, joissa kerätään valtaosa uusista sähköisistä tutkimusaineistoista. Aineistoja koskevat ohjeet eivät ole kovin yksityiskohtaisia. Tutkijat orientoituvat mieluummin aloittamaan uusia tutkimushankkeita kuin arkistomaan vanhoja aineistoja. Siksi julkisrahoitteisesti kootut sähköiset tutkimusaineistot eivät tavallisesti päädy tiedeyhteisön laajempaan käyttöön.

Yksi mahdollisuus on, että rahoittaja suosittelee omissa ohjeissaan tutkimusaineiston luovuttamista jatkokäyttöön rahoitettavan hankkeen päätyttyä. Esimerkiksi Suomen Akatemia on jo 2000-luvun alusta saakka erikseen suositellut siltä rahoitusta saaneiden yhteiskuntatieteellisissä hankkeissa koottujen sähköisten tutkimusaineistojen tallentamista Yhteiskuntatieteelliseen tietoaarkistoon (FSD) tutkimushankkeiden päätyttyä. Suositukseen perustuvia yhteydenottoja on tullut tietoaarkistoon erittäin vähän, vain muutamia vuoden 2006 loppuun mennessä. Tosin suositus on ollut voimassa vasta muutamia vuosia.<sup>2</sup>

Tutkimusrahoittajien yleiset periaatelinjaukset ja tutkimusaineistoja koskevat erityissuositukset ovat kuitenkin tärkeitä. Periaatelinjauksilla, suosituksilla, velvoitteilla tai laajimmillaan

<sup>2</sup> Tietoaarkisto kerää tietoja Suomen Akatemian sekä muiden tutkimusrahoittajien rahoittamien hankkeiden tutkimusaineistoista seuraamalla tiedotusvälineitä, tieteellisiä julkaisuja sekä julkaisu- ja hankerekistereitä. Tietoja olemassa olevista aineistoista saadaan myös ottamalla suoraan yhteyttä tutkijoihin ja tutkimusorganisaatioihin.



erikseen määritellyllä datapolitiikalla<sup>3</sup> voi olla suurtakin merkitystä pitkällä aikavälillä tutkimusaineistojen avoimen saatavuuden kannalta.

Suomessa koko idea tutkimusaineistojen avoimen jatkokäytön aktiivisesta tukemisesta on vielä verraten uusi. Alan tukirakenteet, toimintatavat ja toimintakulttuuri ovat vasta muotoutumassa. Tällaisessa vaiheessa keskeisten tiederahoittajien ja tutkimusorganisaatioiden periaatteellinen ja selvästi julkilausuttu sitoutuminen tutkimusaineistojen pitkän elinkaaren tukemiseen on mitä tärkeintä.

Joissakin maissa tutkimusrahoitusta saavat hankkeet saatetaan velvoittaa luovuttamaan tutkimusaineistojaan tutkimusryhmän omia tarpeita laajempaan käyttöön. Näin tutkimusaineiston luovuttaminen avoimeen jatkokäyttöön toimii tutkimusrahoituksen välttämättömänä ehtona edellyttäen, ettei aineiston jatkokäytölle ole esimerkiksi juuri OECD:n datasuosituksessa mainittua estettä.

Aineistojen avointa jatkokäyttöä edellyttävät tutkimusrahoittajat yleensä tarjoavat mahdollisuuden pieneen rahoituslisään, jonka turvin aineisto muokataan pitkäaikaiseen säilytykseen ja jatkokäyttöön soveltuvaksi. Käytäntö on hyvä, sillä tutkimusaineiston saattaminen jatkokäyttökuuntoon vaatii yleensä tutkimusryhmältä lisätyötä. Lisärahoituksen tyypillinen käyttötarkoitus on dokumentointi- ja muokkaustyö, jota tutkimushankkeessa ei välttämättä jouduttaisi tekemään siinä tapauksessa, että tutkimusdata jäisi vain kokoajensa käyttöön.

Muita rahoittajien keinoja julkisrahoitteisen tutkimusdatan avoimen käytön tueksi voivat olla osallistuminen aineistojen käyttöä mahdollistaviin (yleensä kansainvälisten tutkimusinfrastruktuureiden) jäsenmaksuihin, käyttölisensseihin tai aineistonkeruukustannuksiin ilman että rahoitus kanavoidaan tavanomaisen tutkimushankerahoituksen kautta.

## 2.5 Datat keruu

Kaaviossa 2.1 tutkimusaineiston keruu ja sen valmistelu sijoittuvat elinkaaren viidenneksi vaiheeksi. Tutkimusaineiston laatu määrittää sen, millaisia töitä keruun valmisteluun ja toteuttamiseen liittyy. Alkuvaiheen toimenpiteet ovat erittäin ratkaisevia aineiston elinkaarelle. Alkuvaiheessa tehdyt ja tekemättä jääneet sopimukset ja sitoumukset, itse aineiston koonnin onnistuminen ja eri toimenpiteiden kirjaaminen vaikuttavat suoraan siihen, missä määrin aineistoa voidaan käyttää ensikäytön lisäksi myöhemmin mahdollisiin muihin tutkimustarkoituksiin. Seuraavassa asiaa havainnollistetaan kysely- ja haastatteluaineistojen tyypillisillä keruutoimenpiteillä.<sup>4</sup>

Tutkimusaineistoa voidaan kerätä suoraan tutkittavilta esimerkiksi lomakekyselyllä, haastattelun, kirjoituspyyntönä tai videoiden. Kaikissa tapauksissa tutkittaville tulee antaa tietoa tutkimusaineiston käytöstä ja tämä informaatio on tutkijaa sitova. Jos aineisto sisältää tunnistetietoja, informoinnin sisältö määrittää tutkimusaineiston kohtalon: voiko tutkija itse tai hänen kollegansa käyttää aineistoa tutkimuksen päätyttyä, voiko aineiston arkistoida vai onko aineis-

<sup>3</sup> Tutkimusaineistojen jatkokäyttöä tukevia koti- ja ulkomaisia menettelytapoja ja datapolitiikkoja käsitellään tarkemmin luvussa 5.

<sup>4</sup> Nämä tutkittavien informointiohjeet mallilomakkeineen löytyvät FSD:n verkkosivuilta, ks. [http://www.fsd.uta.fi/laki\\_ja\\_etiikka/keruuhjeet.html](http://www.fsd.uta.fi/laki_ja_etiikka/keruuhjeet.html)

to hävitettävä heti tutkimuksen tulosten oikeellisuuden tarkastamisen jälkeen. Informoinnissa tulee ottaa huomioon kolme pääasiaa:

- mahdollisuus tutkimusaineiston järkevään hyödyntämiseen
- vallitseva lainsäädäntö
- tutkittavien antamien tietojen salassapito.

Informointi voi tutkimusaiheesta ja keruutavasta riippuen olla suullinen tai kirjallinen. Aina-kin seuraavia seikkoja koskevat menettelyt on hyvä kertoa tutkittaville:

- vastuullisen tutkijan/tutkimusryhmän nimi ja yhteystiedot
- tutkimuksen tavoite ja käyttötarkoitus
- tutkimukseen osallistumisen vapaaehtoisuus
- luottamuksella annettujen tietojen suojaaminen
- tutkittavilta kerättävien tietojen käsittely, käyttäjät ja säilyttäminen/hävittäminen
- tutkittavia koskevat mahdolliset muut tiedonlähteet (jos aineistoon yhdistetään rekisteritietoja).

Informoinnin esimerkkejä löytyy muun muassa Yhteiskuntatieteellisen tietoarkiston kotisivuilta<sup>5</sup>. Kun kaikki osa-alueet on tehty tutkittaville selväksi, heillä on mahdollisuus antaa riittävään tietoon ja asian ymmärtämiseen perustuva suostumus (*informed consent*) tutkimukseen osallistumisesta. Kun kyseessä ei ole arkaluonteinen tutkimusaihe ja tiedot kerätään julkisella paikalla, myös suullinen informaatio riittää.

Tutkimusaineistoja voidaan muodostaa myös erilaisista viranomaisrekistereistä. Rekisteritutkimuksiin tai rekisteritietoja osana tutkimusaineistoa hyödyntäviin tutkimushankkeisiin liittyy joukko lainsäädännöllisiä ja rekisterinpitäjien käytäntöihin ja tulkintoihin liittyviä erityispiirteitä, joita ei käsitellä tässä selvityksessä tarkemmin. Viranomaisrekisterit ovat kuitenkin keskeisiä julkisrahoitteisia sähköisiä tutkimusaineistoja ja niiden nykyistä avoimempi käyttömahdollisuus kuuluu itsestään selvänä osana OECD:n datasuositukseen. Suomessa viranomaisten hallinnollisiin tarkoituksiin kerättyjen rekisteriaineistojen tutkimuskäyttöä edistää Stakesin yhteydessä toimiva Rekisteritutkimuksen tukikeskus.

Aina tutkimushankkeessa ei kerätä itse aineistoa, vaan keruun toteuttaa jokin ulkopuolinen organisaatio. Tämä saattaa vaikuttaa paljonkin siihen, kuinka kattavasti ja tarkasti tutkimusaineiston tietosisällöt ja konteksti sekä tiedonkeruun muut oleelliset seikat tulevat tallennetuiksi sähköisiin tutkimusaineistoihin ja niitä kuvaileviin metatietoihin. Tutkimusaineistojen jatkokäytölle olennaisia tietoelementtejä esitellään tarkemmin liitteessä 4.

5 Ks. <http://www.fsd.uta.fi>.

## 2.6 Ensikäyttö

Tutkimusaineiston ensikäytöllä (primary use) tarkoitetaan alkuperäisessä tutkimussuunnitelmassa ilmaista käyttötarkoitusta. Jatkokäytöllä (secondary research, re-use) viitataan puolestaan käyttöön, jossa joko ensi- tai jatkokäyttäjät hyödyntävät tutkimusaineistoa muuhun kuin alkuperäisen tutkimussuunnitelman sisältämään tutkimustarkoitukseen (ks. myös Keckman-Koivuniemi & Kleemola 2006, 7).

Tutkimusdatan elinkaarta ja jatkokäyttöä painotettaessa on syytä korostaa, että tutkimusaineistot kerätään ensisijaisesti ensikäyttöä varten. OECD:n datasuosituksen avoimen saatavuuden ja tehokkaan kokonaiskäytön periaatteen mukaisesti datan elinkaaren ei silti tulisi rajoitua vain ensikäyttöön ilman painavia syitä.

Aina ei ole itsestään selvää, keitä ensikäyttäjät ovat ja kuka päättää aineiston ensikäytöstä sekä mahdolliseen jatkokäyttöön luovuttamisesta. Tämä koskee varsinkin yliopistoissa tehtävää, mutta myös muissa tutkimusorganisaatioissa toteutettavaa tutkimusta. Hyvin usein tutkimusaineistoa ovat olleet suunnittelemassa, keräämässä ja saattamassa käyttökuuntoon useat tutkijat, ohjaajat, esimiehet sekä avustavat henkilöt.

Tietoarkiston käytännön kokemusten nojalla näyttää siltä, että ainoastaan ensikäyttäjär ryhmän jakamaton yksimielisyyden tutkimusaineiston luovuttamisesta jatkokäyttöön on riittävä edellytys aineiston myöhemmälle avoimelle jatkokäytölle. Aineistolle tosin saattaa ilmaantua uusia, ennalta suunnittelemattomia käyttötarpeita tutkimushankkeen tai -organisaation sisällä jo ensikäytön aikana.

Viimeistään ensikäyttövaiheessa tutkimusryhmän ja -organisaation tulee päättää, kuka tai ketkä vastaavat aineiston käyttöoikeuksien myöntämisestä ja millä tavoin lupa myönnetään. Vastuuhenkilöt eivät aina ole samoja kuin tutkimuksen rahoitusvaiheessa hankkeelle vastuullisiksi osoitetut henkilöt.

Joillakin tutkimusaloilla nykytilanne on selkeämpi kuin toisilla. Esimerkiksi Laki lääketieteellisestä tutkimuksesta edellyttää, että tutkimukselle on nimetty tutkimuksesta vastaava henkilö (ks. tarkemmin KTL 2005, 66–67). Joissakin keskeisissä suomalaisissa tutkimusorganisaatioissa, kuten esimerkiksi Tilastokeskuksessa ja Kansanterveyslaitoksessa, on myös jo otettu käyttöön aineistojen ensikäyttöä ja jatkokäyttöön luovuttamista koskevia menettelyohjeita.

Joskus ensikäytön ja jatkokäytön välinen raja on häilyvä. Joskus ensikäyttövaihe muodostuu pidemmäksi kuin mitä alkuperäiseen tutkimussuunnitelmaan on kirjattu. Toisinaan ”pitkittynyt” ensikäyttö kohdistuu vain pieneen osa-aineistoon. Myös alkuperäisen tutkimuksen pitkittyminen voi venyttää ensikäyttöaikaa niin, että se estää aineiston tehokasta avointa käyttöä. Samalla ikääntyvän aineiston käyttöarvo voi laskea. Siksi julkisrahoitteisille sähköisille tutkimusaineistoille voitaisiin harkita asetettavaksi ennakkoon kohtuullinen ensikäyttöaika, jonka jälkeen datat siirrettäisiin jatkokäyttöön tutkimuksesta vastaavan henkilön tai muun vastaavan tahon päätöksellä.

Yksi aineistojen ensikäytön ja myöhemmän käytön suhteeseen liittyvä keskeinen kysymys koskee sitä, mitä ja millä tavoin muokattuja tietosisältöjä jatkokäyttöön tarkoitettun tutkimusaineiston tulee sisältää. Aihe on tärkeä tietosuojanäkökohdian ohella myös siksi, että ensikäytön aikana aineistoa saatetaan täydentää lisäämällä siihen ensikäyttötarkoituksen vaatimia muokattuja tai kokonaan uusia tietoja.

Avoimeen käyttöön tarkoitettun aineiston tulee olla tietosisällöltään eheä ja riittävän helpokäyttöinen. Samalla on huolehdittava, että aineistosta ei ole häivytetty informaatiota tarpeettomasti esimerkiksi turhan ahtaiden tietosuojatulkintojen vuoksi. Huomiota tulisi kiinnittää etenkin siihen, että aineisto on tarpeen mukaan vertailtavissa ja yhdistettävissä muihin tutkimusyhteisön olennaisina pitämiin tietoihin.

## 2.7 Datan arkistointi ja julkaiseminen

Ensikäytön jälkeen julkisrahoitteiset sähköiset tutkimusaineistot jäävät usein lähes kokonaan pimentoon, kun niiden kerääjät ovat jo uusien tutkimustensa parissa. Kun tutkimusdatan ensikäyttöaikaa ei ole määritetty tai sitä ei noudateta, tutkimushankkeet eivät välttämättä pääty lainkaan ensikäyttäjien omasta mielestä. Tutkimusaineistot jäävät hamaan ikuisuuteen vain kerääjiensä haltuun odottamaan mahdollista myöhempää käyttötarvetta. Tällaiseen toimintatapaan sisältyy useita riskejä, joita käsitellään myös OECD:n datajulistuksessa ja -suosituksessa. Sähköiset tutkimusaineistot eivät säily käyttökuntoisina itsestään. Niiden käytettävyys saattaa heikentyä ratkaisevasti jo muutaman vuoden kuluessa. Yksi tapa huolehtia aineistojen käyttökuntoisena säilymisestä on käyttää tutkimusaineistojen arkistointiin keskittyvää yksikköä.

Yhteiskuntatieteellinen tietoarkisto edellyttää arkistoitavalta aineistolta, että

- siihen liittyvät omistus- ja hallintasuhteet ovat riittävän selkeät
- aineistoa koskeva lainsäädäntö, esimerkiksi henkilötietojen käsittelyä koskeva lainsäädäntö, ei estä arkistointia
- alkuperäinen tiedonkeruun käyttötarkoitus ei estä aineiston arkistointia
- aineisto on muutoinkin tietosisällöltään ja teknisiltä ominaisuuksiltaan arkistointikelpoinen.

Tietoarkisto voi tietyin edellytyksin vastaanottaa myös tunnistetietoja sisältäviä tutkimusaineistoja. Mikäli henkilötietolain edellytykset täyttyvät, aineisto voidaan arkistoida tunnistettuna. Muussa tapauksessa tietoarkisto anonymisoi aineiston pysyvää säilyttämistä varten luovuttajan eli yleensä ensikäyttäjän toimeksiannosta.

Aineistoa arkistoidaessa aineistojen luovuttajat täyttävät aineistonkuvauslomakkeen, usein yhdessä tietoarkiston työntekijän kanssa. Riittävän tarkka dokumentointi on paras taie aineistojen käyttökelpoisuuden säilymiselle. Aineiston sisällön ja keruun kuvauksen lisäksi aineistonkuvauslomakkeessa tiedustellaan keskeisiä tietoja alkuperäisestä tutkimushankkeesta ja aineistosta jo tuotetuista julkaisuista. Aineistonkuvauslomakkeen sijasta tai lisäksi aineiston luovuttaja voi toimittaa arkistoon tutkimusraportin tai kopion tutkimusraportin osasta, jossa kuvataan aineistonkeruu ja aineiston sisältö.

Aineistonkuvauslomakkeen lisäksi arkisto ja ensikäyttäjä täyttävät erityisen luovutus sopimuksen, jossa ensikäyttäjä voi asettaa luovuttamalleen aineistolle käyttöehtoja. Lisäksi ensikäyttäjä ilmoittaa luovutus sopimuksessa, miten häntä tulee informoida luovuttamansa aineiston jatkokäytöstä.

Arkistoitava aineisto käydään läpi huolellisesti. Kvantitatiivisen aineiston muuttajat tunnistetaan, yksittäiset muuttajat ja niiden vastaavuus tiedonkeruun sisältöihin tarkistetaan, filt-

terimuuttujien toimivuus sekä puuttuvat tiedot käydään läpi ja lopuksi dataan lisätään mahdollisimman tarkat kuvaustiedot muuttujien sisällöstä ja niiden saamista arvoista. Lisäksi aineistosta yleensä poistetaan tutkijan itse luomat muuttujat niin, että arkistoidussa aineistossa mukana ovat vain kyselylomakkeeseen tai muuhun tiedonkeruulinstrumenttiin suoraan liittyvät muuttujat. Erityistä huomiota kiinnitetään tietosuojangelmiin, joiden ratkaisemiseksi muuttujia voidaan esimerkiksi jättää kokonaan pois tai luokitella uudelleen.

Yksi keskeinen osa aineiston prosessointia on tarvittaessa tehtävä anonymisointi. Se tehdään joko tarkistamalla alkuperäisen tutkijan anonymisointi tai poistamalla tunnisteet arkistossa aineiston prosessoinnin yhteydessä (peruste HetiL 8 § 7 mom.). Anonymisoinnin taso määräytyy siitä, mikä on ollut tutkittaville annettu alkuperäisen tutkimuksen aineistoa koskeva käyttöinformaatio tai kirjallisen tutkimussuostumuksen sisältö (peruste HetiL 7 §).

Kun alkuperäisestä aineistosta on prosessoitu arkistoversio, se kuvaillaan tietoarkiston aineistotietokantaan hyödyntäen alan kansainvälisesti vakiintunutta metadataformaattia (DDI). Lisäksi aineiston sisältöä keskeisesti kuvaavat dokumentit (esim. haastattelukysymykset, kirjoituspyyntö, kyselylomake) digitoidaan, jos ne ovat saatavilla vain paperimuodossa.

Kun tutkimusaineiston arkistointi on valmis, tietoarkisto *julkaisee* aineiston kuvailevat perustiedot aineistotietokannassaan ja antaa aineistolle viittaustiedot. Samat tiedot käännetään ja julkaistaan myös englanninkielisinä. Kotimaisen ja kansainvälisen tiedeyhteisön yhteiseen käyttöön arkistoitu ja julkaistu tutkimusaineisto on tieteellinen lisämeriitti ensikäyttäjälle. Kaikki aineiston jatkokäyttäjät ovat veloitettuja viittamaan käyttämäänsä aineistoon annetulla viittaustiedolla.

Data-arkistot ovat yleensä sisällyttäneet jatkokäyttölupiinsa vastuuvapauslausekkeen, jonka mukaan tutkimusaineiston luovuttajalla ei ole vastuuta aineiston jatkokäytössä puutteellisesti tuotetuista tai tulkituista tutkimustuloksista. Lisäksi jatkokäyttäjää veloitetaan informoimaan tietoarkistoa aineistoissa havaituista puutteista.

## 2.8 Jatkokäytön tuki

Datan arkistointi ei yksin ratkaise jatkokäytön tuen peruskysymyksiä. Tuen voi ryhmitellä yleisiin toiminnallisiin kysymyksiin, informaation välittämiseen ja taloudellisiin kysymyksiin.

Jatkokäytön tuen *yleiset toiminnalliset kysymykset* liittyvät siihen, mihin ja miten aineisto on arkistoitu ja kuinka sitä voidaan käyttää. Yleensä aineisto on jatkokäyttäjälle sitä helpokäyttöisempi, mitä enemmän aikaa ja vaivaa sen arkistointiin on uhrattu. Huolellisinta aineiston arkistointi lienee silloin, kun siitä on vastannut alan erikoisorganisaatio.

Erilaisen käyttäjätuen tarve, saanti ja luonne riippuvat siitä, millä tavoin sähköinen tutkimusdata on käytettävissä. Käyttäjäystävällisimmässä malleissa aineistoa kuvailevat ja sen käyttöön liittyvät tiedot ovat helposti saatavilla Internetissä ja jatkokäyttäjä saa sähköisestä aineistosta kopion itselleen. Tällöin yleensä myös varsinaiset aineistotiedostot siirtyvät jatkokäyttäjän tai hänen tutkimusorganisaationsa tietokoneelle.

Nykyisin myös Internetiä hyödyntävät online -ratkaisut mahdollistavat sähköisten tutkimusaineistojen peruskäytön ilman että itse tutkimusdataa tarvitsee edes lähettää jatkokäyttäjälle. Tietoteknisesti voidaan hallinnoida käyttäjien rekisteröityminen, käyttöoikeuksien mää-

rittäminen ja käyttäjän tunnistaminen sekä hallita aineistojen tietosuojakysymykset. Lisäksi aineistojen käyttöön mahdollisesti liittyvät maksut voidaan hoitaa käyttöliittymillä.

Kaikkein luottamuksellisimpien aineistojen, kuten esimerkiksi rekisteritietojen, käytössä saatetaan kuitenkin vaatia aineistojen käyttöä paikan päällä (on-site) organisaatiossa, joka vastaa aineiston tietojen luottamuksellisuuden säilymisestä sekä neuvoo aineistojen jatkokäyttäjää. Muut jatkokäytön tuen pääkysymykset kuuluvat lähinnä seuraavaan kahteen ryhmään:

*Informaation välittyminen ja turvaaminen:*

- Miten varmistetaan julkisrahoitteisen sähköisen tutkimusdatan riittävän informoitu jatkokäyttö?
- Millaisia ovat jatkokäytön neuvontapalvelujen erilaiset toteuttamismahdollisuudet?
- Millaisia välineitä tutkimusaineistojen jatkokäyttö edellyttää?
- Miten varmistetaan jatkokäytön edellyttämät tietosuojanäkökohdat ja muu tietoturvasuus?
- Millaisia tietoja jatkokäyttäjät voivat liittää saamiinsa aineistoihin?

*Jatkokäytön tuen taloudelliset näkökohdat:*

- Onko julkisrahoitteisesti jo kerran kustannettu tutkimusaineisto saatavilla jatkokäyttöön maksutta vai maksaako se?
- Kuka maksaa ja mistä maksetaan?
- Velvoitetaanko jatkokäyttäjät osallistumaan takautuvasti aineistonkeruukustannuksiin vai maksavatko he pääasiassa vain aineiston saattamisesta jatkokäyttökuntoon?
- Maksavatko neuvontapalvelut ja kuinka ne rahoitetaan?
- Voidaanko joihinkin tiedeyhteisön laajasti hyödyntämiin aineistoihin luoda jatkokäyttöä tukevia rahoitusmalleja, kuten esimerkiksi datalisenssejä?

Eri tutkimusalat ja -aineistot edellyttävät erilaisia tukiratkaisuja. Joillakin aloilla jatkokäytön esteenä ovat datoihin liittyvät kaupalliset intressit, toisilla muun muassa tutkimuskohteiden tietosuojaa koskevat seikat.

## 2.9 Jatkokäyttö

Aineiston jatkokäyttäjällä pitää olla suunnitelma ja käyttötarkoitus haluamalleen aineistolle. Tutkimusdatassa käyttötarkoitus on yleensä tieteellinen tutkimus- tai opetus. Jatkokäytön suunnitelma ei aina ole yhtä perusteellinen kuin uusien aineistojen kokoavien hankkeiden tutkimussuunnitelmat, sillä aiemmin kerättyä aineistoa käytetään vain harvoin jatkotutkimuksen ainoana empiirisenä tutkimusaineistona.

Arkistoituja aineistoja voidaan jatkokäyttää hyvin monin tavoin. Itse OECD:n datajulistus sisältää useita tärkeitä seikkoja aineistojen mahdollisista käyttötavoista. Kleemola ja Keckman-Koivuniemi (2006, 10–14) ovat kirjassaan esittäneet tyypillisimmät arkistoina aineistojen käyttötavat, *joista tietoarkistotalalla on jo kokemusta*.

Arkistoidut aineistot sisältävät paljon täysin käyttämätöntä materiaalia. Niin tutkimusaineistot kuin alun alkaen hallinnollisiin tarkoituksiin kerätyt aineistotkin ovat antaneet mah-

dollisuuksia tutkimusasetelmiin ja -kysymyksiin, joita ei ole lainkaan selvitetty alkuperäisissä tutkimuksissa ja selvityksissä. Arkistoaineistoja on hyvin paljon käytetty myös uuden tutkimuksen tausta-aineistoina. Arkistoitujen aineistojen käyttökelpoisuus on korostunut, kun on haluttu kuvailla tutkimuksen yhteiskunnallisia, sosiaalisia, historiallisia ja kansainvälisiä konteksteja.

Arkistoidut tutkimusaineistot ovat osoittautuneet erityisen arvokkaiksi, kun tutkimukseen on sisällytynyt ajallisia ja kansainvälisiä vertailuja. Arkistoaineistoja on myös käytetty uusien tutkimusten suunnitteluun ja tutkimusmenetelmien kehittämiseen, erityisesti mittareiden ja analyysimenetelmien testaamiseen. Tieteen peruseräiteisiin kuuluvan tulosten todentamisen ja tarkistamisen ohella arkistoaineistoja on hyödynnetty paljon opetuksessa ja opiskelussa. Tieteellisen tutkimuksen opetus ja opiskelu ovat osoittautuneet hyvin tärkeäksi sähköisten arkistoaineistojen sovellusalueeksi. (Mt. 10–14.)

OECD-julistuksessa sähköisten tutkimusaineistojen käytölle ei aseteta rajoituksia. Esimerkiksi Yhteiskuntatieteellisen tietoarkiston periaatteena on, että datan jatkokäytön on liityttävä tieteelliseen tutkimukseen tai opetukseen, käyttötarkoitus ei saa olla kaupallinen. Aineistojaan arkistoitavaksi luovuttavat tutkijat usein arvostavat käyttöoikeuksien täsmentämistä. Sähköisten aineistojen jatkokäytön avoimuuden asteita ja toimintamalleja tarkastellaan yksityiskohdaisesti luvuissa 4 ja 5.

## 2.10 Julkaisu- ja data-tietokannat

Ensi- ja jatkokäyttö tuottavat tutkimusjulkaisuja, joiden keskeiset tiedot tallennetaan erilaisiin julkaisutietokantoihin ja tutkimushankerekistereihin. Julkaisu- ja hanketietokannat sisältävät kuitenkin valitettavan harvoin tietoja julkaisuissa käytetyistä sähköisistä tutkimusaineistoista – poikkeuksena jotkut luonnontieteiden sähköiset julkaisut. Tavallisesti kirjastojen, yliopistojen ja tutkimusorganisaatioiden julkaisu- ja hankerekisterit laaditaan toisistaan erillisinä. Tarjolla on paljon hajanaisia ja toisistaan riippumattomia sekä osin päällekkäisiäkin tutkimustietokantoja.

Tietoarkistot kokoavat monissa maissa tietoja sekä tutkimusaineistoista että julkaisuista, joissa niitä on käytetty. Tiedot julkaistaan data-arkistojen kotisivuilla sähköisissä aineistoluetteloissa. Kansainvälisestikin kattavat aineistotietokannat mahdollistavat tutkimustietojen innovatiivisen vertailun ja yhdistelyn. Vastaavalla tavalla monet suuret aineistonkeruunfrastruktuurit ja niiden tutkimushankkeet ylläpitävät aineistoihinsa kytkeytyviä julkaisutietoja ja tutkimushanketietoja.

Uuden tiedon tuottamisen kannalta tiedeyhteisön on syytä vaatia nykyistä punnekkaimmin välineitä tiedejulkaisujen taustalla olevien tutkimusdatojen löytämiseksi. Tutkimustulosten ja aineistojen yhteys korostuu myös siinä, että useat tieteenalat ja arvostetut kansainväliset tieteelliset aikakauslehdet ovat viime vuosina ryhtyneet edellyttämään tutkimusaineistojen julkaisemista tiedeyhteisön avoimeen käyttöön ennen kuin tutkimustuloksia voidaan julkaista. Näin pyritään edistämään tutkimustulosten tosiasiallista tieteellistä arviointia ja estämään samalla erehdyksiä tai jopa tieteellistä vilppiä.



### 3. Tutkimusdatan säilytys ja avoimuuden esteet

Sähköisten tutkimusaineistojen linkaareen vaikuttavat monet toimijat. Tutkimuksen rahoittajat, tutkimusorganisaatiot, tiedeyhteisö ja myös tutkimusten julkaisijat voivat edellyttää toimia tai tarjota resursseja, jotka edistävät suoraan aineistojen avointa saatavuutta. *Ratkaisuvimmassa asemassa ovat aineistoja kokoavat tutkijat.* Heidän tutkimusdatojensa käytettävyyden, tutkimustoimintansa suunnitelmallisuus sekä suhtautumisensa tutkimusaineistojen avoimuuteen ovat avainkysymyksiä.

Tässä luvussa tarkastellaan liitteessä 2 kuvatun tutkijakyselyn tuloksia sähköisten tutkimusaineistojen arkistoinnista ja jatkokäytöstä keskittyen avoimuuden konkreettisiin, kulttuurisiin ja rakenteellisiin esteisiin. Vaikka tutkijat ovat ratkaisevassa asemassa, heidän työskentelyään määrittää myös lainsäädäntö, viralliset ohjeet ja yliopistojen hallinnolliset käytännöt. Siksi seuraavassa esitellään ensin aineistojen säilyttämistä koskevaa lainsäädäntöä sekä yliopistojen nykikäytäntöjä säilytyksen ja jatkokäytön sääntelyssä.

#### 3.1 Aineistojen säilyttämistä määrittävä lainsäädäntö ja yliopistojen käytännöt

##### 3.1.1 Lainsäädäntö ja tekijänoikeudet

Tutkimusaineistoja ja niiden keruuta määrittävien lakien ja virallisten ohjeiden käytännön merkitys vaihtelee tutkimusaloittain. Esimerkiksi lääketieteellisen tutkimuksen suunnittelussa jo ennen aineistonkeruuta suositellaan tutustumista lakeihin, sopimuksiin ja säädöksiin, joita on kaikkiaan liki 30 (ks. Pelkonen, Mäkelä, Haahela 2004, 2815).

Teknisillä aloilla tutkimusta raamittaa oleellisesti immateriaalioikeus. Sillä tarkoitetaan oikeutta, ”jossa käsitellään mm. kirjallisuudelle ja taiteelle, tieteelle ja tekniikalle, teolliselle muotoilulle ja suunnittelulle sekä tavaramerkille ja muille tunnusmerkeille annettavaa suojaa” (Edelman, Mylly ja Rajala 1998, 21). Teknisillä tieteenaloilla ja biotieteissä merkittäviä immateriaalioikeuksia ovat patenttioikeus ja muut keksintöjen suojaamisen oikeudet. Immateriaalioikeuksien tieteen avoimuutta rajoittavasta vaikutuksesta ja ongelmien mahdollisista ratkaisuista on kirjoittanut esimerkiksi Marjut Salokannel (2003).

Yliopistolaki (1997/645) ja sen nojalla annettu asetus määrittävät tutkimussopimuksia, joissa voidaan muun ohella päättää myös tutkimusaineiston kohtalosta. Tutkimussopimuksia käsittelevissä oppaissa painopiste on luonnontieteiden ja erityisesti tekniikan alan tutkimusaineistoista tehtävien sopimusten laatimisessa (esim. Saarnilehto ja Edelman 1998; Kiuru ja Koponen 2001). Ihmistieteissä sopimusoikeudellinen pääsääntö on, että tutkimusaineisto jää tutkimuksen toteuttamispaikan (yleensä tutkimuslaitoksen) vastuulle ja hallintaan, ellei toisin ole sovittu.

Tutkimusaineistot voivat olla tekijänoikeudellisesti suojattuja, mikäli ne täyttävät teoskynnyksen eli ovat tekijänsä luovan työn omaperäisiä tuloksia. Oikeudellisesta näkökulmasta teoksena ei välttämättä pidettäisi esimerkiksi haastattelun tapaisessa vuorovaikutuksessa syntyvää puhetta tai lomakekyselyllä kerättyä aineistoa. Lomakekyselymykset ovat tavallisesti tie-



tyn perustiedon ja päämäärien pohjalta muotoiltuja ja niiden ainoana tarkoituksena on selvittää tietty asiantila. Näin ollen niissä ei ilmene sellaista omaperäistä henkistä luomistyötä kuin teokselta Suomen tekijänoikeuslain mukaan vaaditaan. Teoskynnys ylittyy, jos voidaan olettaa, ettei kukaan muu vastaavaan työhön ryhtyessään olisi tehnyt samanlaista teosta. Itse asiassa niin omaperäisesti muotoillut lomakekysymykset eivät välttämättä toimisi tarkoitettulla tavalla tutkimusaineiston keruussa.

Edellä mainituista seikoista huolimatta tutkimusaineistoista ei voi sanoa mitään tekijänoikeudellista yleissääntöä, sillä teoskynnyksen ylittäminen arvioidaan aina tapauskohtaisesti. Tutkimusaineistoihin liittyvän tekijänoikeuden juridista arviointia mutkistaa vielä se, että tekijänoikeuslain viidennessä luvussa säädetään eräistä tekijänoikeutta lähellä olevista oikeuksista. Niillä annetaan suojaa esimerkiksi kuvatallenteen, luettelon ja tietynlaisen tietokannan valmistajalle ja valokuvaajalle. Lähioikeuksien tarkoitus on suojata tuotoksia, joiden suunnittelu, kerääminen ja käsittely ovat edellyttäneet huomattavaa panostusta. Suoja kestää 15 vuotta tuotoksen valmistumisesta.

Tekijänoikeudella suojattua tutkimusaineistoa voi käyttää jatkotutkimuksissa, kun tutkimusaineistoa ei saateta sellaisenaan yleisön saataviin. Tutkimusaineistot voivat koostua esimerkiksi videoista, kuvista, teksteistä tai kyselylomakkein kerätystä laajoista aineistoista. On hyvin todennäköistä että yhteen tutkimusaineistokokonaisuuteen voi sisältyä sekä tekijänoikeudellisesti suojattuja että suojaamattomia osia. Toisaalta tutkimusaineistojen tekijänoikeudella ei ole samanlaista merkitystä kuin julkaisuilla, sillä tutkimusaineistoja ei yleensä julkaista tai julkisteta. Tekijänoikeus on kuitenkin hyvin merkittävä tutkimuseettisesti. On ensiarvoisen tärkeää, että aineistojen jatkokäyttäjät viittaavat aineiston alkuperäisiin tekijöihin asianmukaisesti. (Ks. Kuula 2006, 69–70.)

Tekijänoikeutta olennaisemmin tutkimusaineistojen säilyttämistä ja jatkokäyttöä säätelee henkilötietolaki. Tieteellistä tutkimusta varten henkilötietoja saa käsitellä tutkittavan henkilön yksiselitteisesti antamalla suostumuksella, jos käsittely on rekisterinpitäjän toiminnan ja hallinnon kannalta asianmukaista (henkilötietolaki 8.1 § kohta 1). Käytännössä henkilötietojen tulee olla tarpeellisia tutkimuksen suorittamisessa ja niiden keräämisen edellytyksenä täytyy olla tieteellinen tutkimussuunnitelma. Tutkittavan suostumuksen tulee olla vapaaehtoinen, yksilöity ja tietoinen tahdonilmaisu. Kun kerätään arkaluonteisia tietoja, suostumuksen tulee olla lisäksi nimenomainen. (Kleemola 2000.) Nimenomainen suostumus on yleensä kirjallisesti tai muulla tavoin todisteellisesti annettu suostumus.

Kun tutkittaville luvataan vain yksilöity aineiston käyttötarkoitus, sovelletaan jo päättyneen tutkimuksen aineistoon henkilötietolain 14. pykälän neljättä momenttia. Käytännössä vaihtoehdot ovat aineiston tuhoaminen, arkistointi tai anonymisointi. Arkistointiin tulee hakea virallista lupaa Kansallisarkistolta, joka kysyy myös tietosuojavaltuutetun arvion asiasta. Arkistointilupia on tähän mennessä annettu lähinnä terveystieteiden laajoihin tunnistellisiin aineistoihin. Sosiaalitieteiden tai humanististen alojen tunnistellisten aineistojen arkistointilupahakemuksia ei ole toistaiseksi ollut käsiteltävänä.

Henkilötietolain merkitys yhdessä tutkimuseettisten periaatteiden kanssa tekee aineiston keruun ja käsittelyn suunnittelusta hyvin ratkaisevia. Tutkittavien informointi on suunniteltava huolella. Mahdollinen anonymisointi tulee myös suunnitella aineiston jatkokäytön ja arkistointiratkaisujen määrittämällä tavalla. Yhteiskuntatieteellisessä tietoarkistossa on opastettu tutkijoita sekä informoinnissa että anonymisoinnissa (ks. Kuula 2006 ja <http://www.fsd.uta.fi/>).

### 3.1.2 Tutkijoiden suhtautuminen aineistojen avoimuuteen professorikyselyn perusteella

Raporttia varten toteutettiin marras-joulukuussa 2006 verkkokysely, joka kohdennettiin kaikille Suomen yliopistojen humanististen, yhteiskuntatieteellisten ja käyttäytymistieteellisten alojen professoreille. Kyselyyn vastasi yhteensä 150 näiden alojen asiantuntijaa.

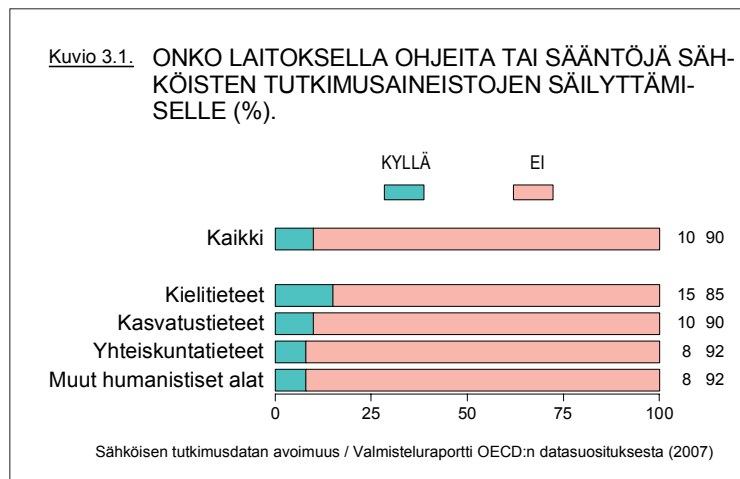
Vastaaajamäärä on tarpeeksi suuri tilastolliseen tarkasteluun, vaikka kyselyn vastausprosentti jäikin vaatimattomaksi (28,4%). Tutkijoilta saadun sähköpostipalautteen sekä tieteenaloittaisen perustulosten ja vastausprosenttien pohjalta voitiin päätellä, että noin puolet tutkimusjoukkoon kuuluvista arvioi sähköisten tutkimusaineistojen säilyttämisen ja käytön ylipäänsä koskevan laitostaan. Asiaa esitellään kyselyaineistojen keruuta koskevassa liitteessä 2.

Taulukko 1.  
Professorikyselyn vastaajien tieteenalat ja tieteenalaryhmittely rakenne (%)

Vastaajan edustama ala	Yhteiskuntatieteet	Kasvatustieteet	Kielitieteet	Muut hum. alat	Kaikki
Kansantaloustiede	6	–	–	–	2
Sosiaalitieteet	42	–	–	–	15
Psykologia	28	–	–	–	10
Kasvatustiede	–	100	–	–	14
Valtio-oppi, hallintotiede	19	–	–	–	7
Viestintä- ja informaatiotieteet	6	–	–	–	2
Filosofia	–	–	–	3	1
Kielitieteet	–	–	100	–	26
Taiteiden tutkimus, kirjallisuus	–	–	–	31	7
Historia ja arkeologia	–	–	–	22	5
Kulttuurien tutkimus	–	–	–	19	5
Muu ala	–	–	–	25	6
<b>Yhteensä (n)</b>	<b>100 (51)</b>	<b>100 (20)</b>	<b>100 (39)</b>	<b>100 (36)</b>	<b>100 (147)</b>

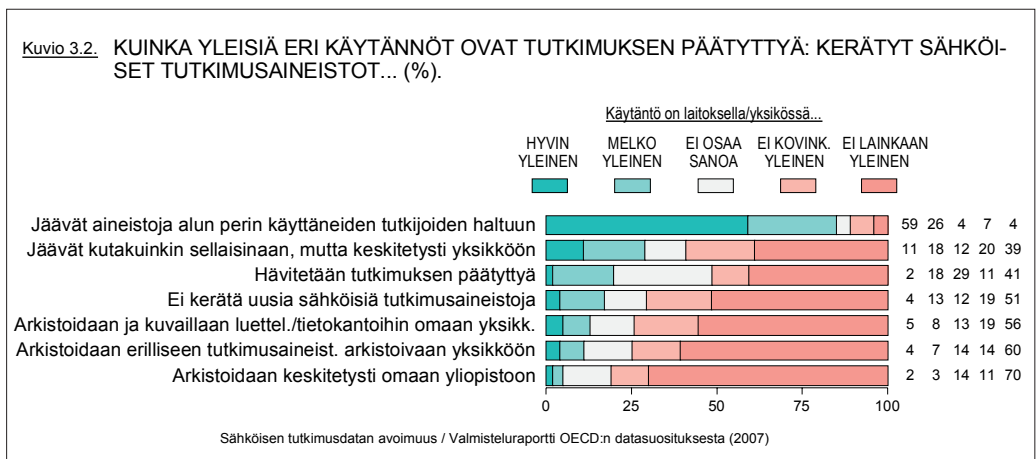
Kyselyn vastaajista noin 80 prosenttia oli joko Helsingin, Jyväskylän, Tampereen tai Turun yliopiston professoreita. Näiden lisäksi vastaajia oli kuudesta muusta yliopistosta. Vastanneista 39 prosenttia oli naisia. Taulukko 1 kertoo vastaajien tieteenalan ja sen miten tieteenalat on ryhmitelty tätä selvitystä varten. Professorikyselyn aineiston pienen koon vuoksi jatkossa esitellään lähinnä kokonaistuloksia. Nämä tarjoavat suuntaa antavan kuvan mainittujen alojen ja niillä toimivien yliopistolaitosten tilanteesta. Luultavasti tiedot ovat monin osin yleistettävissä myös monille muille tieteenaloille.

Sähköisten aineistojen nykyisiä säilytyskäytäntöjä tiedusteltiin muutamalla peruskysymyksellä. Ensimmäinen kuului yksinkertaisesti, ”onko laitoksella ohjeita tai sääntöjä sähköisten tutkimusaineistojen säilyttämiselle?” Peräti yhdeksän kymmenestä vastaajasta totesi, ettei heidän laitoksellaan ole asiasta mitään sääntöjä tai ohjeita (kuvio 3.1).



Vastauksissa ilmeni pientä tieteenaloittaista vaihtelua, mikä voi selittyä pienen vastaajamäärän aiheuttamasta satunnaisvaihtelusta. Muista aloista poikkesi hieman edukseen kielitieteet, jonka vastaajista 15 prosenttia totesi laitoksellaan olevan joitakin asiaan liittyviä sääntöjä tai ohjeita.

Tämän jälkeen vastaajille esitettiin toinen heidän laitoksensa vallitsevia käytäntöjä koskeva kysymys. Siinä vastaajat arvioivat seitsemää erilaista säilyttämistapaa tai asiantilaa siltä kannalta, kuinka yleisiä ne ovat heidän laitoksellaan tai yksikössään. Vastausvaihtoehdot vaihtelivat hyvin yleisestä käytännöstä ei lainkaan yleiseen käytäntöön. Kysymysosioista saadut tulokset on esitetty kuviossa 3.2.



Laskettaessa yhteen hyvin tai melko yleisen käytännön vastaukset tulokset vahvistavat sen yleisesti arvellun oletuksen, että yleensä tutkimusaineistot jäävät tutkijoiden itsensä haltuun (yht. 85 %). Tällöin aineistoja säilytettäneen enimmäkseen siten, että vain tutkijat itse voivat käyttää niitä. Kuitenkin vajaa kolmannes vastaajista (29 %) katsoi vähintään melko yleiseksi käytännöksi sen, että aineistot jäävät sellaisinaan, mutta keskitetysti, tutkijan laitokseen tai yksikköön. Vain joka kahdeksas sanoi, että heidän laitoksellaan sähköiset aineistot yleensä arkistoidaan ja kuvaillaan luetteloihin tai tietokantoihin. Lisäksi ainoastaan viitisen prosenttia vastaajista katsoi, että laitoksen sähköisiä aineistoja arkistoidaan keskitetysti omaan yliopistoon.

Myöskään aineistojen arkistointi erilliseen, tutkimusaineistoja arkistoiwaan yksikköön ei osoittautunut yleiseksi käytännöksi. Sitä piti yleisenä käytäntönä ainoastaan joka kymmenes vastaaja. Yhteiskuntatieteellisen tietoarkiston toiminta ja arkistointimahdollisuudet nostivat jonkin verran yhteiskuntatieteellisen alan lukuja, mutta vain vähän. Tälläkin alalla ainoastaan joka seitsemäs (16 %) piti tätä käytäntöä hyvin tai melko yleisenä. Tämä säilytystapa osoittautui harvinaisimmaksi kasvatustieteissä, jonka vastaajista vain yksi piti tapaa yleisenä alallaan. Myös kielitieteissä ja muilla humanistisilla aloilla, kuten kulttuurintutkimuksessa, onkin kasvatustieteistä poiketen joitakin valtakunnallisia keskitettyjä säilytysratkaisuja.

Kaikkiaan käytännöt viittaavat siihen, että sähköisten tutkimusaineistojen säilyttämisestä ohjeistetaan Suomen yliopisto-organisaatioissa joko vähän tai ei ollenkaan. Aineistot jäävät ylivoimaisesti useimmiten tutkijoille itselleen, eivätkä laitospohjaiset säilytysratkaisut ole kovin yleisiä. Tämä selittää osin myös sen, miksi sähköisiä aineistoja ei juurikaan hävitetä tutkimuksen päätyttyä, vaan aineistot jäävät aktiivikäytön jälkeen kokoajiansa haltuun ilman varsinaisia säilytysuunnitelmia.

Tutkimusaineistojen säilyttämistä ei olekaan yleensä ohjeistettu kovin näkyvästi. Arkistolaitoksen ylitarkastaja Markku Leppäsen (2006) mukaan Tieteellisen informoinnin neuvoston asettama Tutkimusdatajaosto pohti jo vuosina 1989–1991 tutkimusaineistojen säilyttämiseen, seulontaan ja myöhempään tutkimuskäyttöön liittyviä kysymyksiä. Se suositti, että pysyvästi säilytettävien data-aineistojen valinta lankeaa viime kädessä tutkimuksen harjoittajille. Jaosto esitti tieteenalakohtaisten data-arkistojen perustamista, aineistojen huolellista dokumentointia ja erityisen tutkimusaineistorekisterin perustamista. (Mt.) Tietoarkiston perustaminen 1990-luvun loppupuolella oli osin seurausta tästä prosessista.

Suomen arkistolaitos ei myöskään ole antanut yleistä määräystä tutkimustoiminnan asiakirjallisten tietojen ja tutkimusaineistojen pysyvästä säilytyksestä, vaan päätökset ovat olleet virasto- ja laitospohjaisia. Käytännössä yliopistotasolla annettava ohjeistus sisältyy yliopistojen arkistonmuodostussuunnitelmiin. Tavallisesti yliopistotason suunnitelmiin sisällytetään varsin tarkat ohjeet yleis-, talous- ja opintoasiain hallinnollisten asiakirjojen säilyttämisestä. Tutkimusaineistoille on toisinaan arkistonmuodostussuunnitelmissa määritetty 1–10 vuoden säilytysaika. Lopullinen aineiston arvonmääritys ja säilytysaika päätetään – tai tulisi päättää – laitostasolla tutkimuskohtaisesti.

Sopimusoikeudellisesti ja arkistolain perusteella yliopistojen arkistojen tulisi kattaa julkisin varoin toteutettujen tutkimushankkeiden asiakirjojen lisäksi julkisin varoin hankitut tutkimusaineistot, ellei muusta ole tutkimusta aloitettaessa erikseen tutkimussopimuksissa tai muutoin kirjallisesti sovittu. Käytännössä tutkimusaineistojen käsittely ja säilyttäminen on kuitenkin suunniteltava ja toteutettava laitostasolla.

Tutkimusaineiston arkistointi voi sisältyä yliopistolaitoksilla laadittaviin laatukäsikirjoihin. Vaikka käsikirjojen laatimisohejissa ei tavallisesti ole otettu huomioon tutkimusaineistoja, jotkut laitokset ovat sisällyttäneet aineistojen hallinnan osaksi laatutyötään. Esimerkiksi Kuopion yliopiston terveyshallinnon ja -talouden laitoksen laatukäsikirjassa (2006) kirjoitetaan tutkimusaineistoista seuraavasti: ”Aineiston arkistoinnissa tulee huomioida henkilötietolaissa olevat määräykset (rekisterinpitäjä, velvollisuudet: Henkilötietolaki 22.4.1999/523). Laitoksella on määritelty arkistointivastaava, jonka tehtäviin kuuluu huolehtia siitä että tutkimusaineisto säilytetään asianmukaisesti ja että se on dokumentoitu riittävästi. Arkistointivastaava terveyshallinnon ja -talouden laitoksella on tutkimussihteeri. Kunkin aineiston kohdalla harkitaan erikseen (tutkimusryhmä ja arkistointivastaava) säilytyksestä sekä siitä, voidaanko se luovuttaa yhteiskuntatieteelliseen tietoarkistoon.”

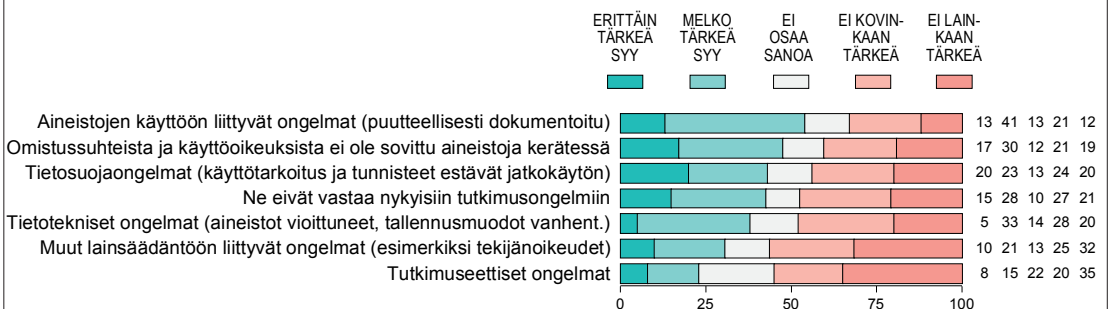
Positiivinen poikkeus tutkimusaineistojen keruun ja säilyttämisen ohjeistamisessa on myös Turun yliopiston kulttuurien tutkimuksen laitos. Perusopetukseen on sisällytetty aineistojen keruun, käsittelyn ja arkistoinnin perusteet. Lisäksi laitoksen verkkosivuilla on selkeät ohjeet aineistojen dokumentointiin, arkistointiin ja jatkokäyttöön (ks. Mahlamäki 2002a, 2002b ja 2002c sekä <http://kultut-arkistot.utu.fi/tkuaineistotalennus.htm>).

### 3.2 Keskeisimmät avoimuuden esteet ja haitat

OECD:n datajulistuksessa ja -suosituksessa pidetään sähköisten tutkimusaineistojen jatkokäyttötarvetta varsin itsestään selvänä. Näin varmaan onkin, jos aineistojen jatkokäytön tuki-infrastruktuurit ja ajatus tutkimusaineistojen avoimuudesta ovat jo vakiintuneet ja tiedeyhteisö tunnistaa avoimen jatkokäytön hyödyt. Suomessa näin ei vielä ole. Tämän vuoksi kyselyssä kartoitettiin kahdella kysymyspatterilla, miten suurina tai pieninä vastaajat kokevat erilaiset avoimuuden esteet ja haitat. Professoreilta tiedusteltiin eri syiden tärkeyttä sille, miksi jo päättyneiden tutkimusten sähköisiä aineistoja ei käytetä jatkotutkimuksissa. Toisella kysymyksellä puolestaan kartoitettiin mahdollisten avoimuuden haittojen määrää ja laatua.

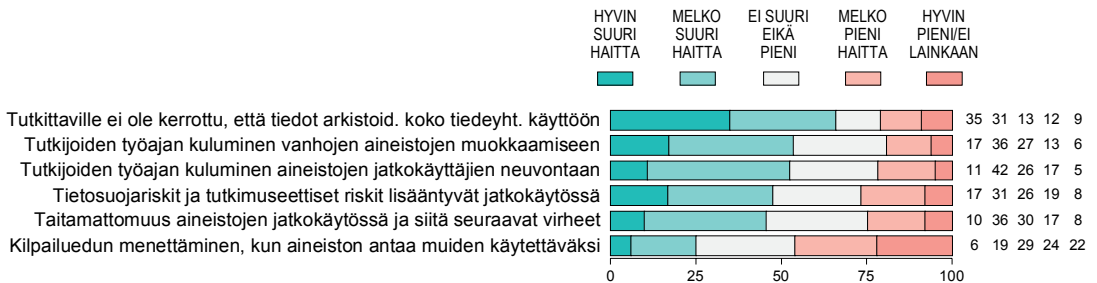
Kuviossa 3.3 on esitettyinä jatkokäytön esteitä kartoittavien kysymysten tulokset.

**Kuvio 3.3.** KUINKA TÄRKEITÄ ERI SYYT OVAT SILLE, ETTÄ JO PÄÄTTYNEIDEN TUTKIMUSTEN SÄHKÖISIÄ AINEISTOJA EI JATKOKÄYTETÄ OMALLA TUTKIMUSALALLA (%).



Yksi päätulos on, että todellisia esteitä on vastaajien mielestä useita eikä mikään tekijä nouse selvästi ylitse muiden. Jatkokäytön mahdollisia haittoja luodanseen kysymyksen tulokset on puolestaan esitelty kuviossa 3.4.

**Kuvio 3.4. KUINKA SUURINA TAI PIENINÄ PITÄÄ AVOIMUUDEN LISÄÄMISEN MAHDOLLISIA HAITTOJA ERI ASIOISSA (tekstejä osin lyhennetty, %).**



Sähköisen tutkimusdatan avoimuus / Valmisteluraportti OECD:n datasuosituksesta (2007)

Seuraavassa kuvioiden 3.3 ja 3.4 tuloksia kootaan aihepiireittäin viittaamalla myös alan tutkimuskirjallisuuteen sekä professorikyselyssä annettuun vapaamuotoiseen palautteeseen.

### 3.2.1 Taitamattomuus aineiston käytössä ja aineistojen soveltumattomuus uusiin tutkimusongelmiin

Kuvion 3.4 mukaan vastaajista lähes joka toinen (46 %) piti melko tai hyvin suurena haittana sitä, että jatkotutkimuksissa ei osata käyttää aineistoja oikein. Vastaus voi kertoa siitä, että tutkijat mieltävät vain itsensä riittävän päteviksi käyttämään aineistojaan. Ihmistieteissä tällaiset seikat koskevat yhtä hyvin kvantitatiivisia kuin laadullisiakin tutkimusaineistoja, joista jälkimmäiset ovat esimerkiksi teksti-, kuva- ja äänitiedostoja.

Jotkut aineistot voidaan myös ymmärtää sellaisiksi, että niitä voidaan analysoida väärin. Näin asiaa kommentoitiin kyselyssä:

*”Näkisin avoimuudessa suuria mahdollisuuksia aivan uudenlaisten kysymysten tutkimiseen. Toisaalta pulmia ei ole vaikea paikallistaa. Tuntuu painajaismaiselta ajatella, että ryhmäni datoihin voitaisiin vedota virheellisesti ja perustella näkökantaa, jota pidän paitsi vääränä, myös epäeettisenä.”*

Esimerkiksi laadullisen tutkimussuhteen erityisluonnetta voidaan käyttää perusteena sille, että kukaan muu ei aineistoa voi ymmärtää. Natasha Mauthner et al. (1998) korostavat, että laadullinen aineisto ei sovi arkistoaineistoksi, koska laadullisessa tutkimuksessa nojaututaan tulkinalliseen ja refleksiiviseen tutkimussuhteeseen. Vain alkuperäinen tutkija kykenee ymmärtämään aineistoa. (mt. 743.)

On totta, että aineistokeruutilanteessa mukana olevalla on aistein havaittavissa ja osin tulkittavissa tunteet, ilmeet, huudahdukset ja vastaavat. Inhimilliseen vuorovaikutukseen voi sisältyä myös piirteitä, joita on vaikea sanoa ilmaista. On kuitenkin hyvä muistaa, että aineiston kerää ja käsittelee usein joku muu kuin itse tutkija. Lisäksi myös tutkittavansa itse kohdanneet tutkijat työskentelevät pääasiassa tutkimusaineistoksi muokatun materiaalin parissa analyysiä tehdessään. Tutkimuksen konventioiden mukaisesti kaikki aineiston tulkinnot – myös autenttisisissa tilanteissa muodostuneet – on kyettävä ilmaisemaan ja perustelemaan tutkimusteksteissä sanallisesti.

Nigel Fielding (2000) ja Louise Corti (2006a) näkevät laadullisen aineiston jatkokäytön mahdollisuudet lähinnä käytännöllisenä, eivät epistemologisena kysymyksenä. Jotta tutkimusaineistoa voitaisiin käyttää jatkotutkimuksiin, on aineiston keruuta ja itse aineistoa kuvaava kontekstietä kirjattava kattavasti ja huolellisesti. Aineistojen jatkokäytön taitamattomuuden pelko voi kertoa myös siitä, että nykyisellään aineistoja ei dokumentoida jatkokäyttöä varten. Ilman dokumentointia aineistojen jatkokäyttö on vähintään puutteellista, pahimmillaan myös virheellistä.

Kuviossa 3.3 noin kaksi viidesosaa (43 %) vastaajista piti toisten kokoamien aineistojen soveltumattomuutta nykyisiin tutkimusongelmiin tärkeänä syynä sille, että niitä ei käytetä jatkotutkimuksissa. Myös kyselyn kommentteissa mainittiin, että aivan kaiken sähköisen aineiston valmistaminen arkistointikuntoon ei ole perusteltua, jos aineisto on kerätty hyvin spesifiin tarpeeseen.

Markku Leppänen (2006) on puolestaan todennut, että datan arkistointiin liittyvässä seulonassa ei voi käyttää mitään yksiselitteisiä kriteereitä sen määrittämiseen, mitkä aineistot on säilytettävä pysyvästi. Hän kuitenkin listaa muutamia keskeisiä seulonnan arviointikriteerejä: aineiston ainutkertaisuus, käytettävyys ja käyttörajoitukset, kustannustekijät samoin kuin sen yhteiskunnallinen, kulttuurinen, historiallinen ja tieteenhistoriallinen arvo. Lisäksi arkistointipäätöksen tekemiseen voi vaikuttaa aineiston käyttökelpoisuus esimerkiksi tutkimusmenetelmien opetuksessa.

### 3.2.2 Epäselvyydet aineiston omistajuudesta

Kuvion 3.3. mukaan vastaajista kolmannes (31 %) piti lainsäädäntöön, esimerkiksi tekijänoikeuteen liittyviä ongelmia tärkeänä tai melko tärkeänä syynä sille, ettei tutkimusaineistoja käytetä jatkotutkimuksissa. Liki puolet (47 %) katsoi tärkeäksi tai melko tärkeäksi syyksi aineistojen jatkokäytön vähäisyydelle sen, että aineistojen omistus- ja käyttösuhteista ei ole sovittu aineistoja kerätessä.

Tekijänoikeuksien, muiden immateriaalioikeuksien ja sopimusehtojen ohella – ja monilla aloilla itse asiassa usein niitä enemmän – tutkimusaineiston omistajuuteen vaikuttavat epäviralliset akateemiset käytännöt ja ajattelutavat. Tutkimuksen tekeminen on luovaa työtä tutkimussuunnitelman laatimisesta aina lopullisten tulosten kirjoittamiseen. Tutkimusaineiston suunnittelu, kerääminen ja käsittely ovat empiirisessä tutkimuksessa hyvin keskeinen osa tutkimusta. Aineiston saattaminen analysoitavaan muotoon voi sisältää monia vaativia työvaiheita. Niiden aikana tehdään tärkeitä teknisiä ja sisällöllisiä päätöksiä aineistosta. Siksi tutkijoille voi



olla vaikea ajatella, että huolella suunniteltu ja tehty aineisto ei lopulta olisikaan heidän itsensä tai tutkimusprojektin yksinomaisessa käytössä ja omistuksessa.

Usein aineisto mielletään osaksi sitä henkistä pääomaa, jota tutkimustyössä kartutetaan. Näin aineiston ajatellaan kuuluvan sen tehneelle tutkijalle tai tutkimusryhmälle ja käytännössä tämän mukaan myös toimitaan: tutkijat ottavat usein hankkimansa aineistot mukaansa työpaikkaa vaihtaessaan. Myös kyselyn tulokset kertovat, että aineistot mielletään käytännössä tutkijoiden yksityisomaisuudeksi. Oppiaineesta riippumatta pääasiallinen käytäntö on, että sähköiset aineistot jäävät aineistoja alun perin käyttäneiden tutkijoiden haltuun. Vastaajista 85 prosenttia ilmoittaa sen laitoksellaan hyvin tai melko yleiseksi käytännöksi.

Yleisellä tasolla *sopimusoikeudellinen pääsääntö kuitenkin on, että tutkimusaineisto jää tutkimuksen toteuttamispaikan (yleensä tutkimuslaitoksen) vastuulle ja hallintaan, ellei toisin ole sovittu*. Tästä poikkeuksena Suomen Akatemia suosittaa rahoittamiensa yhteiskuntatieteellisten tutkimushankkeiden tutkimusaineistojen arkistoinnista Yhteiskuntatieteelliseen tietoaarkistoon. Suosituksesta huolimatta tutkimusaineistojen kohtalosta ei tavallisesti sovita ennen tutkimusta eikä sen jälkeen. Epäselvyyttä aineistojen omistusoikeuksista eivät hälvännä ainakaan yliopistojen sopimuskäytännöt, sillä niihin sisältyy edelleen vain harvoin tutkimusaineistoa koskevia kohtia.

Usein tutkimusaineiston hallintaoikeus rinnastetaan tekijänoikeuteen. Tekijänoikeus ja hallinta- ja käyttöoikeus ovat kuitenkin eri asioita, mikä käy ilmi tekijänoikeusneuvoston lausunnoista. Yksi niistä käsitteli tutkijan valitusta koskien vanhan tutkimusaineiston siirtoa alan tutkimuslaitokseen ilman aineistoa keskeisesti suunnittelemassa olleen tutkijan lupaa. Asiasta lausuntoja antoivat itse tekijänoikeusneuvoston lisäksi muun muassa ympäristöministeriö ja Suomen Akatemia tehdyn tutkimuksen rahoittajina. Ympäristöministeriön mukaan aineiston omistusoikeus oli valtiolla ja tekijänoikeuksia useilla ihmisillä, koska tutkimusaineistoon sisältyi lukuisia joukko dia- ja valokuvia. Akatemia totesi puolestaan, että tutkimusmäärärahoilla hankitut laitteet ja muu tarpeisto sekä tutkimusaineisto jäävät valtion omistukseen ja kuuluvat yleensä sille korkeakoululle, jossa tutkimus on tehty. (Tekijänoikeusneuvoston lausunto 1999:15, 3-5.)

*Tekijänoikeudet tutkimusaineistoon eivät tavallisesti muutu, vaikka aineiston hallinta- ja käyttöoikeudet muuttuvat*. Esimerkiksi Yhteiskuntatieteelliseen tietoaarkistoon arkistoitujen aineistojen tekijänoikeudet säilyvät alkuperäisillä tutkijoilla. Aineiston hallinta sen sijaan siirtyy tietoaarkistolle, joka säätelee lupakäytännöllään myös jatkokäyttöä (ks. tarkemmin Kuula 2006, 68-75). Juridista tekijänoikeutta olennaisempia ovat käytännössä jo tutkimusten alussa tehtävät suunnitelmat ja sopimukset tutkimusaineistojen käsittelystä, säilyttämisestä ja jatkokäytöstä.

Tekijänoikeudet voivat olla kuitenkin toisinaan avoimuuden este, kun analysoitavana aineistona käytetään julkisia tekijänoikeussuojaa saavia sähköisiä aineistoja. Ongelmat voivat korostua esimerkiksi taiteentutkimuksessa ja kulttuurintutkimuksessa. Esimerkiksi sähköisiksi muunnetut lehtikirjoitukset, mainokset ja vastaavat eivät ole nykyisen tekijänoikeuslain mukaisesti arkistoitavissa erikseen tutkimuskäyttöä varten. Tähän ongelmaan toivottiin kyselyn kommentissa etsittävän ratkaisuja, sillä tekijänoikeudellinen suoja sulkee jo nyt pois tutkimusaiheita.



### 3.2.3 Akateeminen kilpailu

Joka neljäs professorikyselyn vastaajista näki kilpailuedun menettämisen tutkimusaineistojen avoimuudessa suurena tai melko suurena häittana (kuvio 3.4). Pieneksi haitaksi sen arvioi samoin neljännes ja 22 prosenttia piti sitä hyvin pienenä tai ei lainkaan haitallisena. Aihetta sivuttiin myös kyselyn kommentteissa:

*”Näkisin, että olemassa oleva verkottunut yhteistyö on ehto käyttämiselle. Minun on henkilökohtaisesti vaikea sopeutua ajatukseen avoimen valtakirjan antamisesta.”*

Akateeminen kilpailu on nostettu datan avoimuuden esteeksi useissa julkaisuissa (ks. Clubb et al. 1985, 57–58; Sieber 1991, 142–143). Vaivalla kerättyjä tutkimusaineistoja ei välttämättä haluta antaa kenen tahansa tutkijan hyödynnettäväksi. Erityisen hankalaksi saatetaan kokea ajatus siitä, että saman alan kilpatutkijat menestyisivät julkaisu- ja rahoituskilvassa aineistoilla, jotka itse on ensin tehnyt ja myöhemmin avoimeen käyttöön saattanut.

Toinen avoimuuden este voi olla pelko epäasiallisesta kritiikistä, jota aineiston jatkokäyttäjät voivat esittää alkuperäisen tutkimuksen tuloksista, mikäli pääsevät aineistoa tutkimaan. Tällainen pelko oletettavasti korostuu, mikäli tutkija on kokenut tai nähnyt epäreilua kilpailua. Tosin avoimuutta voi estää myös pelko siitä, että tosiasialliset tutkimuksen heikkoudet paljastuvat. Puutteellinen aineisto tai käytettyjen menetelmien virheet paljastuvat, mikäli joku tarkistaa epäilemiään tutkimuksen tuloksia aineistosta.

*Aineiston tekijöiden ensikäyttöoikeus omaan aineistoonsa on tunnustettu ja arvostettu oikeus.* Tutkimustulosten originaalisuus on tieteessä keskeinen arvo ja käytännössä kunnia uudesta tieteellisestä tuloksesta menee sen ensimmäiselle esittäjälle (Kiikeri ja Ylikoski 2004, 127–129.) Aineiston suunnitteleilla ja tehneillä tutkijoilla on myös oikeus kirjoittaa siitä keskeisimmät tulokset ilman pelkoa siitä, että joku muu tekee sen ensin. Tämä seikka nostettiin esille myös kyselyn avovastauksissa. Tutkijalla tulee olla oikeus tuottaa kokoamastaan aineistosta olennaisimmat tulokset. Niiden julkaisemisen jälkeen aineiston saattaminen toisten jatkokäyttöön tulee kuitenkin tieteen avoimuuden peruseriaatteita, joita ovat *objektiivisuus, kriittisyys, autonomisuus ja edistyvyys* (Niiniluoto 2002).

Muun muassa objektiivisuuteen ja kriittisyyteen liittyy mahdollisuus tarvittaessa tarkistaa epäselviä tuloksia empiirisestä aineistosta. Näin on siitakin huolimatta, että tutkimuksen toistaminen täsmälleen entisen kaltaisena ja samanlaisissa olosuhteissa ei tavallisesti ole mahdollista. Erityisen vaikeaa se on kvalitatiivisessa tutkimuksessa, mutta aivan suoraviivaista se ei ole kvantitatiivisessakaan tutkimuksessa (Ray ja Valeriano 2003, 79).

Vaikka kvantitatiivisessa tutkimuksessa ei eläydyttäisikään vastaajien kokemukseen, siinäkin tulee tulkita ja ymmärtää selitettävää todellisuutta. Lisäksi analyysi voi sisältää niin monia teknisiä ja analyttisiä vaiheita, että tutkimuksen toistaminen täysin samanlaisena ei käytännössä ole mahdollista. Silti tutkimusaineistojen avoimuus voi parantaa tutkimuksen laatua. Jo tieto siitä, että tuloksia voidaan tarvittaessa tarkistaa empiirisestä aineistosta, pakottaa systemaattisuuteen ja huolellisuuteen aineiston analyysissä.

Yhtä olennaista on antaa muille tutkijoille mahdollisuus tehdä uudenlaisia analyysejä samasta empiirisestä aineistosta. Periaate on olennainen osa tieteen edistävyyttä, ja kokemukset osoittavat, että juuri sen merkitys korostuu tutkimusaineistojen avoimuuden myötä. Esi-

merkiksi Paul Thompsonin 1970-luvulla keräämä laaja laadullinen haastatteluaiaineisto briteistä (The Edwardians: Family Life and Work Experience Before 1918) on sähköiseen muotoon saatettuna ollut käytössä jo yli sadassa jatkotutkimuksessa. Aineistosta tehdään edelleen useita hyvin erilaisia tutkimuksia, sillä se sisältää monipuolista ja ainutkertaista tietoa teollistumisen vaikutuksista perhe-elämään. (Corti 2006b.) Thompsonin aineisto on erinomainen esimerkki siitä, miten tieteen kollektiivisen luonteen ulottaminen myös tutkimusaineistoihin tukee uuden tieteellisen tiedon tuottamista. Yhden tutkijan hallussa pysyvistä samasta aineistosta ei koskaan olisi voitu tehdä niin erilaisia ja lukuisia tutkimusjulkaisuja. Tieteen edistävyyttä ja kollektiivista luonnetta kommentoitiin myös kyselyn lopussa:

*”Jos tutkija kerää tietoa vain itselleen, hän on kuin säästöpossu, jota ei koskaan tyhjennetä. Todellinen, luotettava tieto lisääntyy, kun sitä jaetaan ja arvioidaan avoimesti ja kriittisesti.”*

*”Itselläni on pelkästään hyviä kokemuksia siitä, että olen omia laajojakin aineistoja antanut muiden tutkijoiden käytettäväksi – siitäkkin huolimatta että niitä itsekin käytän. Kun jotain antaa, niin yleensä jotain myös saa!!”*

Oman aineiston antaminen avoimeen käyttöön voi olla menestystekijä myös tieteellisessä kilpailussa. Gleditsch, Metelis ja Strand (2003) ovat tutkineet kansainvälisen politiikan tutkimuksen alan vuosina 1999–2001 julkaistuihin artikkeleihin viittaamista. Artikkeleihin, joiden tutkimusaineistot oli saatettu avoimesti saataville, viitattiin kaksi kertaa useammin kuin artikkeleihin, joiden aineistoa ei ollut saatavilla. Ray ja Valeriano (2003) puolestaan huomauttavat, että vain joka kymmenenteen kaikista heidän alansa (kansainvälinen politiikka) artikkeleista ylipäänsä viitataan. Tämän vuoksi he näkevät tutkijoiden mahdollisuuden tulla viitatuksi aineistonsa kautta pelkästään positiivisena seikkana. Datan julkaiseminen ja saattaminen avoimeen käyttöön on osoitus siitä, että ensikäyttäjät on valmistanut käyttökelpoisen ja tutkimuksellisesti kiinnostavan aineiston (mt. 84). Sieberin (1991) mukaan rahoitushakemuksissakin on vain (kilpailu)etua siitä, että tutkimuksessa kerättävän aineiston voidaan ennakoida olevan käytettävissä monipuolisesti erilaisissa jatkotutkimuksissa.

### 3.2.4 Aineistojen tietotekniset ja dokumentoinnin puutteet

Iso osa aineistoista jää käyttökeltottomiksi, kun niitä ei alun alkaen järjestetä ja dokumentoida arkistointia ja jatkotutkimuksia varten. Tämä näkyy myös kyselyn tuloksissa. Kyselyyn vastanneet professorit kokivat jatkokäytön suurimmaksi esteeksi sen, että aineistojen tietosisällöt ja tiedostot ovat puutteellisesti dokumentoituja ja järjestettyjä. Tätä piti tärkeänä tai melko tärkeänä syynä peräti 54 prosenttia kaikista vastaajista (kuvio 3.3). Sähköisiä aineistoja voi tutkimuslaitoksilla olla paljonkin. Aineistoilla ei silti välttämättä ole minkäänlaista käyttöarvoa, jos niitä ei ole järjestetty ja dokumentoitu mahdollista jatkokäyttöä silmällä pitäen.

Aineistojen jatkokäyttökuntoon saattamisen ongelmat ilmenevät myös vastaajien arvioissa avoimuuden mahdollisia haittoja. Vastaajista 53 prosenttia pitää melko tai hyvin suurena aineistojen avoimuuden haittana sitä, että tutkijoiden työaika meni vanhojen aineistojen muokkaamiseen käyttökelpoiksi (kuvio 3.4). Yhtä moni arvioi suureksi haitaksi jatkokäyttä-

jien neuvontaan mahdollisesti kuluvan työajan. Vastauksissa korostuu vastaajien työn kuormittavuus ja pelko siitä, että aineistojen avoimuus lisää ylimääräistä työtä. Sama asia otettiin esille myös kyselyn loppukommenteissa:

*”Pelkona on se, että tutkijoiden ja laitosten hallintotyö jälleen (!) kasvaa tämän prosessin myötä. Tehtävä on annettava jonnekin muualle kuin kuormitetulle laitostasolle. Tutkijoita on kuitenkin ehdottomasti konsultoitava, sillä he ovat aineistojensa parhaat asiantuntijat.”*

*”Suurin ongelma olisi ehkä se, että tutkijoilta ja/tai tutkimusvastaavilta kuluu paljon aikaa ja vaivaa tämän organisoimiseen ja toteuttamiseen.”*

Tietotekniset ongelmat estävät aineistojen jatkokäyttöä, kun siihen ei ole suunnitelmallisesti varauduttu. Erityisesti sähköiset tallenteet vanhenevat nopeasti. Rungas kolmannes vastaajista (38 %) piti erittäin tai melko tärkeänä syynä jatkokäytön vähäisyydelle aineistojen vioittuneisuutta tai vanhentunutta tallennusmuotoa tai sovellusohjelmaa (kuvio 3.3). Nopea tietotekninen kehitys muodostuukin usein ongelmaksi, vaikka muutoin aineistojen avoimeen jatkokäyttöön suhtauduttaisiin myönteisesti. Ongelmia aiheuttaa myös anonymisointi, mikäli se on edellytyksenä aineiston jatkokäytölle.

Kun aineisto säilytetään mahdollista jatkokäyttöä varten, on käytännössä ratkaistava, missä ja mitkä tiedot aineistosta ovat saatavilla ja missä muodossa aineisto säilytetään (vrt. liite 4). Molemmat asiat ovat hyvin oleellisia riippumatta siitä, arkistoidaanko aineisto omaan tutkimuslaitokseen vai erilliseen tutkimusaineistojen arkistointiin. Aikanaan arvokkaiksi määrittelyistä ja siksi säilytetyistä aineistoista ei ole juurikaan tutkimukselle iloa, jos perustietoja aineistoista ja niiden säilytyspaikoista ei ole saatavilla edes laitoksen omille tutkijoille. Jos tiedot aineistoista jäävät vain ensikäyttäjille, häviävät tiedon palaset sitä mukaa kun kyseiset henkilöt siirtyvät muihin tehtäviin tai eläkkeelle.

Dokumentoinnin ohella keskeinen käytettävyysongelma on ohjelmistojen vaihtuessa esiin nouseva sähköisten aineistojen konvertoinnin tarve. Dokumentointi ja konvertointi edellyttävät taitoja, oikeanlaisia työvälineitä ja ohjeita. Siksi erityisesti vanhojen aineistojen saattaminen avoimeksi on hyvin haasteellista ja toisinaan mahdollista vain erillisten lisäresurssien turvin. Kysely nosti professorikunnassa esille myös pelkoja siitä, että aineistojen avoimuus ylhäältä annettuna määräyksenä voidaan pahimmillaan kokea vain ylimääräisenä hallinnollisena ja teknisenä työnä. Siksi korostettiin tarvetta luoda pelisääntöjä ja niiden mukaisia täsmällisiä ohjeita ja tietoteknisiä välineitä avoimuuden toteuttamiseen käytännössä.

*”Hyvä vaan, että avoimuutta lisätään. Mutta myös ohjeistus ja taltiointisäännöt ovat paikallaan.”*

*”Många materialinsamlingar som enskilda forskare gjort kan vara sådana att forskaren själv är den enda som förstår vad materialet står för och hur det skall hanteras. Om man vill göra det tillgängligt för andra bör man också följa gemensamma regler vid kodning och redovisning av materialets fördelar och nackdelar.”*

*”Kyse on perustavaa laatua olevasta muutoksesta, jonka on ulotuttava tutkijoiden asenteisiin. Tällä hetkellä asia koetaan todennäköisesti ylimääräisenä työnä, joka muutenkin aikaansa paljoon muuhun kuin tutkimukseen käyttävältä tutkijalta vielä vaaditaan. Kestänee sukupolven verran, ennen kuin tästä saadaan luonnollisesti toimiva käytäntö, jos sitä nyt lähdetään tosissaan viemään eteenpäin.”*

### 3.2.5 Tietosuojakysymykset ja aineistojen luottamuksellisuus

Myös eettiset tietosuojakysymykset ovat erittäin keskeisiä tutkimusaineistojen avoimen saatavuuden kannalta. Asia korostuu tilanteissa, joissa tutkittaville ei kerrota aineiston säilyttämisestä jatkotutkimuksia varten. Aihepiirin yleinen merkitys ilmenee myös kyselyn tuloksissa ja erityisesti se korostuu kohdassa, jossa vastaajat tarkensivat omin sanoin eettisiä avoimuuden ongelmia.

Kaksi kolmesta kyselyyn osallistuneesta professorista (66 %) mainitsi avoimuuden lisäämisen merkittävänä ongelmana tapaukset, joissa tutkittaville ei ole kerrottu, että heiltä kerättävät tiedot arkistoidaan tutkimuksen päätyttyä koko tiedeyhteisön käyttöön (kuvio 3.4). Avoimuuden lisäämisen haitoista se koettiin suurimmaksi, mutta toisaalta asia on luonteeltaan hieman erilainen kuin kuviossa 3.4 listatut muut haittatekijät. Lähes puolet vastaajista arvioi myös tietosuojariskien ja tutkimuseettisten riskien lisääntyvän aineistojen avoimuuden myötä.

Mikäli tutkimusaineisto sisältää henkilötunneita, sen säilyttämiseen jatkokäyttöä varten on sovellettava *henkilötietolakia*. Sen mukaisesti yksilöityyn tutkimustarkoitukseen – esimerkiksi johonkin projektiin – kerättyä henkilötietoja sisältävää tutkimusaineistoa ei voi käyttää muihin tutkimuksiin, jos tutkittaville on kerrottu aineistoa käytettävän vain kyseiseen tutkimukseen. Kirjaimellisesti lakia noudattaen tunnistelliset tutkimusaineistot on tuhottava, kun siitä tehtyjen tieteellisten tulosten oikeellisuus on varmistettu. Ensikäyttövaiheen jälkeen aineistojen tieteellisen jatkokäytön voi turvata anonymisoimalla aineistot soveltuvin osin.

Erityisesti sairaanhoitopiirien eettiset toimikunnat voivat edellyttää tutkimussuostumuksia, joissa luvataan, että kerättävän tunnistellisen aineiston käyttö rajataan *vain nimettyyn tutkimukseen*. Tämä käytäntö kerrottiin avoimuuden esteeksi myös muutamissa professorikyselyyn vastanneiden kommentteissa. Eettiset toimikunnat todennäköisesti nojaavat ohjeissaan henkilötietolain käytötarkoitussidonnaisuuteen soveltaen sitä tiukasti.

Käytötarkoitussidonnaisuus voi tarkoittaa myös aineiston käytön rajaamista aineiston tieteelliseen tutkimukseen nyt ja tulevaisuudessa. Olennaisinta on, että tutkimustarkoitukseen kerättyjä tietoja ei saa luovuttaa mihinkään muuhun tarkoitukseen, esimerkiksi viranomaisille tai muille tahoille, jotka tekevät tutkimukseen osallistuvista ihmisistä päätöksiä. On myös varmistettava, että aineistoja ei käsitellä ja säilytetä huolimattomasti ja että tutkittavista (tutkimushenkilöistä, koehenkilöistä) ja heidän yksityisasiastaan ei puhuta lavertelevasti tai muutoin epäasiallisesti. Tutkittavista voi puhua ja kirjoittaa vain tutkimustarkoituksessa ja silloinkin ei-tunnisteellisesti aina kun se on mahdollista.

Toisinaan luottamuksellisuus rinnastetaan salassapitoon. Professorikyselyn tarkentavissa kommentteissa asia kerrottiin niin, että tutkittaville on luvattu, että vain nimetty tutkija käyttää aineistoa. Tarkkaan ottaen tutkimusaineistojen luottamuksellisuus ei kuitenkaan ole sama asia kuin salassapito. Luottamuksellisuus tutkimusaineistoista puhuttaessa viittaa yksittäisiä

ihmisiä koskeviin *tietoihin* ja näiden tietojen *käytöstä annettuihin lupauksiin*. Lupaukset aineiston käyttötavoista, käyttäjistä, käyttöajasta, muokkaustavoista ja säilyttämisestä annetaan tutkittaville jo tehtäessä sopimusta tutkimukseen osallistumisesta. Luottamus tutkimusaineiston yhteydessä tarkoittaa ensisijaisesti tutkittavan luottamusta siihen, että aineistoa käytetään, käsitellään ja säilytetään, kuten on sovittu. Käytännössä luottamuksellinen on myös tutkimusaineisto, joka arkistoidaan data-arkistoon. Sen luottamuksellisuus tarkoittaa tarkoin määrättyjä aineiston lupakäytäntöjä ja käyttöehtoja.

Yleensä ottaen tutkittavat suhtautuvat hyvin positiivisesti aineiston arkistointiin tieteellistä jatkokäyttöä varten. Onhan jo esimerkiksi haastatteluun osallistuminen osoitus halukkuudesta edesauttaa aihepiirin tutkimusta. Yhteiskuntatieteellisessä tietoaarkistossa on muutamien tutkimusten tutkittavilta kysytty jälkikäteen lupaa aineiston arkistointiin tilanteessa, joissa tutkija on alun perin luvannut vain itse käyttää aineistoa. Lähes kaikki ovat antaneet luvan arkistoinnille tieteellistä jatkotutkimusta ja opetuskäyttöä varten. Valtaosa perustelee antamaansa arkistointilupaa sillä, että he ovat nähneet osallistumisen kulloisenkin aihepiirin tutkimiseen mielekkäänä jo alun perin ja näkevät aineiston arkistoinnin tiedeyhteisön käyttöön yksinomaan myönteisenä seikkana.

Tutkittaville annettava informaatio saattaa ratkaista sen, haluavatko ihmiset osallistua tutkimukseen ja se ratkaisee myös tunnisteellisen aineiston kohtalon. Tutkittaville annettu informaatio yhdessä aineiston sisällön kanssa määrittävät sen, voiko aineiston arkistoida vai onko aineisto kenties hävitettävä heti tutkimuksen tulosten oikeellisuuden tarkastamisen jälkeen. Aineiston luottamuksellisuudesta tulisi voida sopia tutkittavien kanssa niin, että aineisto on tutkimuksen päätyttyä arkistoitavissa tieteellistä jatkokäyttöä varten.

Launiksen (2007, 31) mukaan eettistä arvokeskustelua edellytetään edelleen sen ratkaisemiseen, mikä muutos tieteellisessä käyttötarkoituksessa on sellainen, että tutkittavilta täytyy kysyä erikseen uusi lupa jatkotutkimuskäyttöä varten. Tutkittavien informoinnin lähtökohdaksi tulisi ottaa *koko aineiston elinkaaren suunnittelu*, jotta tietosuojakäytökset eivät estäisi aineistojen tieteellistä jatkokäyttöä. Myös tähän seikkaan tarvitaan yhteisiä pelisääntöjä ja ohjausta.

## 4. Aineistojen jatkokäytön nykytilanne

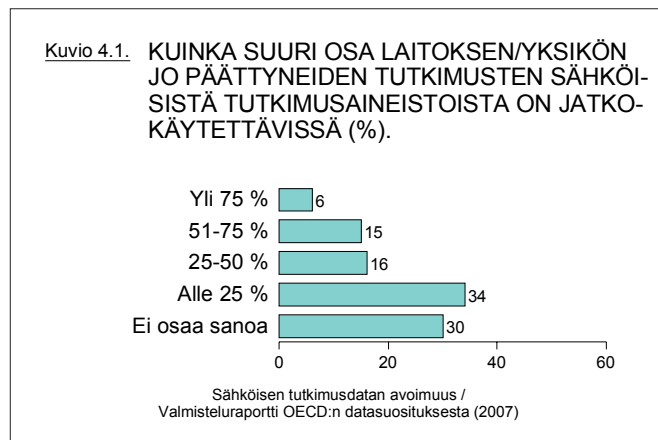
On kutakuinkin mahdotonta arvioida, missä määrin eri alojen ja organisaatioiden suomalaiset, julkisrahoitteisesti kootut sähköiset tutkimusaineistot olisivat tarvittaessa jatkokäytettävissä. Yliopistoissa, valtion virastoissa ja tutkimuslaitoksissa sekä kuntasektorilla on oletettavasti kymmeniä tuhansia sähköisiä tutkimusdatoja. Vain pieni osa niistä on suunnitelmallisen pitkäaikaissäilytyksen ja/tai aktiivisen jatkokäytön piirissä.<sup>6</sup>

Ilman huolellista säilytystä ja dokumentointia sähköiset tutkimusaineistot kuitenkin yleensä vanhenevat käyttökelvottomiksi 5–10 vuoden kuluessa kokoamisesta. Vaikka läheskään kaikkia julkisrahoitteisia tutkimusaineistoja ei pidä säilyttää, nykyistä huomattavasti useammat olisi järkevää saattaa tehokkaammin avoimen jatkokäytön piiriin.

Monet edellisessä luvussa esitetyistä avoimuuden esteistä lienevät yleistettävissä yliopiston ohella muidenkin tutkimusorganisaatioiden tilanteeseen. Seuraavaksi käsitellään suhtautumista tutkimusaineistojen avoimuuteen ja jatkokäytön etuihin. Asiaa tarkastellaan nyt tutkijoiden ja toimintayksikköjen näkökulmasta vain niiden akateemisten alojen ja toimijoiden osalta, jotka olivat professorikyselyn kohteena. Lisäksi esitellään joitakin tietoja yliopistojen, tutkimusrahoittajien ja valtion tutkimuslaitosten suhtautumisesta tutkimusaineistojen avoimuuteen.

### 4.1 Aineistojen käyttö ja käytettävyys

Kulttuurin ja yhteiskunnan tutkimusalojen professorikyselyyn vastanneet arvioivat oman laitoksensa aineistojen jatkokäyttöä useissa eri kysymyksissä. Yksi näistä koski sitä, kuinka suuri osa heidän laitoksensa tai yksikkönsä jo päättyneiden tutkimusten sähköisistä aineistoista on (vastaajan oman arvion mukaan) realistisesti ottaen jatkokäytettävissä. Kysymys koski viimeksi kuluneiden kymmenen vuoden aikana kertyneitä aineistoja. Tulokset on eritelty kuviossa 4.1.



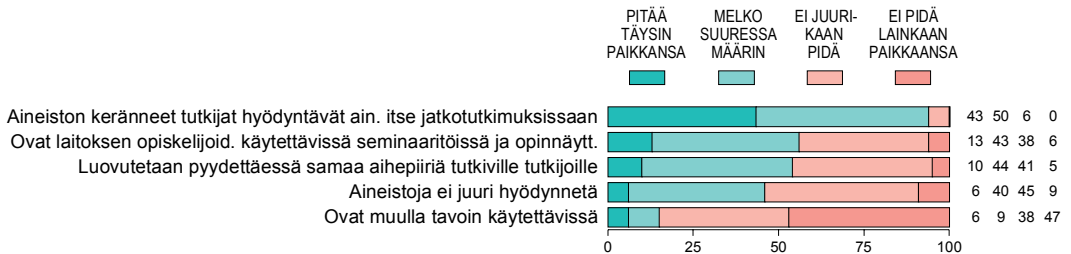
<sup>6</sup> Esimerkiksi Yhteiskuntatieteellisen tietoarkiston aineistovarannossa on vajaat 800 data-aineistoa (toukokuu 2007).

Vain joka viides vastaaja (21 %) oli sitä mieltä, että heidän laitoksensa sähköisistä tutkimusaineistoista vähintään puolet olisi edelleen jatkokäytettävissä. Koska samalta laitokselta saattaa olla useita vastaajia, prosenttiluvut eivät tässä viittaa aivan tarkkaan laitosten lukumääriin. Lisäksi ”en osaa sanoa”-vastausten osuus nousi melko suureksi.

Vastauksista voi päätellä, että *valtaosalla laitoksista useimmat kymmenen vuoden aikana kertyneistä aineistoista eivät ole enää jatkokäytettävissä*. Vaikka asian syitä ei tiedusteltu kyselyssä tarkemmin, edellisessä luvussa käsitellyt jatkokäytön esteet viittaavat todennäköisimmiksi syiksi aineistojen teknisen vanhenemisen tai tuhoutumisen sekä dokumentoinnin puutteen.

Lähes kaikkien kyselyyn vastanneiden laitoksilla käytetään kuitenkin joitakin sähköisiä aineistoja vaihteleviin jatkokäyttötarkoituksiin. Näitä teemoja kartoitettiin väittämillä, joiden paikkansapitävyyttä professorit arvioivat neliportaisella skaalalla (kuvio 4.2).

**Kuvio 4.2. MISSÄ MÄÄRIN SEURAAVAT VÄITTEET PITÄVÄT YLEISESTI OTTAEN PAIKKANSA LAITOKSELLAYKSIKÖSSÄ: TUTKIMUKSEN PÄÄTTYTTYÄ... (%)**



Sähköisen tutkimusdatan avoimuus / Valmisteluraportti OECD:n datasuosituksesta (2007)

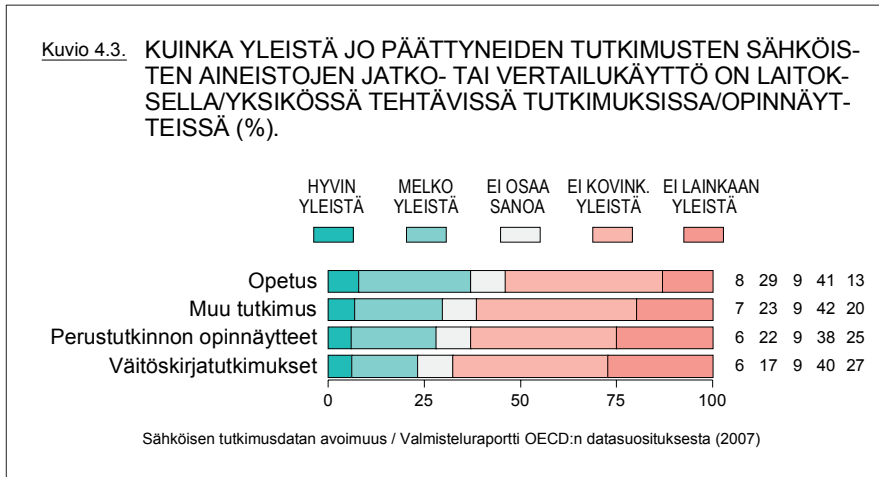
Kuvion 4.2 tulokset osoittavat odotetusti, että useimmiten jatkokäyttäjät ovat ensikäyttäjät itse. He hyödyntävät datojaan myöhemmissä tutkimuksissaan. Kaksi viidestä vastaajasta (43 %) oli sitä mieltä, että tätä asiaa koskeva väite pitää täysin paikkansa. Samalla lähes kaikki muut vastaajat sanoivat, että se pitää melko suuressa määrin paikkansa.

Kolme muuta aineistojen käyttöä koskevaa väitettä ei saanut taakseen läheskään näin varauksetonta samanmielisyyttä. Vain noin joka kymmenes professori katsoi täysin pitäväksi väitteen siitä, että heidän laitoksellaan tai yksikössään tutkimusdatoja luovutetaan pyydettyessä samaa aihepiiriä tutkiville tutkijoille. Tosin yli 40 prosenttia oli sitä mieltä, että väite pitää paikkansa melko suuressa määrin.

Entä ovatko laitoksen tutkimusaineistot sitten edes laitoksen omien opiskelijoiden käytettävissä esimerkiksi seminaaritöissä tai opinnäytteissä? Tämän katsoi pitävän täysin paikkansa 13 prosenttia vastaajista ja 43 prosenttia katsoi sen pitävän paikkansa melko suuressa määrin. Toisaalta lähes joka toisella laitoksella tällaisia käytäntöjä ei näytä olevan luuri lainkaan. Tulosta varmentaa havainto siitä, että *joka toinen vastaaja katsoi pitävän paikkansa, ettei sähköisiä aineistoja juurikaan hyödynnetä heidän laitoksellaan tutkimusten päätyttyä*.



Näitä tietoja täydentävät tulokset siitä, kuinka yleistä jo päättyneiden tutkimusten sähköisten aineistojen jatko- tai vertailukäyttö on opetuksessa sekä laitoksilla tehtävissä tutkimuksissa ja opinnäytteissä (kuvio 4.3).



Asiaa tiedusteltiin erikseen opetuksen, perustutkinnon opinnäytteiden, väitöskirjatutkimusten ja muun tutkimuksen osalta. Suosituimmaksi osoittautui opetusikäyttö, jota yli kolmannes vastaajista piti laitoksellaan hyvin tai melko yleisenä käytäntönä. Tosin hyvin yleistä tämäkin oli ainoastaan kahdeksassa prosentissa vastauksista.

Muutkaan sähköisten aineistojen jatkokäyttötarkoitukset eivät olleet tätä useammin hyvin yleisiä. Kun mukaan lasketaan ”melko yleistä” -vastaukset, noin joka neljännen vastaajan mielestä heidän laitoksellaan aineistoja jatkokäytetään yleisesti väitöskirjoihin, muuhun tutkimukseen ja myös perustutkinnon opinnäytteisiin.

Tuloksia tulkittaessa on pidettävä mielessä, että sähköisiä aineistoja hyödyntävät tutkijat ja laitokset lienevät yliedustettuja vastaajajoukossa. Täten nyt raportoidut käytön ja käytettävyyden tasot tuskin antavat ainakaan todellisuutta karumpaa kuvaa yliopistolaitosten kokonaistilanteesta. Aineistoja kyllä jatkokäytetään tutkimuksessa ja opetuksessa eri tavoin, mutta vallitseva toimintakulttuuri ei tue riittävästi jatkokäytön mahdollisuuksia. *Yleisintä on, että tutkijat käyttävät itse, ja ilmeisesti nimenomaan vain itse, kokoamiaan sähköisiä tutkimusaineistoja.*

## 4.2 Tutkijoiden suhtautuminen avoimuuden hyötyihin

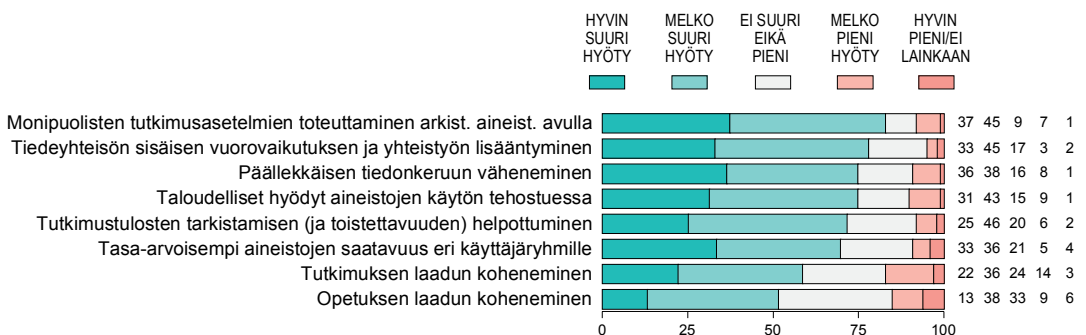
Millä tavoin tutkimusdatan avoimuutta olisi sitten mahdollista lisätä ja mitä etuja siitä koituisi? OECD:n datajulistus ja -suositus ottavat asiaan kantaa melko yleisellä ja periaatteellisella tasolla. Edellä on jo käynyt ilmi, että tutkittujen alojen professorit tunnistavat monia syitä olla lisäämättä tutkimusaineistojen avoimuutta. Niistä osa liittyi tutkijan ajankäytön rajoitteisiin ja



osa huoliin tutkimusdatan asianmukaisesta jatkokäytöstä. Monet olivat huolissaan myös tietosuojakysymysten hallinnasta.

Toisaalta julkisrahoitteisten sähköisten tutkimusaineistojen avoimuuden lisäämisellä voidaan edistää datan taloudellista ja muulla tavoin tehokasta käyttöä tilanteessa, jossa niukat tutkimusresurssit pakottavat rationalisoimaan uusien aineistojen keruuta. Kuviossa 4.4 listataan professorikyselyn vastaajien mielipiteitä tästä asiasta sekä seitsemästä muusta avoimuuden lisäämisen mahdollisesta hyödyistä.

Kuvio 4.4. KUINKA SUURINA TAI PIENINÄ PITÄÄ AVOIMUUDEN LISÄÄMISEN HYÖTYJÄ ERI ASIOISSA (%).



Sähköisen tutkimusdatan avoimuus / Valmisteluraportti OECD:n datasuosituksesta (2007)

Kaikki tiedustellut edut osoittautuivat sellaisiksi, että enemmistö vastaajista piti niitä viisiporraisella arviointiskaalalla joko hyvin tai melko suurina. Toisessa ääripäässä vastausvaihtoehdot 'melko pieni hyöty' ja 'hyvin pieni hyöty tai ei lainkaan hyötyä' kokosivat yhteen laskienkin useimmissa kohdissa vain alle 10 prosenttia vastauksista. Täten *tutkimusdatan avoimuuden hyödyt arvioitiin kauttaaltaan avoimuuden haittoja suuremmiksi* (vrt. kuvio haitoista 3.4).

Suurimmaksi hyödyksi kulttuurin ja yhteiskunnan tutkimuksen alan professorit arvioivat monipuolisten tutkimusasetelmien toteutumisen aineistojen jatkokäytön tuella. Yli 80 prosenttia vastaajista piti tätä vähintään melko suurena hyötynä. Sama argumentti nousee nykyisin esille usein myös luonnontieteitä, lääketiedettä ja teknisiäkin tieteitä koskevissa visioissa. Muiden muassa akatemiaprofessori Risto Nieminen (2007) on nostanut esiin idean verkottuneesta ja yhdistelevään datan käyttöön perustuvasta tieteestä. Hän kirjoittaa:

”Termillä ’Web 2.0’ on ryhdytty kuvaamaan teknologioita, sovelluksia ja liiketoimintamalleja, jotka ovat tehneet mahdolliseksi sellaiset Internetin menestystarinat kuin Google, Amazon, Skype, eBay, YouTube ja iTunes. ’Tiede 2.0’ sopii yleisnimeksi niille uusille tavoille tehdä tutkimusta, jotka hyödyntävät Internetiä, sen datavarantoja ja palveluita, virtuaalisista observatorioista proteiinitietokantoihin.

Tiede 2.0 on neljäs paradigma tutkimustyön historiassa. Parituhatta vuotta sitten tiede perustui luonnosta tehtyihin havaintoihin. Muutama sata vuotta sitten mukaan tulivat ensin systemaattiset kokeet ja sitten teoriat, jotka mahdollistivat tulosten systematisoinnin

ja ennusteet. Laskennallinen tiede ja tutkimus, kolmas paradigma, on vakiinnuttanut asemansa viime vuosikymmeninä. Se mahdollistaa monimutkaisten järjestelmien, kuten maapallon ilmaston, numeerisen simuloinnin. Neljäs paradigma on tutkimusmatkailu verkon välityksellä datan valtameressä. Dataa tuotetaan havainnoimalla, mittaamalla ja simuloimalla. Paradigma Tiede 2.0 siis yhdistää kolme muuta. Siitä käytetään myös muodikkaasta englanninkielistä nimeä eScience.”

Kulttuurin ja yhteiskunnan tutkimuksen alojen professorikyselyssä datan avoimuuden hyödyiksi tunnistettiin monia muitakin seikkoja. Listalta löytyi neljä myönteistä puolta, jotka noin joka kolmas vastaaja arvioi hyvin suureksi. Nämä olivat päällekkäisen tiedonkeruun väheneminen, taloudelliset hyödyt aineistojen jatkokäytön tehostuessa, tiedeyhteisön sisäisen vuorovaikutuksen ja yhteistyön lisääntyminen sekä tasa-arvoisempi aineistojen saatavuus eri käyttäjärhyille. Suhteellisesti vähiten suurta hyötyä avoimuudesta katsottiin koituvan tutkimuksen ja opetuksen laadulle, mutta näissäkin kohdin enemmistö vastaajista piti hyötyjä vähintään melko suurina.

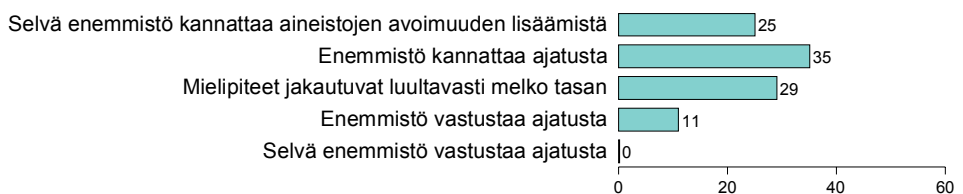
### 4.3 Tutkijoiden suhtautuminen jatkokäyttömahdollisuuksien lisäämiseen

Avoimuuden haittoja ja hyötyjä koskevat mielipiteet eivät ole keskenään ristiriitaisia. Jokainen on helppo tunnistaa sekä avoimuuden hyötyjä että mahdollisia haittoja. Aineistoja koonnut tutkija saattaa myös suhtautua penseästi omien aineistojensa jatkokäyttöön muiden tutkimuksissa, mutta samalla kannattaa periaatteellisesti tutkimusdatan avoimen käytön lisäämistä. Yleisen ja omakohtaisen suhtautumisen selvittämiseksi professoreilta tiedusteltiin

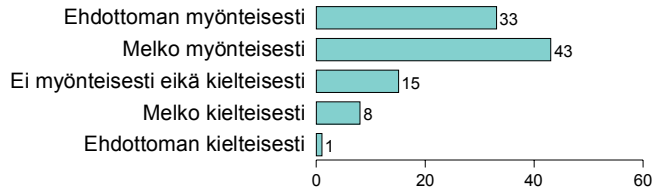
- miten he arvioivat oman alansa tutkijoiden (keskimäärin) suhtautuvan julkisrahoitteisten sähköisten tutkimusaineistojen jatkokäytön avoimuuden lisäämiseen
- kuinka he suhtautuvat omissa tutkimushankkeissa kertyneiden sähköisten tutkimusaineistojen avoimuuteen.

Vastausten jakaumat on esitetty kuvioissa 4.5 ja 4.6.

**Kuvio 4.5. MITEN ARVIOI OMAN ALAN TUTKIJOIDEN (KESKIMÄÄRIN) SUHTAUTUVAN JULKISRAHOITTEISTEN SÄHKÖISTEN TUTKIMUSAINEISTOJEN JATKOKÄYTÖN AVOIMUUDEN LISÄÄMISEEN (%).**



**Kuvio 4.6. KUINKA ITSE SUHTAUTUU OMISSA TUTKIMUSHANKKEISSAAN KERTYNEIDEN SÄHKÖISTEN TUTKIMUS-AINEISTOJEN AVOIMUUTEEN (%).**



Sähköisen tutkimusdatan avoimuus / Valmisteluraportti OECD:n datasuosituksesta (2007)

Joka neljäs (25 %) kulttuurin ja yhteiskunnan tutkimuksen alan professorivastaaja katsoi, että selvä enemmistö oman alan tutkijoista kannattaa aineistojen avoimuuden lisäämistä (kuvio 4.5). Suurimman vastausryhmän muodostivat ne, joiden mielestä enemmistö oman alan tutkijoista kannattaa ajatusta. Tasan mielipiteiden arvioi jakautuvan noin 30 prosenttia professoreista. Vain joka kymmenes arvioi, että enemmistö vastustaisi ajatusta.

Kyselyyn vastanneiden tutkijoiden suhtautuminen omien aineistojen käytön avoimuuteen oli linjassa yleistä tilannetta koskevien mielipiteiden kanssa (kuvio 4.6). Joka kolmas (33 %) oli ehdottoman myönteinen omien sähköisten aineistojen avoimelle jatkokäytölle ja 43 prosenttia suhtautui asiaan melko myönteisesti. Neutraalin vastuksen valitsi 15 prosenttia ja kielteisen alle 10 prosenttia vastaajista.

Tieteenalaryhmittäisessä tarkastelussa ei löytynyt suuria eroja. Kasvatustieteilijöistä hieman suurempi osa suhtautui asiaan kielteisesti, mutta heidänkin joukossaan selvä enemmistö (n. 60 %) oli myönteinen omien aineistojen avoimuudelle.

Kolmas sähköisten aineistojen avoimeen käyttöön liittynyt yksittäiskysymys oli rakenteeltaan ja sisällöltäänkin kyselyn muista kysymyksistä poikkeava. Siinä vastaajille kuvattiin esimerkkitalanne, jonka pohjalta heidän tuli arvioida kantaansa tutkimusaineistojen avoimuutta koskeviin toimenpiteisiin. Kysymyksen johdanto kuului näin:

”Kuinka sitovia eri tahojen suositusten ja ohjeiden tulisi mielestänne olla tutkimusaineistojen Open Access -asiassa? Ajatellaan asiaa esimerkkitapauksen kautta.

Oletetaan, että tutkijaryhmä kokoaa julkisrahoitteisesti laajan sähköisen tutkimusaineiston, jota se ei enää käytä aktiivisesti viiden vuoden kuluttua aineistonkeruusta. Aineisto ei sisällä tietosuojongelmia eikä siihen kohdistu avointa käyttöä estäviä tekijänoikeudellisia tai kaupallisia intressejä.

Aineisto on pienin töin toimitettavissa tiedeyhteisön avoimeen käyttöön soveltuvan tietopalveluorganisaation kautta, joka huolehtii aineiston käyttöluupa-asioista ja informoi tutkijaryhmää aineiston myöhemmästä käytöstä.”

Tämän jälkeen vastaajia pyydettiin valitsemaan jokin kolmesta seuraavasta vaihtoehdosta:

”Pitääkö mielestänne esimerkkitapauksessa...

Vaihtoehto A:

Velvoittaa tutkimusryhmä luovuttamaan aineisto tiedeyhteisön avoimeen käyttöön.

Vaihtoehto B:

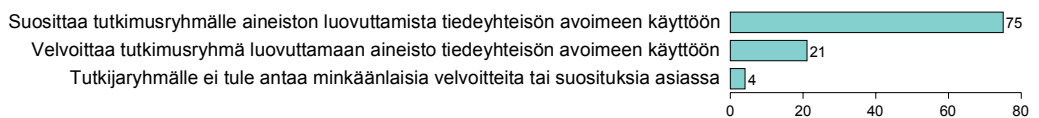
Suosittaa tutkimusryhmälle aineiston luovuttamista tiedeyhteisön avoimeen käyttöön.

Vaihtoehto C:

Tutkijaryhmälle ei tule antaa minkäänlaisia velvoitteita tai suosituksia asiassa.”

Vastaajien valitsemien vaihtoehtojen vastausjakaumat ovat kuviossa 4.7.

**Kuvio 4.7.** KUINKA SITOVIA ERI TAHOJEN SUOSITUSTEN JA OHJEIDEN TULISI OLLA: PITÄÄKÖ KUVATUSSA ESIMERKKITAPAUKSESSA... (kysymys kokonaisuudessaan, ks. teksti; %).



Sähköisen tutkimusdatan avoimuus / Valmisteluraportti OECD:n datasuosituksista (2007)

Kuvion 4.7 kuvaama tulos on myönteinen käytön avoimuudelle, vaikka kysymyksessä ei yksilöity tarkkaan sitä, mikä taho asiassa antaisi ohjeita tai suosituksia. *Kolme neljästä vastaajasta katsoi, että tutkimusryhmälle tulisi suosittaa aineiston luovuttamista tiedeyhteisön avoimeen käyttöön.* Sitovia velvoitteita kannatti joka viides ja vain neljä prosenttia oli sitä mieltä, ettei mitään velvoitteita tai suosituksia pitäisi antaa.

Samaa asiaa tiedusteltiin myös yliopistoille ja tutkimusrahoittajille suunnatuissa postikyselyissä, ja vastausten suunta oli täysin vastaava. Velvoitteita olisi näissä molemmissa vastaajaryhmissä ollut valmis ottamaan käyttöön vain yksi vastaaja. Täten ylivoimainen *enemmistö yliopistoista ja tutkimusrahoittajista olisi suosittanut kuvatussa tilanteessa tutkimusaineiston luovuttamista tiedeyhteisön avoimeen käyttöön.*

#### 4.4 Esimerkkejä avoimuuden lisäämisestä ja jatkokäytön edistämisestä

Data-arkistot tukevat aineistojen elinkaaren hallintaa ja kattavaa aineistojen tietopalvelua, mutta data-arkistojen perustaminen ja toiminnan vahvistaminen ei kuitenkaan ole ainoa mahdollinen vaihtoehto aineistojen avoimuuden järjestämisessä. Monet tutkimusorganisaatiot ja tut-

kimushankkeet ovat järjestäneet aineistojen käytön tai jatkokäytön tiedeyhteisölle itse. Tämä voi olla perusteltua varsinkin silloin, kun aineistot sisältävät arkaluonteisia tietoja ja ne vaativat anomymisointia, tai kun aineistojen jatkokäyttö edellyttää ensikäyttäjien konsultointia.

Joitakin tietoja olemassa olevien sähköisten tutkimusaineistojen nykyisistä käyttömahdollisuuksista löytyy valtion tutkimuslaitosten osalta liitteestä 3. Vaikka aineistojen avoimuus on vielä suhteellisen uusia asia, on olemassa hyviä esimerkkejä tiedeyhteisön avoimeen käyttöön perustuvista laajoista aineistoista. Kuvatut esimerkit koskevat suomalaisia sähköisiä tutkimusaineistoja ja sellaisia kansainvälisiä tutkimusdatoja, joissa on mukana tietoja suomalaisista.

#### 4.4.1 Kouluterveystutkimus<sup>7</sup>

Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskuksen vuosittaista kouluterveyskyselytutkimusta on tehty vuodesta 1995 alkaen. Kouluterveyskyselyn tuloksia on saatavilla valtakunnallisesti sekä lääneittäin ja maakunnittain. Tulokset on koottu alueittain omille sivuilleen. Kyselyn verkkosivut sisältävät tuoreimmat kuviot kyselyn indikaattoreista sekä raportit ja taulukot eri vuosilta (ks. <http://info.stakes.fi/kouluterveyskysely/FI/index.htm>).

Kouluterveystutkimuksen tuloksia hyödynnetään käytännön tasolla edistettäessä nuorten hyvinvointia. Välittömän yhteiskunnallisen hyödyn lisäksi kysely on esimerkki hyvin laajasta ja monipuolisista tutkimuksellisia käyttömahdollisuuksia tarjoavasta pitkäaikaisaineistosta. Aineistoja käytetäänkin sekä Stakesissa että eri yliopistoissa niin erillisissä tutkimuksissa kuin opinäytetöissä pro graduista väitöskirjoihin. Kyselyn toteuttava Stakes huolehtii aineiston säilyttämisestä käyttökuntoisena ja hallinnoi sen jatkokäyttöä.

Kouluterveystutkimuksen aineistojen käyttöluvapakemuksessa tulee ilmoittaa aineiston käyttötarkoitus, tarvittavat kyselyvuodet ja aihepiirit. Lupahakemukseen liitetään tutkimussuunnitelma ja aineiston käyttäjä sitoutuu lupahakemuksessa esitettyihin aineiston käytön ehtoihin.

#### 4.4.2 Kansallinen terveystutkimus<sup>8</sup>

Kansallisen terveystutkimuksen avulla on tutkittu suomalaisten terveyttä jo vuodesta 1972 alkaen, jolloin kartoitettiin Pohjois-Karjalan ja Pohjois-Savon maakuntien väestön terveydentilaa Pohjois-Karjala -projektia varten. Viisi vuotta myöhemmin tutkimuksella mitattiin projektin vaikutuksia näiden alueiden väestön terveyteen. 1982 alkaen tutkimuskohteita on vähitellen laajennettu ja vuonna 2007 tutkimus toteutetaan jo viidellä alueella: Pohjois-Karjalassa, Pohjois-Savossa, Oulun läänissä, Helsingissä ja Vantaalla, Turussa ja Loimaalla sekä yhdeksässä Varsinais-Suomen kunnassa.

7 Tiedot saatiin puhelimitse hankkeen vastaavana tutkijana toimivalta erikoistutkija Minna Pietikäiseltä (1.3.2007).

8 Tiedot saatiin puhelimitse erikoistutkija Markku Peltoselta KTL:stä (2.3.2007).

Kansallisen terveystutkimuksen toteuttaa Kansanterveyslaitos ja tutkimus tunnetaan tutkijayhteisössä nimellä FINRISKI. Tutkimuksen aineistojen jatkokäyttöä koordinoi FINRISKI:n johtoryhmä, jolle aineistojen tutkimisesta kiinnostuneet välittävät tiedot omista tutkimuksellisista kiinnostuksen kohteista. Johtoryhmä ensisijaisesti huolehtii siitä, että täysin samanlaisia tutkimuksia ei tehdä päällekkäisesti. Muutoin periaatteena on FINRISKI-aineistojen mahdollisimman laajan ja monipuolisen tutkimuskäytön edistäminen.

Aineistojen käyttö ei ole rajautunut vain Suomen yliopistoihin. Tieteellisesti poikkeuksellisen arvokkaita FINRISKI-aineistoja käytetään myös muissa maissa kansainvälisissä terveystutkimuksissa. Aineistojen pohjalta tehtäviä tutkimusjulkaisuja ilmestyy vuosittain noin 40–50.

Kansanterveyslaitoksen sähköisten tutkimusaineistojen hallinnan ja jatkokäytön edistämistä käsitellään julkaisussa *Hyvä tutkimustapa Kansanterveyslaitoksessa* (2005). Käsikirjassa ohjeistetaan aineistojen hallintaa, dokumentointia ja tallentamista. Sähköisten aineistojen käsittelyä organisoidaan osittain tietoverkon kautta. Käyttöoikeuksia ja tietoverkkoyhteyksiä järjestetään tarvittaessa ulkopuolisille tutkijoille, jotka pystyvät näin analysoimaan Kansanterveyslaitoksen atk-laitteille sijoitettuja tiedostoja. Tunnisteettomia aineistoja voidaan myös tilata ja lähettää sähköpostitse. (Mt. 48–53.)

Lisätietoja:

<http://www.ktl.fi/portal/suomi/yhteistyoprojektit/terveystutkimus/>

#### 4.4.3 European Social Survey

European Social Survey (ESS) on Euroopan tiedesäätiön (ESF) aloitteesta syntynyt kansainvälinen tutkimusohjelma, jonka tavoitteena on kyselytutkimusten avulla kartoittaa yhteiskunnalliseen kehitykseen liittyvää arvojen, asenteiden ja käyttäytymisen muutosta. Hankkeessa kerätään joka toinen vuosi vertailukelpoista tutkimusaineistoa yli 20 Euroopan maasta. Tutkimusaiheet vaihtelevat poliittisesta osallistumisesta tiedotusvälineiden seuraamiseen sekä poliittisiin ja sosiaalisiin asenteisiin. ESS-aineistonkeruu toteutettiin ensimmäisen kerran vuosina 2002–2003 ja toisen kerran vuosina 2004–2005. Suomi on ollut mukana alusta lähtien.

ESS:n menetelmälliset tavoitteet ovat poikkeuksellisen korkeat. Vertailukelpoisuuden toteutumista ohjaa keskitetysti kansainvälinen koordinoitiryhmä. Eri maiden aineistojen vertailukelpoisuus varmistetaan yhteisillä säännöillä, joista tärkeimmät koskevat otantaa, haastattelulomakkeen kääntämistä, kenttätöitä sekä lopullisen aineiston muokkaamista arkistoinnin ja jatkokäytön edellyttämään muotoon. Kaikissa osallistujamaissa käytetään ainoastaan yksinkertaisia satunnaisotoksia ja kenttätö perustuu käyntihaastatteluihin.

European Social Survey -hankkeen ensimmäisen ja toisen kierroksen aineistot sekä yhdistelmäaineisto ovat saatavana vapaasti tieteelliseen tutkimuskäyttöön englanninkielisinä Norjan tietoarkiston verkkosivujen kautta. Aineistojen kerääjillä ei ole minkäänlaista ensikäyttöoikeutta kansainväliseen vertailuaineistoon, vaan yhdistetty aineisto tulee täysin avoimeen käyttöön heti valmistumisensa jälkeen. Aineistojen käyttö ja lataaminen omalle koneelle on maksutonta, mutta edellyttää rekisteröitymistä (ks. <http://www.europeansocialsurvey.org>).

#### 4.4.4 Luxembourg Income Study

Luxemburgin yliopiston yhteydessä toimiva kansainvälinen Luxembourg Income Study (LIS) -projekti kerää tietoja yhteensä 30 maasta. LIS tarjoaa tutkijoiden käyttöön kaksi kansainvälistä vertailuaineistoa: tulonjakotutkimuksen (LIS eli Luxembourg Income Study) ja työllisyystutkimuksen (LES eli Luxembourg Employment Study). Molemmat aineistot ovat nk. mikrodataa.

LIS-datan perustutkimusyksikkönä on kotitalous. Useista maista on jo saatavilla useampia poikkileikkaustietoja noin viiden vuoden välein vuodesta 1980 alkaen. Aineistot tarjoavat runsaasti mahdollisuuksia vertailla tulonjakoa ja tulonsiirtojen vaikuttavuutta eri maissa, vaikka kaikki muuttujat eivät ole suoraan vertailukelpoisia. Muuttujat sisältävät tietoja ammatista, siviilisäädystä, perhekoosta, työnteosta, kotitalouden tuloista (mm. työtulot ja tulonsiirrot) ja kulutusmenoista (mm. asumis-, koulutus-, päivähoito- ja sairaskulut).

LES-datan muuttujat sisältävät tietoja henkilön työmarkkina-asemasta, päätyön ja mahdollisen toisen työn luonteesta, aikaisemmasta työhistoriasta, työnhausta, koulutuksesta ja tuloista. LES perustuu Labour Force Survey -tutkimuksiin eri maissa. Tarjolla on kaksi aaltoa: 1990 ja 1995. Vuodesta 2000 alkaen LES-dataa ei enää saa erikseen. Käytettävissä on laajennettu LIS-data, johon on lisätty työhön ja työmarkkinoihin liittyviä muuttujia.

Aineistoja käytetään lähettämällä tietyn tilastomatematiikan ohjelmiston komentotiedosto (SPSS-, SAS- tai STATA-syntaksi) LIS:n antamaan sähköpostiosoitteeseen. Vastauksen ja ajon tulokset saa omaan sähköpostiosoitteeseensa lähes välittömästi. Kaikki LIS-aineistoihin perustuvat tutkimukset on julkaistu LIS:n Working Papers -sarjassa (ks. <http://www.lis-project.org/>).

## 5 Tutkimusaineistojen avoimuuden ja datapolitiikan vaihtoehtoja

Tässä luvussa esitellään koti- ja ulkomaisia toimintamalleja, joilla on pyritty tukemaan julkisrahoitteisen tutkimusdatan avointa saatavuutta. Tarkastelu palaa edellä esillä olleisiin perusvaihtoehtoihin jatkokäytön varmistamisessa, ja arvioi samalla eri mallien hyviä ja huonoja puolia. Näin argumentit sekä laajenevat että syventyvät selvityksen toisessa luvussa esitettyyn elinkaariajatteluun nähden.

Tämän jälkeen kuvataan olemassa olevia, varsinkin ulkomaisia tiedepoliittisia ratkaisuja tutkimusdatan avoimuuden turvaamiseksi. Näin pohjustetaan keskustelua OECD-suosituksen toimeenpanon tarkoituksenmukaisista etenemistavoista Suomessa.

### 5.1 Aineiston jatkokäyttö tutkijan tai tutkimusyksikön hallinnoimana

Tutkimusaineistot voidaan antaa jatkokäyttöön muille tutkijoille niin, että aineiston kerännyt tutkija tai tutkimusyksikkö pitää aineiston itsellään. Hyvä puoli ratkaisussa on se, että tutkija voi informoida aineiston jatkokäyttäjiä ja tarvittaessa kontrolloida myöhempää käyttöä. Samalla alkuperäiset tutkijat pysyvät tietoisina siitä, minkä tyyppisissä jatkotutkimuksissa aineistoa käytetään.

Jotta tutkimusaineistot olisivat aidosti tiedeyhteisön käytettävissä, aineiston perustiedot tulisi olla yleisesti saatavilla. Tutkimusyksiköissä säilytettävistä aineistoista nämä perustavat tiedot voidaan tarjota vaikkapa yksiköiden tai tutkimushankkeiden omilla www-sivuilla.

Jatkokäyttäjien näkökulmasta systemaattinen ja hallittu aineistojen etsiminen onnistuu kuitenkin parhaiten keskitetympien tietokantojen avulla. Kirjastojen ja tietoarkistojen avoimesti käytettävät luettelotietokannat ovat yksi esimerkki toimivasta informaation välittämisestä. Esimerkiksi Yhteiskuntatieteellinen tietoarkisto kuvailee tietokantoihinsa myös yksittäisten tutkijoiden tai tutkimusyksiköiden hallussa olevia tutkimusaineistoja.

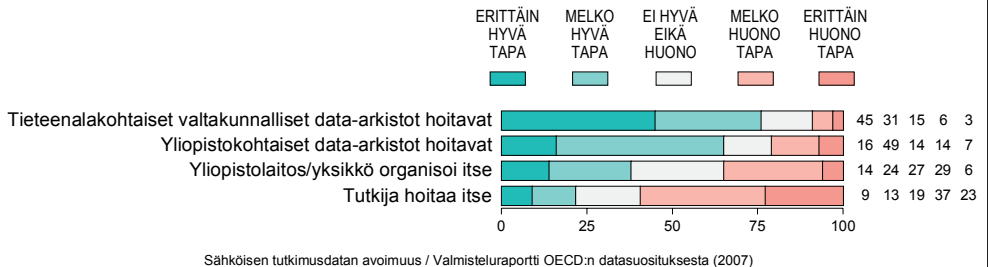
Tutkijan tai tutkimusyksikön itse hallinnoimassa aineistojen jatkokäytössä on ongelmana usein myös se, että tutkijoilla ei ole riittävästi aikaa neuvoa aineiston jatkokäyttäjiä eikä hoitaa aineiston kokoamiseen ja siirtoon liittyviä käytännön asioita. Tutkijat eivät myöskään välttämättä osaa tai ehdi konvertoida aineistojansa pitkäaikaisiin tallennemuotoihin, saati huolehtia aineiston pitkäaikais säilytyksestä. Sama pätee yllättävän usein myös tutkimusorganisaatioihin, vaikka asiasta olisikin olemassa yksikön sisällä sovittuja toimintatapoja.

Myös KY-alan professorivastaajat pitivät vuoden 2006 lopussa huonoimmin avoimuutta edistävänä ratkaisuna vaihtoehtoa, jossa tutkija itse hoitaa omien aineistojensa arkistoinnin ja jatkokäyttöön luovuttamisen (kuvio 5.1). Peräti 60 prosenttia vastaajista piti sitä melko tai erittäin huonona tapana toteuttaa aineistojen avoimuutta käytännössä.

Vastaajien mielipiteet jakautuivat, kun he arvioivat yliopistolaitosta tai tutkimusyksikköä tutkimusaineistojen säilyttämisen ja jatkokäyttöön luovuttamisen organisoijana. 38 prosenttia piti sitä erittäin tai melko hyvänä tapana. Vastaavasti 35 prosenttia piti sitä huonona vaihtoehtona. Laitostasolla aineistojen kuvailujen ja avoimesti saatavien aineistoluetteloiden organi-



**Kuvio 5.1. MITEN HYVINÄ TAI HUONONA PITÄÄ ERI TAPOJA TOTEUTTAA SÄHKÖISTEN TUTKIMUSAINEISTOJEN ARKISTOIMINEN JA SIIHEN LIITTYVÄT TOIMINNOT (kysymys kokonaisuudessaan, ks. teksti; %).**



sointi voidaan kuitenkin tehdä hiukan keskitetympin kuin tilanteessa, jossa jokainen tutkija yksin vastaa aineistoistaan. Mutta tavallisesti tutkimusyksiköissä ei ole henkilöresursseja aineistojen pitkäaikaissäilytyksen tietotekniikan kehittämiseen ja ylläpitoon, eikä myöskään jatkokäytön toteuttamiseen ja hallinnointiin. Sen organisointi laitostasolla edellyttäisi joka tapauksessa yhteneviä käytäntöjä ja ohjeita sekä aineistojen hallintaan suunnattuja resursseja.

## 5.2 Data-arkistot

Kyselyyn vastanneet pitivät data-arkistoja parhaimpana tapana toteuttaa käytännössä tutkimusaineistojen arkistointi ja jatkokäyttöön luovutus (kuvio 5.1). Kaksi kolmesta vastaajasta piti erittäin tai melko hyvänä tapana tilannetta, jossa yliopistokohtaiset data-arkistot hoitavat aineistojen arkistoinnin ja jatkokäyttöön luovutuksen.

Parhaimmaksi tavaksi kuitenkin arvioitiin valtakunnalliset tieteenalakohtaiset data-arkistot. Melkein joka toinen vastaaja (45 %) piti valtakunnallisia tieteenalakohtaisia tietoarkistoja erittäin hyvänä tapana ja lisäksi lähes kolmannes (31 %) melko hyvänä tapana toteuttaa käytännössä tutkimusaineistojen avoimuuteen liittyvät käytännön työt. Myös tieteenalakohtaisten data-arkistojen perustaminen edellyttäisi valmistelemaa työtä, yhteisiä pelisääntöjä, käytäntöjä, ohjeita ja resursseja. Tätä puolta kommentoi yksi kyselyn vastaajista:

”Tavoite on kovin epärealistinen ennen kuin asiaa koskevat yleiset periaatteet on pystytty TIETEENALAKOHTAISESTI määrittelemään suosituksen kartoittamissa, ongelmia tuotavissa kysymyksissä. Samalla on ratkaistava realistiset keinot tavoitteeseen pääsemiseen. Niistä ylivoimaisesti paras olisi se, että julkinen valta rahoittaisi tutkijoille sellaisten tietoteknisten välineiden käyttöön saannin, jotka ajavat samalla sekä yksityisen (tutkijan) että julkisen datan käyttöön saannin tavoitteita niin tehokkaalla tavalla, että tutkijat siihen tarttuvat jo senkin takia, että ne auttavat heitä merkittävästi heidän itselleen päätavoitteekseen asettamassaan omassa tutkimuksessaan. Kaikki se työ, joka EI yksittäisen tutkijan oman tutkimuksen kannalta ole erittäin hyödyllistä, mutta on edellytys datan julkiseen käyttöön

saattamiseen, tulisi hoitaa sitä varten perustettujen tahojen (esim. Yhteiskuntatieteellinen tietoarkisto) riittävällä tuella.”

Tieteenalakohtaisten data-arkistojen etuna on tietoteknisen ja sisällöllisen osaamisen, työn ja palvelujen keskittäminen. Se vähentää aineistojen avoimuudesta koituvaa lisätyöpainetta tutkija- ja laitostasolla, jossa aikaa ei koeta olevan riittävästi edes varsinaiseen tutkimustyöhön.

Tutkimusaineistojen avoimuutta ja hallittua jatkokäyttöä onkin Suomessa jossain määrin jo organisoitu perustamalla erilaisia data-arkistoja tai muita sähköisten aineistojen akateemisia tietovarantoja. Erityisesti paperiaineistojen arkistointiin ovat humanistisilla aloilla keskittyneet perinnearkistot (esim. Joensuun perinnearkisto, Suomalaisen Kirjallisuuden Seuran kansanrunousarkisto, Kirkollisen kansanperinteen arkisto). Tampereen yliopiston musiikintutkimuksen laitoksen arkisto ja Turun yliopiston kulttuurien tutkimuksen laitoksen TKU- ja TYKL-arkistot sekä Suomalaisen Kirjallisuuden Seura ovat tehneet alallaan pioneerityötä myös aineistojensa digitoinnin osalta (ks. Mahlamäki 2002a, 2002b, 2002c ja Klemettinen 2006). Mainitut arkistot keräävät nykyisin aineistoja myös valmiiksi digitaalisessa muodossa.

Kielitieteitäkin palvelevat Suomessa mittavat sähköiset tutkimusaineistot. Valtakunnallisessa Kotimaisten kielten tutkimuskeskuksessa on yli 76 miljoonaa sanaa erilaisia kielenaineksiä tietokoneessa käsiteltävässä, sähköisessä muodossa. Tutkimuskeskuksessa käytössä olevat aineistot ovat syntyneet hyvin eri tavoin ja erilaisia tarpeita silmällä pitäen. Korpuksia varten aineistoja on yhdenmukaistettu, jotta kokoelmat olisivat tietoteknisesti helppokäyttöisiä. Aineistot jakautuvat nykysuomen aineistoihin (1900 ->), 1800-luvun kirjasuomen aineistoihin (1810–1900), vanhan kirjasuomen aineistoihin (-> 1810) sekä murre- ja nimiaineistoihin. Myös suomenruotsin aineistoja on koottu tekstipankiksi.

Sähköisessä muodossa olevia kielentutkimuksen aineistoja on arkistoituna myös Tieteen tietotekniikan keskuksen (CSC). Aineistoista iso osa on ns. rakenteisia, mikä mahdollistaa erityyppiset haut. Suuri osa korpuksista on tekstiaineistoja, mutta kielentutkimuksen tarpeisiin on luotu myös digitoitu audiovisuaalinen suomalainen radio- ja tv-korpus.

CSC:n extranet-palvelun avulla tutkijoilla on mahdollisuus etäkäyttää CSC:n tietokantoja tavallisen verkkoselaimen avulla. Käyttöliittymäohjelmisto on kehitetty sellaiseksi, että sen avulla metakoneympäristön palvelut ovat käytettävissä niillekin, jotka eivät ole tottuneet käyttämään metakoneen unix-käyttöjärjestelmää. CSC:n palvelu on esimerkki elektronisten tutkimusaineistojen tiedeyhteisölle suunnatusta avoimesta käyttömahdollisuudesta.

Sähköisten aineistojen avoimuutta tukee myös Rekisteritutkimuksen tukikeskus (ReTki), jota rahoittavat Kansaneläkelaitos, Kansanterveyslaitos, Stakes ja Työterveyslaitos. ReTkin tavoitteena on edistää viranomaisten tilastollisiin ja hallinnollisiin tarkoituksiin keräämien kansallisten rekisterien tutkimuskäyttöä erityisesti terveys- ja sosiaalitieteissä. Sen perustehtäviin kuuluu tiedon ja koulutuksen tarjoaminen rekistereistä ja niiden tutkimuskäytöstä. (Ks. lisätietoja <http://retki.stakes.fi/FI/index.htm>)

Yhteiskuntatieteelliset tietoarkistot ovat myös alan vakiintuneita toimijoita monissa maissa. Yhdysvalloissa ja useissa Euroopan maissa tietoarkistoja perustettiin jo 1960-luvulta alkaen. Suomen Yhteiskuntatieteellinen tietoarkisto aloitti toimintansa vuonna 1999. Se toimii yhteiskuntatieteellisen tutkimuksen ja opetuksen valtakunnallisena palveluyksikkönä, joka arkistoi ja välittää elektronisia tutkimusaineistoja tutkimus- ja opetuskäyttöön. Tietoarkisto on opetusministeriön rahoittama yksikkö, joka on sijoitettu erillislaitoksena Tampereen yliopis-

ton yhteyteen. Se vastaanottaa ensisijaisesti aineistoja, jotka soveltuvat yhteiskuntatieteelliseen tutkimukseen. Soveltuvien osien tietoarkistoon arkistoidaan myös humanististen alojen, kasvatustieteiden ja terveystieteiden aineistoja. Aineistojen tulee olla teknisesti ja juridisesti arkistointikelpoisia. Arkistoitavat aineistot tarkistetaan ja tarvittaessa anonymisoidaan. Kaikki aineistot kuvaillaan sähköisiin aineistoluetteluihin, joihin voi tutustua avoimen verkon kautta.

### 5.2.1 Käyttöluvut data-arkistoissa

Sähköisten tutkimusaineistojen Open Access rajautuu useimmiten käytännössä vain tiedeyhteisön harjoittamaan tutkimukseen ja opetukseen. Tutkimusaineistoja arkistoiivat perinnearkistot, CSC ja Yhteiskuntatieteellinen tietoarkisto ovat rajanneet arkistoitujen tutkimusaineistojensa käytön tieteelliseen tutkimukseen ja opetukseen. Siksi aineistojen käyttöä kontrolloidaan käyttöluvhakemuksin.

Professorikyselyn vastaajista joka toinen (48 %) piti avoimuuden haittana lisääntyviä tietosuojariskejä ja tutkimuseettisiä riskejä (kuvio 3.4). Rajattu käyttöoikeus tutkimukseen tai opetukseen, lupakäytännöt ja jatkokäyttäjää juridisesti sitovat käyttöehdot ovat data-arkistojen tapa ratkaista jatkokäytön haittoiksi koetut riskit. Käytännöt nojautuvat myös sosiaaliseen luottamukseen: käyttäjät sitoutuvat käyttöehtoihin ja toimivat muutoinkin hyvien tieteellisten käytäntöjen mukaisesti.

Käyttöluvhakemuksessa kerrotaan aineiston käyttötarkoitus eli suunniteltu tutkimus, hakijan ja muiden aineiston käyttäjien yhteystiedot sekä tutkimusorganisaatio, jossa/jonne tutkimus tehdään. Opiskelijoiden pyytäessä aineistoja seminaaritöihin, kandidaatintutkielmiin tai pro graduihin, edellytetään tavallisesti opiskelijan allekirjoituksen lisäksi myös työn ohjaajan allekirjoitusta käyttöluvhakemukseen. Näin halutaan varmistaa aineistojen legitiimi käyttö.

Kun aineistonsa arkistoitavaksi antanut tutkija ei ole asettanut omia erityisiä käyttörajoituksia, data-arkistot luovuttavat aineistoa tieteellisen tutkimuksen ja opetuksen tarpeisiin lupakäytäntöjensä mukaisesti.

Käyttöluva tarvitaan tavallisesti myös verkon kautta käytettäviin aineistoihin. CSC:n kielikorpusten online-tutkimuskäyttöä varten täytyy tunnukset hakea erillisellä käyttöluvhakemuksella (ks. <http://www.csc.fi/asiakkaaksi/korkeakoulut/kayttolupahakemukset/kielipankki>). Myös Turun yliopiston kulttuurien tutkimuksen laitoksen arkiston digitoitujen teksti-, ääni-, kuva- ja videoaineistojen online-tutkimuskäyttö edellyttää erillistä käyttölupaa tunnusten saamiseksi (ks. <http://kultut-arkistot.utu.fi/index.html>).

Mikäli tutkimusaineistoon ei sisälly mitään eettisiä tai tietosuoaongelmia, sitä voidaan käyttää myös täysin avoimesti. Tällaista materiaalia löytyy Internetistä erittäin runsaasti, ja myös tutkimusaineistojen säilytykseen ja jatkokäyttöön erikoistuneet yksiköt tarjoavat niitä jossakin määrin. Esimerkiksi Yhteiskuntatieteellisen tietoarkiston menetelmäopetuksen tietovarannon harjoitusaineistot ovat avoimesti käytettävissä (ks. <http://www.fsd.uta.fi/metelmaopetus/>). Tampereen yliopiston musiikintutkimuksen laitoksen Kansanperinteen arkistoaineistoista vapaasti verkon kautta käytettävissä on näyteaineistoja, mutta varsinaiseen tietokantaan saa tunnukset vain erillistä käyttöluvhakemuksesta vastaan (ks. <http://www.uta.fi/laitokset/mustut/arkistosivut/aanet.htm>).

## 5.2.2 Käyttöehdot data-arkistoissa

Käyttölupahakemukset rajaavat tutkimusaineistojen käytön tutkimukseen ja opetukseen, mutta erilliset käyttöehtositoumukset tarkentavat käytön ehtoja. Yleensä jatkokäyttäjä ei saa käyttää tutkimusaineistoja kaupallisiin tarkoituksiin. Ehdottoman kiellettyä on aineistoissa ilmenevien tietojen käyttö yksittäisiä ihmisiä koskevaan päätöksentekoon, valvontaan tai tarkkailuun.

Jatkokäyttäjän allekirjoittamassa käyttöehtositoumuksessa vahvistetaan ehdot, joihin aineiston käyttäjä sitoutuu. Aineistoa saa käyttää ainoastaan käyttölupahakemuksessa ilmoitettuun käyttötarkoitukseen, jonka jälkeen aineisto on hävitettävä. Lisäksi aineiston käyttäjän on huolehdittava aineiston säilyttämisessä ja käytössä riittävästä tietoturvasuhteesta ja vastattava siitä, että aineisto ei ole muiden kuin käyttölupahakemuksessa yksilöityjen henkilöiden tai tahojen saatavilla. Käyttäjän on huolehdittava aineiston käsittelyssä ja julkistettavissa tuloksissa niiden henkilöiden ja tahojen tietosuojasta, joita koskevia tietoja aineisto sisältää.

Sopimuksessa aineiston käyttäjä sitoutuu myös siihen, että hän ei pyri tunnistamaan tietosuojaan rikkovalla tavalla aineiston tietoihin liittyviä henkilöitä tai tahoja. Kun aineisto on arkaluonteinen ja sisältää huomattavasti epäsuoria tunnisteita, käyttöehtositoumuksen lisäksi voidaan edellyttää allekirjoitettua vaitiolositoumusta. Tutkimusetiikan kannalta hyvin oleellinen käyttöehto on aineiston alkuperän mainitseminen julkaisuissa. Käytännössä se tarkoittaa aineiston alkuperäisten tutkijoiden ja aineiston nimen mainitsemista tieteen konventioiden mukaisin viittauksin.

Aineistonsa arkistoitavaksi luovuttavat tutkijat määrittävät luovutus sopimuksessa aineistonsa jatkokäyttöehdot. Kun tutkija luovuttaa tutkimusaineiston Yhteiskuntatieteelliseen tietokantoon arkistoitavaksi, hän voi halutessaan asettaa aineiston jatkokäytön alkamisajaksi päivämäärän, jolloin tietää omien ensitulosten jo tulleen julkaistuksi. Toiseksi tutkija voi rajata luovuttamansa aineiston jatkokäytön vain tutkimukseen rajaten pois aineiston tieteellisen opetuskäytön mahdollisuuden. Jos tutkija niin haluaa, häneltä voidaan myös kysyä erikseen lupa jokaiseen aineiston jatkokäyttöön. Tällöin täytyy kuitenkin sopia alustavasti menettelytavoista niissä tilanteissa, kun tutkija ei kenties itse syystä tai toisesta pysty lupapäätöstä tekemään. Kun erityisiä käyttörajoituksia ei aseteta, arkisto luovuttaa aineistoa tieteellisen tutkimuksen ja opetuksen tarpeisiin arkiston omien lupakäytäntöjensä mukaisesti.

Data-arkistojen käyttöehtositoumusten mukaisesti aineiston jatkokäyttäjän on tuhottava aineisto analyysien valmistumisen jälkeen. Mikäli myöhemmin tulee tarvetta analyysien tarkistamiseen, data-arkistot toimittavat aineiston pyytäjälle uudelleen. Käytännön mukaisesti tutkijat eivät kerää koneelleen suoria tai epäsuoria tunnisteita sisältäviä datatiedostoja pitkäaikaiseen säilytykseen, koska data-arkistot hoitavat niiden säilytyksen ja hallinnan. Näin tuetaan tutkimusaineistojen hallinnan, tietosuojan ja tietoturvan hyviä käytäntöjä.

## 5.2.3 Tietosuoja data-arkistoissa

Tietosuoja tarkoittaa ihmisten yksityisyyden kunnioittamista ja suojelemista oikeudellisia säännöksiä noudattavin periaattein ja toimintakäytännöin. Jokaisen tutkijan velvollisuus on noudattaa tietosuojalainsäädäntöä riippumatta siitä käyttääkö hän data-arkiston aineistoja vai itse

keräämiensä aineistoja. Lain huolellisuusvelvoite edellyttää, ettei aineistoissa esiintyvien tutkittavien yksityisyyden suojaa loukata. Suojaamisvelvoite puolestaan edellyttää, että tutkimusaineistojen sisältämät henkilötiedot suojataan niin, etteivät asiattomat pääse niihin käsiksi.

Tietosuojaan liittyy oleellisesti tietoturva. Henkilötietolain sisältämän määritelmän mukaisesti tietoturva tarkoittaa teknisiä ja organisatorisia toimenpiteitä, joita tehdään henkilötietojen suojaamiseksi asiattomalta pääsylvä tietoihin ja vahingossa tai laittomasti tapahtuvalta tietojen hävittämiseltä, muuttamiselta, luovuttamiselta, siirtämiseltä taikka muulta laittomalta käsittelyltä (henkilötietolaki 32 §). Tietoturvaan kuuluvat ratkaisut siitä, millä tietokoneilla tutkimusaineistoja säilytetään, ovatko ne salasanoilla suojattuja, saako ja millä edellytyksillä kopioita aineistosta viedä tutkimustilojen ulkopuolelle ja miten muutoin huolehditaan tietoteknisin ja teknisin ratkaisuin luottamuksellisten tietojen suojaamisesta.

Tietokonelaitteistojen ja tallennusmenetelmien vanheneminen johtaa aika ajoin tarpeeseen muuntaa aineistovaranto toiseen tietojärjestelmään sopivaksi. Tähän data-arkistoissa vaurudutaan seuraamalla tiiviisti alan kehitystä. Myös arkistoitujen aineistojen tallenneformaattit valitaan siten, että ne ovat mahdollisimman helposti muutettavissa toisiksi formaateiksi. Käytännössä tämä tarkoittaa avoimien, standardien mukaisten ja useiden valmistajien tukemien formaattien suosimista suljettujen ja valmistajariippuvaisten formaattien sijaan.

Digitaalisista tutkimusaineistoista tehdään säännöllisesti varmuuskopioita, joista 1–2 säilytetään muualla kuin data-arkiston yhteydessä. Esimerkiksi Yhteiskuntatieteellisessä tietoarkistossa tehdään arkistoaineistoista varmuuskopiot erilliselle levyjärjestelmälle joka yö. Levyjärjestelmästä otetaan erilliset varmuuskopiot kaksi kertaa kuukaudessa ja niistä yhdet kappaleet arkistoidaan tietoarkiston ulkopuolelle. Digitaaliset käyttöaineistot toimitetaan tutkijoille tavallisimmin sähköpostin liitetiedostona tai postitse CD-ROM-levyllä. Jos aineisto on suuri ja/tai se sisältää tunnisteita, se voidaan myös toimittaa SFTP-siirtona. Siirron ajankohdasta sovitaan tilaajan kanssa ja välittömästi siirron onnistumisen varmistuttua siirtoon käytetyt tunnukset poistetaan käytöstä.

Yliopistojen ja tutkimuslaitosten tehtävä on taata tietosuojan infrastruktuuri, mutta viime kädessä sen käytäntöön soveltamisesta vastaavat tutkijat itse. Data-arkistojen aineistojen käytön lopullisesta tietosuojasta vastaaminen kuuluu jokaiselle aineiston käyttäjälle. Asianmukaiset käytännöt perustuvat luottamukseen, joka pysyy tutkijoiden noudattaessa data-arkistojen toimittamien aineistojen käyttöehtoja.

### 5.3 Datakeskukset (data enclaves)

Normaalioloissa digitaalisten tutkimusaineistojen jakelu ja asianmukainen käyttö perustuvat kirjallisiin sopimuksiin ja sosiaaliseen luottamukseen. Datakeskuksissa (engl. data enclave tai data center) käyttörajoitukset tehdään konkreettisiksi. Aineiston käyttö rajataan tilallisesti ja tietoteknisesti. Tutkijan tulee mennä itse datakeskuksen tiloihin, jotka ovat lukittuja ja tietoturvallisia paikkoja. Aineistoja ei saa siirtää datakeskusten ulkopuolelle, joten analyysit tehdään paikan päällä. Datakeskusten tietokoneilta ei ole myöskään verkkoyhteyksiä, jotta aineistojen siirto tahallisesti tai edes vahingossa ei olisi mahdollista.

Datakeskuksissa käytettävät aineistot ovat yleensä laajoja ja tunnisteellisia. Huomattavan paljon suoria tunnisteita sisältävien aineistojen ohella datakeskuksiin voidaan sijoittaa aineis-

tot, joissa on suuret määrät yksittäisiä ihmisiä koskevia epäsuoria tunnisteita. Erityisesti Yhdysvalloissa datakeskusten käyttöä suosittellaan aineistoille, joiden tietoihin pystyisi yhdistämään muita vapaasti saatavilla olevia tietoaaineistoja. Riskinä pidetään mahdollisuutta tunnistaa tai alueellisesti paikantaa ihmisiä tai ihmisryhmiä yhdistettyjen tietojen avulla. Tunnistamisen suurimpana ongelmana on koettu sen mahdollistama aggressiivinen suoramarkkinointi (Heeney 2004, Weigand ja Heeney 2003).

Suomessa datakeskuksena toimii esimerkiksi Tilastokeskuksen Tutkimuslaboratorio. Se on valvottu ympäristö, jossa aineistoja voidaan analysoida Tilastokeskuksen tietoturvakäytäntöjä ja sopimusehtoja noudattaen. Tutkimuslaboratoriossa voi analysoida yrityksiä, toimipaikkoja ja niiden henkilöstöä koskevia tietoja. Joitakin Tutkimuslaboratorion aineistoja voi myös yhdistellä. Tutkimuslaboratoriosta voi vuokrata tutkijapaikan, jolla saa pääsyn erikseen määriteltyihin suojattuihin aineistoihin, opastusta aineistojen mahdollisuuksista ja rajoituksista sekä tarvittavan tietotekniikan ja käyttäjätuen.

Datakeskusten etuna on mahdollisuus päästä analysoimaan muutoin salassa pidettäviä, arkaluonteisia ja/tai tunnistellisia aineistoja. Muita etuja voivat olla asianmukaiset analyysiohjelmat, tietotekninen tuki ja aineistoja koskeva neuvonta paikan päällä. Datakeskusten huomattava etu on myös se, että aineistojen ja rekisteritietojen yhdistäminen mahdollistaa monipuolisten tutkimuskysymysten asettamisen.

Nolte ja Keller (2001) pitävät datakeskusten keskeisimpänä ongelmana perustamisen ja ylläpidon suuria kustannuksia. Usein keskuksset ovat myös käyttäjilleen kalliita. Esimerkiksi Tilastokeskuksen Tutkimuslaboratorio on ensisijaisesti tarkoitettu laajojen akateemisten tutkimushankkeiden toteuttamiseen, sillä sen tiloissa analyysia tekevän tutkimushankkeen perustamiskustannukset ovat 883 euroa ja yhden työpisteen vuokraus 18 euroa tunti (ks. lisätietoja Tilastokeskuksen kotisivuilta <http://www.stat.fi/index.html>). Tutkimustyön kustannuksia lisää myös se, että analyysia tehdäkseen tutkijan on päästävä konkreettisesti paikan päälle datakeskukseen. Tässä mielessä datakeskukset muistuttavat perinteisiä arkistoja, joissa säilytettävää paperimateriaalia tulee mennä tutkimaan paikan päälle itse arkistoon.

## 5.4 Kohti periaatteellista avoimuutta: tieteellisten julkaisujen ja rahoittajien datapolitiikat

Yhteiskunta- ja käyttäytymistieteissä sähköisten tutkimusaineistojen avoimuudesta alettiin laajemmin keskustella 1990-luvulta alkaen. Keskusteluissa esitettiin hyvin paljon varauksia avoimuutta kohtaan. Esimerkiksi Gibson (1995) epäili politiikan tutkimuksen heikentyvän laadullisesti, mikäli aineistot tulevat avoimiksi. Hänen mukaansa tutkimusaineistojen avoimuus saa aikaan sen, että laajat aineistot ovat kyllä aktiivisessa käytössä, mutta uusia ja tärkeitä aineistoja ei enää kerätä. Lisäksi hän arveli metodologisen kehityksen hidastuvan, kun käytetään vain samoja aineistoja (mt. 475.) Samansuuntainen kommentti saatiin myös yhdeltä KY-alan professorikyselyn vastaajalta: ”Ei-luovat tutkijat menettävät lopullisesti kyvyn tuottaa itse tutkimusmateriaalia”.

Niin sanottujen ”vanhojen” aineistojen käytöstä saadun kokemuksen valossa avoimuus on alettu nähdä päinvastoin edellytyksenä kehittää tutkimusmenetelmiä. Keckman-Koivuniemi ja Kleemola (2006) kertovat, miten aineistoja käytetään mittareiden ja analyysimenetelmien

testaamiseen. Olemassa olevia aineistoja käytetään myös apuna kerätessä uusia aineistoja. Niistä tutkitaan, miten erilaiset operationalisoinnit ovat käytännössä toimineet. Aineistot ovat myös erittäin käyttökelpoisia sekä kvantitatiivisten että kvalitatiivisten tutkimusmenetelmien opetuksessa. (Mt. 1–13.)

Luonnontieteissä aineistojen avoimuus on kokemuksen kautta nähty keskeisenä tekijänä tieteellisen tutkimuksen edistymiselle. Esimerkiksi molekyylibiologiassa aineistojen avoimuutta vastustettiin aluksi, mutta aineistojen yhteiskäytöstä saadut tulokset ovat johtaneet innostuksen kautta laajaan hyväksyntään. Aineistojen avoimuus havaittiin ratkaisevan merkitykselliseksi molekyylibiologian kehitykselle. (Gazzaniga ja Rockinore 2001.)

Tutkijat toivoivat Yhdysvalloissa 1990-luvun alussa sekä yhtenäisiä käytäntöjä että ratkaisuja moniin konkreettisiin asioihin ennen kuin aineistojen avoimuudesta tehtäisiin velvoittava käytäntö. Tuolloin pidettiin ratkaisemattomina kysymyksiä (Sieber 1991, 142–145):

- ketkä lopulta omistavat aineistot
- kenen vastuulla on dokumentoida aineistot
- saako aineiston avoimeksi saattamisesta tieteellistä meriittiä
- viitataan aineiston tekijöihin jatkotutkimuksissa
- missä määrin rahoittaja voi määrätä aineistosta ja sen jatkokäytöstä
- miten aineistoa koskevat vastuut määritetään yhteisprojekteissa.

Täsmälleen samat kysymykset odottavat periaatteellisia ja käytännön ratkaisuja nyt Suomessa, sillä juuri nämä kysymykset ovat keskeisiä OECD:n datasuosituksen kansallisen toimeenpanon kannalta.

Tutkimusaineistojen avoimuutta on käytännössä edistetty sekä tieteellisten julkaisujen että rahoittajien linjauksilla viimeisen vuosikymmenen aikana. Erityisesti rahoittajien datapolitiikoilla on pyritty ratkaisemaan osa tutkijoiden esittämistä ongelmista. Esimerkiksi aineiston dokumentoinnin vastuu voidaan asettaa alkuperäisille tutkijoille ja rahoittajat voivat edellyttää, että jatkotutkimuksissa viitataan asianmukaisesti aineiston alkuperäisiin tekijöihin. Seuraavaksi esitetään lyhyesti esimerkkejä tieteellisten julkaisujen ja tutkimusrahoittajien datapolitiikoista. Toimintamallit ovat Yhdysvalloista ja Isosta-Britanniasta.

#### 5.4.1 Tieteellisten julkaisujen datapolitiikat

Luonnontieteiden tieteellisissä aikakauslehdissä edellytetään usein, että julkaistavassa artikkelissa analysoitu aineisto on muiden tutkijoiden saatavilla. Aineiston tulee mielellään olla arkistoitu tai suoraan verkosta ladattavissa. Tätä edellytetään tavallisesti kemian, tähtitieteen, biologian ja fysiikan aloilla. Esimerkiksi seuraavilla aikakauslehdillä on aineiston avoimuutta tukevat viralliset ohjeet ja suositukset eli datapolitiikka (data policy):

- Biophysical Journal
- Cell
- Journal of the American Chemical Society
- Journal of Biological Chemistry



- Journal of Magnetic Resonance
- Journal of Medicinal Chemistry
- Journal of Structural and Functional Genomics
- Nature
- Nature Structural & Molecular Biology
- Protein Science
- Science
- Structure

Ihmistieteiden julkaisuissa tutkimusaineistoja koskevat ohjeet ovat luonnontieteitä väljemmät. Toisinaan edellytetään, että lehden toimituskunnan on tarvittaessa saatava tarkistettavaksi julkaistavassa artikkelissa analysoitu aineisto. Lisäksi voidaan edellyttää, että tutkimusaineisto joko arkistoidaan tai mikäli se on tietosuojasyistä suotavampaa, tutkija tai tutkimuslaitos säilyttää aineistoa julkaisuajankohdasta esimerkiksi viisi vuotta eteenpäin. Näin halutaan varmistaa, että aineistoa koskevat analyysit voidaan tarvittaessa tarkistaa. Tällaista datapolitiikkaa noudatetaan esimerkiksi seuraavissa psykologian julkaisuissa:

- British Journal of Social Psychology
- British Journal of Psychology
- Journal of Occupational and Organizational Psychology
- Legal and Criminological Psychology
- British Journal of Developmental Psychology
- Psychology and Psychotherapy: Theory, Research and Practice
- British Journal of Health Psychology.

Ray ja Valeriano (2003) antavat lukuisia esimerkkejä tieteellisten lehtien datapolitiikoista. Esimerkiksi *International Interactions* (kustantaja Taylor & Francis) pyytää artikkeleita tarjoavia tutkijoita arkistoimaan aineistonsa tai muulla tavoin osoittamaan, että aineisto on artikkeleita lukevien tutkijoiden saatavilla<sup>9</sup>. Mikäli aineisto ei sisällä ihmisiä koskevia tunnisteeellisiä tietoja, voidaan myös vaatia sen julkaisemista kokonaisuudessaan. Aikakauslehdet *Journal of Peace Research* (PRIO – International Peace Research Institute) ja *Political Analyses* (Oxford journals) pyytävät artikkeleiden kirjoittajia tarjoamaan lukijoille mahdollisuuden päästä tarkistamaan analyysit alkuperäisestä aineistosta. (Mt. 77–78.)

Kaikki lehdet eivät kuitenkaan kontrolloi julkilausumansa datapolitiikan noudattamista käytännössä. Esimerkiksi Rayn ja Valerianon tutkimista *International Interactions* -lehden artikkeleista vuosilta 1999–2001 kolmannes (33 %) ei sisältänyt informaatiota siitä, mistä ja mi-

9 Lehden verkkosivujen kohdassa *Instructions for Authors* asia kerrotaan näin: "Authors of data-based articles in this journal should indicate how investigators may obtain their data, as well as other information necessary to replicate the reported results, e.g., in which public archives such information has been deposited or, in the case of multiple authors, which author to write. Authors should remove information from their datasets that must remain confidential, such as the identity of survey respondents. Investigators always have the right to first publication and, on rare occasions, may wish to embargo the data (or a portion of it) while they finish publishing from it; these authors must specify a date by which a replication dataset will be publicly available."



ten artikkelissa käytetty empiirinen aineisto on saatavilla. Yli puolet (54 %) ilmoitti ohjeiden mukaisesti, mistä aineiston saa ja loppuissa artikkeleista ei käytetty empiiristä aineistoa. Osa aineistoista saattoi olla linkitettyä suoraan artikkelin tekstiin. (Ray ja Valeriano 2003, 81.)

Tutkijakaksikko jatkoi aineistojen avoimuuden selvittämistä pyytämällä aineistot tutkijoilta, jotka olivat ilmoittaneet itsensä aineiston säilyttäjäksi ja luovuttajaksi. Näistä 23 tutkijasta vain kolme toimitti aineiston välittömästi. Muilla oli monenlaisia käytännön esteitä ja muutama oli ylipäänsä haluton luovuttamaan aineistonsa analyysien tarkistamiseen. Yllättävän moni kertoi kyllä olevansa halukas antamaan aineistonsa, mutta ilmoitti sen vievän aikaa. (Mt. 81.)

Kun tutkijat säilyttävät itse aineistonsa, sen eteenpäin luovuttaminen tarkoittaa käytännössä myös aineistoa koskevan käyttöinformaation laatimista ja itse aineiston järjestämistä luovuttettavaan kuntoon. Jos tutkijat eivät ole kuvailleet ja järjestäneet aineistoa valmiiksi tutkimuksen kuluessa tai heti sen päätyttyä, aineiston antaminen jatkokäyttöön tulee sitä työläemmäksi, mitä pitempi aika on kulunut itse tutkimuksen päättymisestä.

Itse aineistonsa säilyttäjiksi ja eteenpäin luovuttajiksi ilmoittautuvat tutkijat ovat Kingin (1995) mielestä sinänsä hyvällä asialla. Ongelmana on kuitenkin se, että tutkijat eivät hallitse riittävästi sähköisen arkistoinnin perusasioita. Lisäksi tutkijat ovat tyypillisesti määräaikaissa tehtävissä ja vaihtavat työpaikkoja, joten tietynä ajanhetkenä artikkelin yhteydessä ilmoitetut yhteystiedot voivat muuttua. Siksi King pitää järkevämpänä ratkaisuna arkistoida päättyneiden tutkimusten aineistot data-arkistoihin. (Mt. 446–447.)

## 5.4.2 Tutkimusrahoittajien datapolitiikat

Tutkimusaineistojen avoimuutta voidaan perustella sekä tieteen periaatteisiin kuuluvalla tarkistettavuudella että tiedon kumuloitumisella. Nämä tieteen arvoihin sisältyvät periaatteet yhdistetään taloudellisuuteen tutkimusrahoittajien datapolitiikoissa.

Suomen Akatemian tutkimusrahoituksen yleisissä ehdoissa on seuraavanlainen suositus: ”On suositeltavaa, että Akatemian rahoittamat tutkimushankkeet luovuttavat hanketta varten kokoamansa yhteiskuntatieteellisen aineiston Tampereen yliopistossa sijaitsevaan yhteiskuntatieteellisen tietoarkiston käyttöön. Luovutus tapahtuu mahdollisimman pian Akatemian rahoituksen päättymisen jälkeen ja aineiston vastaanottamisesta päättää tietoarkisto”. Suositusta tiukempia ehtoja aineistojen avoimuudelle ovat antaneet esimerkiksi Yhdysvalloissa National Institutes of Health ja National Science Foundation. Isossa-Britanniassa avointa datapolitiikkaa tukevat Medical Research Council ja Economic and Social Research Council.

### 5.4.2.1 *National Institutes of Health (NIH)*

Yhdysvalloissa lääketieteellistä tutkimusta ja terveydenhuoltoon kohdistuvaa muiden tieteenalojen tutkimusta rahoittava *National Institutes of Health* kohdentaa datapolitiikkansa tutkimushankkeisiin, joiden vuosittainen budjetti on vähintään 500.000 dollaria. Datapolitiikan tavoitteena on saattaa tutkimusaineistot koko tiedeyhteisön käyttöön. Tutkimusaineistot voivat sisältää ihmisiä koskevia tietoja (niin laadullisia kuin määrällisiäkin aineistoja) tai laboratorioaineistoja, sillä NIH rahoittaa useita tieteenaloja.

Ohjeiden korostetaan koskevan perus- tai pohja-aineistoa, ei niistä muokattuja tiivistelmiä, summamuuttujia, taulukoita, raportteja tai vastaavia. NIH:n avoimuusperiaate koskee erityisesti tutkimusaineistoja, jotka ovat harvinaisia ja ainutkertaisia. Kansainvälisissä tutkimushankkeissa otetaan huomioon tutkimuspartnereiden omien maiden lainsäädännön asettamat rajoitukset.

Käytännössä mittavien tutkimushankkeiden tutkimussuunnitelmiin ja hakemusasiakirjoihin tulee sisällyttää *konkreettinen suunnitelma siitä, miten kerättävä tutkimusaineisto arkistoidaan tai muutoin saatetaan koko tiedeyhteisön käytettäväksi, kun tutkimuksen päätulokset on saatettu valmiiksi*. Mikäli kerättävien tietojen arkaluontoisuus ja tietosuojakysymykset estävät koko aineiston saattamista tiedeyhteisön käyttöön, tulee perustelut ja aineiston avoimuuden rajoitukset selittää rahoitushakemuksessa. Mikäli tutkimushenkilöitä aiotaan informoida aineiston käytöstä vain tekeillä olevaan tutkimukseen, perustelut ratkaisulle tulee erikseen selittää.

NIH itse ei suosita tutkimusaineistoille tiettyjä tallenneformaatteja tai kuvailustandardeja, vaan kehottaa tutkimushankkeita noudattamaan oman tieteenalansa käytäntöjä aineistojen tallentamisessa, kuvailussa ja mahdollisessa arkistoinnissa. Tutkimusaineiston pitkäaikais säilytystä ja kuvailua koskevista ratkaisuista pitää informoida tutkimushankkeen väliraporteissa. Jo päättyneiden NIH:n rahoittamien hankkeiden aineistojen arkistointi on ehtona saada rahoitusta uusiin tutkimushankkeisiin.

NIH tunnustaa alkuperäisille tutkijoille ensikäyttöoikeuden ja oikeuden käyttää kerättyä aineistoa myös jatkotutkimuksissa. Ohjeiden mukaisesti tutkimusaineisto tulee kuitenkin saattaa koko tiedeyhteisön käytettäväksi heti, kun tutkimuksen päätulokset on julkaistu. Pitkittäistutkimusten aineistot tulee saattaa avoimiksi keruuaalloittain. Alkuperäisillä tutkijoilla ei ole valtaa jatkossa määrittää muiden tutkijoiden aineistoon kohdistamia tutkimuskysymyksiä ja analyysimenetelmiä.

Ennen aineiston saattamista tiedeyhteisön avoimeen käyttöön NIH suosittaa aineiston anonymisointia. Toisinaan anonymisointi kuitenkin vähentää oleellisesti aineiston tieteellistä käyttöarvoa ja mahdollisuuksia tarkistaa aineiston avulla tehdyt tulokset. Tällöin suositellaan aineiston käytön rajaamista. Aineiston käyttöön voidaan edellyttää lupahakemusta ja verkon kautta lataamiseen autentikointia eli käyttäjän ja käyttöoikeuksien tunnistamista ja varmistamista. Sensitiivisten tunnistellisten aineistojen käyttö voidaan myös rajata tapahtuvaksi erillisissä datakeskuksissa (data enclave), joiden tietokoneilta ei ole verkkoyhteyksiä.

#### 5.4.2.2 National Science Foundation

Niin ikään yhdysvaltalainen *National Science Foundation* (NSF) kannustaa ja rohkaisee tieteen avoimuuteen. Sen sosiaali-, käyttäytymis- ja taloustieteellinen osasto (Directorate for Social, Behavioral, and Economic Sciences SBE) on luonut näiden alojen tutkimusaineistojen avoimuutta tukevan politiikan, joka nojautuu NSF:n yleisiin periaatteisiin.

SBE tunnistaa, että eri oppiaineiden aineistot vaihtelevat suuresti aina arkeologian näytteistä surveydataan, suulliseen perimätietoon tai erilaisia kulttuureita kuvaaviin audiovisuaalisiin tallenteisiin. Tämän vuoksi *aineiston avoimeksi saattamiselle ei anneta yhtenäisiä ohjeita*. Rahoitettavissa tutkimushankkeissa kehoitetaan oppiaine- ja aineistokohtaisesti etsimään ratkaisut tutkimusaineistojen avoimuudelle ja säilyttämiselle. Tutkimushankkeissa tulee ennalta suunnitella, miten kerättävät aineistot voidaan toimittaa aineistosta kiinnostuneille tutkijoille.

Tekijänoikeudellisista tai tietosuojaan liittyvistä syistä aineistojen avoimuutta voidaan myös rajoittaa. (National Science Board 2005, 59–61.)

NSF:n rahoituspoliittisissa yleisohjeissa kehoitetaan tutkimushankkeiden vastuututkijoita huolehtimaan tutkimuksen päätyttyä aineistojen saatavuudesta muille alan tutkijoille. *Aineistojen jatkokäytön organisointiin ja suunnitteluun kehoitetaan jo tutkimuksen rahoituksen hakuvaiheessa. Siihen kiinnitetään erityistä huomioita rahoitushakemusten arvioinnissa. Siksi tutkimushankkeiden aikataulutukseen ja suunnitteluun kehoitetaan lisäämään aineistojen valmistamiseen, mahdolliseen anonymisointiin ja dokumentointiin tarvittava työmäärä ja resurssit. Poikkeukset aineistojen avoimuudelle sallitaan, kun ne on erikseen perusteltu. Tutkimusryhmiä kannustetaan jakamaan myös uudet tietotekniset sovellukset aina, kun se on tekijänoikeudellisesti erikseen sopien mahdollista. Ylipäänsä kaikkien tieteen tekemiseen liittyvien innovaatioiden avoimen jakamisen katsotaan edistävän tieteen kehittymistä.* (Ks. lisätietoa NSF: Grant Policy Manual 2005, kohta 734, Dissemination and Sharing of Research Results.)

#### 5.4.2.3 Medical Research Council/UK

Isoissa-Britanniassa *Medical Research Council* (MRC) kertoo datapolitiikkansa perustuvan OECD:n raporttiin ”Promoting Access to Public Research Data for Scientific, Economic and Social Development”. Lähtökohtana on julkisesti rahoitettujen tutkimusaineistojen avoimuus ja poikkeukset siten kuin OECD on ne määritellyt. (Kyseessä on yksi OECD:n datasuosituksen valmisteluvaiheen asiakirja.)

Datapolitiikkaa sovelletaan kaikkeen MRC:n rahoittamaan tutkimukseen. Erikseen ei kuitenkaan määritetä sitä, koska ja millä tavalla tarkkaan ottaen tutkimusaineistojen avoimuus tulee järjestää. Ensikäyttöoikeuden merkitys aineiston primaaritutkijoille eli ensikäyttäjille tunnustetaan, mutta tutkijoita kehoitetaan kuitenkin jo ennalta varautumaan aineistojen avoimuuden järjestämiseen.

Kerättävän tutkimusaineiston säilyttämisestä ja tiedeyhteisön avoimeen käyttöön saattamisesta *tulee kirjata tutkimushakemuksiin suunnitelma otsakkeella ”Data sharing and preservation strategy”*. Tämä päätös astui voimaan vuoden 2006 alusta. *Mikäli aineistoa ei voi tutkimuksen päätyttyä antaa muun tiedeyhteisön avoimeen käyttöön, se tulee rahoitushakemuksessa erikseen perustella.*

Hakemuksiin kirjattavan aineistoa koskevan suunnitelman tulee olla ytimekäs ja konkreettinen. Sen tulee sisältää yhteenvedo laadullisista ja määrällisistä aineistoista, joita tutkimuksessa aiotaan kerätä (esim. haastattelut, kyselyt, koeasetelmien aineistot, tallenteet, kuvat jne.). Lisäksi tulee kertoa jo tiedossa oleva jatkokäyttö suunnitelluille aineistoille. Myös suunnitelma aineistojen kuvailusta sekä pitkäaikais säilyttämisen ja jatkokäytön avoimuuden ratkaisuihin tulee kirjata hakemukseen. Hakemuksessa tulee selittää, mikä lisäarvo tuotetaan uudella aineistolla verrattuna jo olemassa oleviin tutkimusaineistoihin.

MRC toteaa datapolitiikkansa esittelyssä, että se uskoo rahoittajien, akateemisten instituutioiden ja aineistojen jatkokäyttäjien tunnustavan ja huomioivan asianmukaisella tavalla aineistojen alkuperäiset tekijät. Aineistojen avoimuuden tärkeyden tunnustaminen ja kunnioittaminen nähdään tärkeäksi keinoksi edesauttaa jatkotutkimuksia. Jatkokäytössä aineistoja tulee tutkia eettisesti ja tieteellisesti kestäväillä tavoilla kuten primaaritutkimuksissakin. Jatko-tutkimusten tekeminen jo itsessään katsotaan seikaksi, joka lisää käytettävien aineistojen ar-

voa. Jatkokäyttäjien yhteistyö primaaritutkijoiden kanssa katsotaan suotavaksi, jotta vastuut aineiston käytössä saadaan tarkoin sovittua.

MRC:n ohjeiden mukaisesti rahoituksen saaneet tutkimushankkeet, niiden johtajat ja vastuututkijat ovat velvollisia huolehtimaan datan keruun suunnittelusta ja käsittelystä tavalla, jossa huomioidaan koko aineiston elinkaari ja sen tulevaisuuden jatkokäyttö. Aineiston kerääjien vastuulla on myös aineiston huolellinen kuvailu. Tunnisteellisten aineistojen osalta kehoitetaan noudattamaan lakeja ja eettisiä ohjeita. Tutkittaville ja yhteiskunnalle uusista tieteellisistä tuloksista koituvien hyötyjen tulee olla suurempia kuin tutkittavien tunnistamisesta koituvat riskit. (Medical Research Council 2007.)

#### 5.4.2.4 Economic and Social Research Council/UK

*Economic and Social Research Council* (ESRC) rahoittaa Isossa-Britanniassa talous- ja sosiaalitieteiden, osin terveystieteiden ja humanististen alojen tutkimusta. Se laati datapolitiikkansa päälinjat jo 1996. Datapolitiikan noudattaminen on ehtona rahoituksen saamiselle.

ESRC tukee datapolitiikallaan arkistoinnista vastaavaa katto-organisaatiota, jonka nimi on Economic and Social Data Service ESDS. ESDS:n alaisuuteen integroidut yksiköt vastaavat tutkimusaineistojen arkistoinnista, sähköisistä tutkimusaineistojen luetteloista ja jatkokäyttöön jakamisesta. Nämä yksiköt on muodostettu vastaamaan eri aineistotyypeistä. Oma yksikkönsä on hallinnossa kerättäville aineistoille, kansainvälisille aineistoille, pitkäikäisaineistoille, laadullisille aineistoille (Qualidata) ja kvantitatiivisille aineistoille (UK Data Archive eli UKDA).

ESRC edellyttää kaikkien rahoituksen saaneiden tutkimushankkeiden tarjoavan arkistoitavaksi Qualidata-yksikköön niin paperiset kuin sähköisetkin laadulliset tutkimusaineistonsa. Aineistot voivat olla monentyyppisiä (esim. haastattelulitteraatiot, päiväkirjat, kenttämuistiinpanot, autenttisten tilojen nauhoitukset, valokuvat, lehtileikkeet jne.). Kun tutkimushankkeessa kerätään kvantitatiivista aineistoa, toimitaan vastaavasti UK Data Archive -yksikön kanssa. Arkistointi tulee suunnitella niin, että aineisto on eettisesti ja juridisesti muiden tutkijoiden jatkokäytettävissä. Siksi tutkijoita kehoitetaan ottamaan yhteyttä Qualidataan tai UKDA:an jo rahoitushakemuksen suunnitteluvaiheessa.

Mikäli suunnitellussa tutkimushankkeessa on tarkoitus kerätä uusia aineistoja, tulee rahoitushakemuksessa esittää sille perustelut ja osoittaa, että aihetta käsittelevää aineistoa ei ole jo data-arkistoissa saatavilla. Tällä halutaan varmistaa, ettei päällekkäistä tai muuten turhaa tiedonkeruuta tapahdu.

*ESRC pidättää viimeisen tutkimuksen rahoituserän siihen saakka, kunnes hankkeessa kerätyn aineiston arkistointi aloitetaan.* Mikäli aineistoa ei voi eettisesti tai muista syistä arkistoida lainkaan, asiasta pitää kirjoittaa perustelut jo rahoitushakemukseen. ESRC tukee tutkimusaineistojen arkistointia myös niin, että rahoitushakemukseen voi sisällyttää resurssit, joita tarvitaan arkistointikuntoon saattamiseksi.

Tutkimusaineistoja arkistoivissa yksiköissä opastetaan tutkijoita aineistojensa dokumentointiin ja tiedostoformaattien valintaan. Tietoarkistolla (UK Data Archive) on aineistojen käsittelyn oppaat sekä kvantitatiivisille että kvalitatiivisille aineistoille. Aineistojaan arkistoivat tutkijat määrittävät arkiston kanssa tehtävässä luovutussopimuksessa aineiston jatkokäytön ehdot. (ESRC Research Funding Guide 2005.)

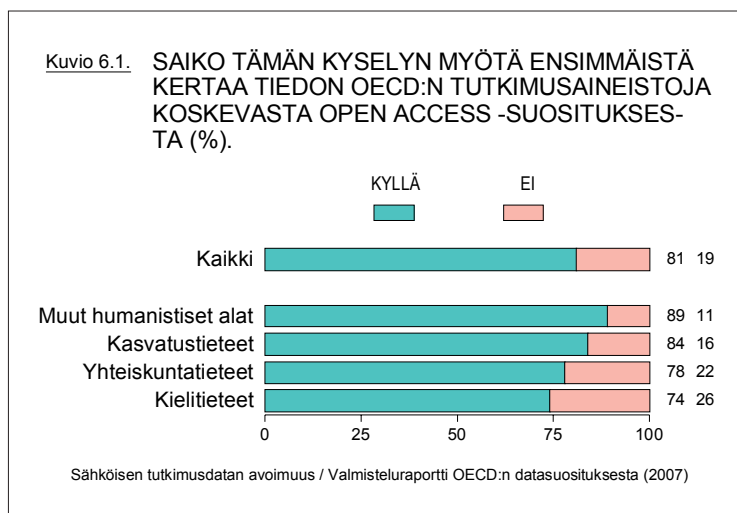
## 6 Suhtautuminen OECD:n datasuositukseen ja sen kansallisen toimeenpanon toteuttamiseen

Edellä on käynyt eri tavoin ilmi, kuinka monitahoinen kysymys julkisrahoitteisten sähköisten tutkimusaineistojen avoimen saatavuuden lisääminen on. Esimerkit tutkimusrahoittajien ja tieteellisten aikakauslehtien datapolitiikoista kuitenkin osoittavat, että avoimuutta voidaan edistää sekä julkilausutuin periaattein että konkreettisesti asettamalla aineistojen avoin saatavuus tutkimusrahoituksen ehdoksi.

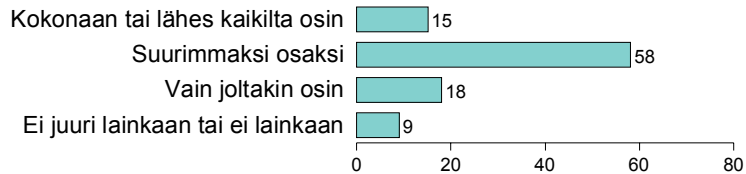
Seuraavaksi tarkastellaan tehdyn kyselyn perusteella, mitä KY-alojen professorit Suomessa ajattelevat OECD-suosituksesta ja kuinka toteuttamiskelpoisina sen päätavoitteita pidetään. Lisäksi kerrotaan vastaajien näkemyksiä siitä, keiden pitäisi olla kansallisessa toimeenpanossa suunnittelemassa ja päättämässä sekä osallistumassa keskusteluun. Professorikyselyn ohella tarkastellaan joiltakin tutkimusorganisaatioilta ja tutkimusrahoittajilta saatuja vastauksia.

### 6.1 Tutkijoiden yleisiä näkemyksiä datasuosituksesta

Vuoden 2006 lopussa toteutetussa professorikyselyssä vastaajilta tiedusteltiin, saivatko he kyselyn myötä ensimmäistä kertaa tiedon OECD:n datasuosituksesta. Tähän vastasi myöntävästi neljä viidestä vastaajasta (81 %). Erot jäivät pieniksi KY-alojen välillä (kuvio 6.1). Toinen yleisluonteinen kysymys kuului, ”missä määrin OECD:n suositukset ovat mielestänne toimeenpantavissa omalla alallanne?”. Tähän saatuja vastauksia esitellään tieteenalaryhmittäin kuviossa 6.2.



**Kuvio 6.2. MISSÄ MÄÄRIN OECD:N SUOSITUKSEN PERIAATTEET OVAT TOIMEENPANTAVISSA OMALLA ALALLA (%).**



Sähköisen tutkimusdatan avoimuus / Valmisteluraportti OECD:n datasuosituksista (2007)

Yhteensä 15 prosenttia kaikista vastaajista oli sitä mieltä, että periaatteet ovat toimeenpantavissa kokonaan tai lähes kaikilta osin. Ylivoimaisesti eniten vastauksia keräsi seuraava vaihtoehto, jonka mukaan periaatteet ovat toteutettavissa ”suurimmaksi osaksi”. Vajaa viidennes valitsi vaihtoehdon ”vain joltakin osin” (18 %) ja vain yhdeksän prosenttia oli sitä mieltä, etteivät periaatteet ole toimeenpantavissa omalla alalla juuri lainkaan tai ei lainkaan.

Myönteisimmin toimeenpanon mahdollisuuksiin uskoivat yhteiskuntatieteellisten alojen ja kielitieteiden professorit, joista kolme neljästä piti asiaa mahdollisena ainakin suurimmaksi osaksi. Lisäksi viidennes (22 %) vastanneista yhteiskuntatieteilijöistä ajatteli, että periaatteet olisivat toteutettavissa kokonaan tai lähes kaikilta osin. Kielitieteiden vastaajien joukossa tämä osuus oli vähän pienempi (14 %), mutta toisaalta heidän joukossa oli selvästi eniten niitä, jotka katsoivat periaatteiden toteuttamisen mahdolliseksi suurimmaksi osaksi (65 %).

Esimerkiksi kasvatustieteilijöihin verrattuna näiden kahden tieteenalaryhmän suhtautuminen oli selvästi myönteisempää. Kasvatustieteiden professoreista nimitäin vain viisi prosenttia tarttui kaikkein myönteisintä suhtautumista kuvaavaan vastausvaihtoehtoon. Vastauskaalan toista ääripäätä tarkasteltaessa havaitaan, että kaikkein varauksellisimmin asiaan suhtautuivat muiden humanististen alojen edustajat.

Eroihin lienee monia syitä. Yhteiskuntatieteilijöiden myönteistä suhtautumista selittänee alan tietoarkiston olemassaolo ja toiminta vuodesta 1999 lukien. Tutkimusaineistojen avoimuutta on pidetty alalla esillä viime vuosina. Kielitieteiden aineistoissa on usein vähemmän tietosuojakysymyksiä kuin esimerkiksi yhteiskunta- ja kasvatustieteiden aloilla suosituissa laadullisissa tutkimusaineistoissa. Ehkä kasvatustieteiden tuloksia selittävät alalla tyypilliset, arkaluonteiseksi koetut oppilasaineistot. Niissä vastaajat ovat helpommin tunnistettavissa kuin esimerkiksi koko väestöä koskevissa satunnaisotoksiin perustuvissa aineistoissa.

## 6.2 Yliopistojen, tutkimusrahoittajien ja valtion tutkimuslaitosten mielipiteitä OECD:n datajulistuksesta ja -suosituksesta

Suomalaisten tutkimusorganisaatioiden alustavaa suhtautumista OECD:n datajulistukseen tiedusteltiin ensi kerran jo keväällä 2005. Tuolloinen kartoitus liittyi datasuosituksen valmis-

telun alkuvaiheeseen Pariisin julistuksen jälkeen. Vastaaventyypisiä mielipidekartoituksia toteutettiin muissakin järjestön jäsenmaissa. Suomessa kysely lähetettiin tässä vaiheessa seuraaviin valtion tutkimuslaitoksiin ja -organisaatioihin.

- Kansanterveyslaitos
- Tekes
- Geodeettinen laitos
- Geologian tutkimuskeskus
- Ilmatieteen laitos
- Kansaneläkelaitos
- Kansanterveyslaitos KTL
- Kotimaisten kielten tutkimuskeskus
- Kuluttajatutkimuskeskus
- Maa- ja elintarviketalouden tutkimuskeskus
- Merentutkimuslaitos
- Metsäntutkimuslaitos
- Museovirasto
- Riista- ja kalatalouden tutkimuslaitos
- SITRA
- Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus STAKES
- Suomen Akatemia
- Suomen Pankin tutkimusosasto
- Suomen ympäristökeskus
- Säteilyturvakeskus
- Tilastokeskus
- Työterveyslaitos
- Valtion taloudellinen tutkimuskeskus
- Valtion teknillinen tutkimuskeskus VTT
- Tieteen tietotekniikan keskus CSC

Puolet näistä yhteensä 25 organisaatiosta vastasi asiaa valmistelleen OECD-sihteeristön muotoilemaan suppeaan kyselyyn. Kysymykset ja vastaukset on koottu liitteeseen 3. Tässä yhteydessä niistä kootaan vain muutamia yleisiä päätelmiä.

Valtion tutkimusorganisaatioiden vastauksissa näkyy se tosiseikka, että eri aloilla aineistot ovat hyvin erilaisia ja että niiden jatkokäyttömahdollisuuksia säätelevät monet lait. Ajatusta sähköisten aineistojen avoimuuden lisäämisestä kannatetaan periaatteessa. Monilla aloilla on paljon seikkoja, jotka rajoittavat käytännössä aineistojen jatkokäyttöä.

Useissa vastauksissa korostuivat informoidun jatkokäytön ja käyttäjätuen vaatimukset. Lisäksi monet kyselivät maksajaa kustannuksille, joita avoimuuden lisäämisestä koituisi. Keskeisin vastausten sanoma lienee, että asia vaatii alakohtaista pohdintaa sekä ala- ja organisaatiokohtaisia ratkaisuja.<sup>10</sup>

10 Kartoituksessa mukana olleet VTT ja KTL olivat Yhteiskuntatieteellisen tietoarkiston ohella mukana vastaamassa myös OECD-sihteeristön laatimaan jatkokyselyyn datasuosituksesta vuonna 2006. Tämän suppean tiedustelun tuloksia ei käsitellä tässä yhteydessä, koska vastausten päälinjau oli sama kuin aiemmassa tiedustelussa.



Kyselyyn osallistuneiden valtion tiedontuottajaorganisaatioiden mielipiteet eivät edusta tyhjentävästi eri alojen näkemyksiä. Myöskään yliopistojen ja tutkimusrahoittajien mielipiteitä ei voida nyt tarkastella kattavasti, koska asiaa varten suunnitellut aineistot jäivät puutteelliseksi. Professorikyselyä sisällöllisesti muistuttavat tiedonkeruut toteutettiin loppuvuodesta 2006 postikyselyinä myös yliopistoille ja keskeisille tutkimusrahoittajille. Lomakkeita lähetettiin yksi kuhunkin organisaatioon ja vastausaktiivisuus jäi matalaksi (ks. liite 2). Vastaaminen lienee koettu lausunnon antamiseksi hankalasti mielletävään ja tosiasiaa keskeneräiseen asiaan. Siksi monet eivät halunneet ottaa vielä kantaa asiaan. Kaikkiaan lomakkeita palautui tutkimusrahoittajilta ja yliopistoilta molemmilta kymmenkunta.

Yliopistojen vastaajat olivat tutkimuksesta vastaavia rehtoreita tai heidän nimeämiään henkilöitä. Kahdeksasta vastaajasta vain kaksi ilmoitti, että heidän yliopistollaan oli erityisiä ohjeita tai sääntöjä sähköisten tutkimusaineistojen säilyttämiselle. Vastaajista viisi sai kuulla kyselyn myötä OECD:n datasuosituksesta ensi kertaa. Yliopistovastaajista neljä oli sitä mieltä, että OECD:n datasuosituksen periaatteet ovat toimeenpantavissa heidän omalla tutkimusalallaan vain joiltakin osin. Kaksi katsoi suosituksen olevan toimeenpantavissa joko kokonaan tai suurimmaksi osaksi. Ainoastaan yksi vastaaja katsoi, ettei suositus ole lainkaan toimeenpantavissa.

Yhdeksää eri tutkimusrahoittajaa edustaneiden vastaajien mielipiteet olivat hieman suopeammat suositusta kohtaan. Heistä neljä sanoi, että suositus olisi toteutettavissa heidän rahoittamillaan aloilla suurimmaksi osaksi. Kolme valitsi vaihtoehdon ”toimeenpantavissa vain joiltakin osin”, kun taas kaksi katsoi, ettei toimeenpano ole lainkaan mahdollista. Viimeksi mainitut kaksi lomaketta tulivat hammaslääketieteen alalta.

### 6.3 OECD:n datasuosituksen toimeenpano: mielipiteitä keinoista ja ensisijaisista vastuutahoista

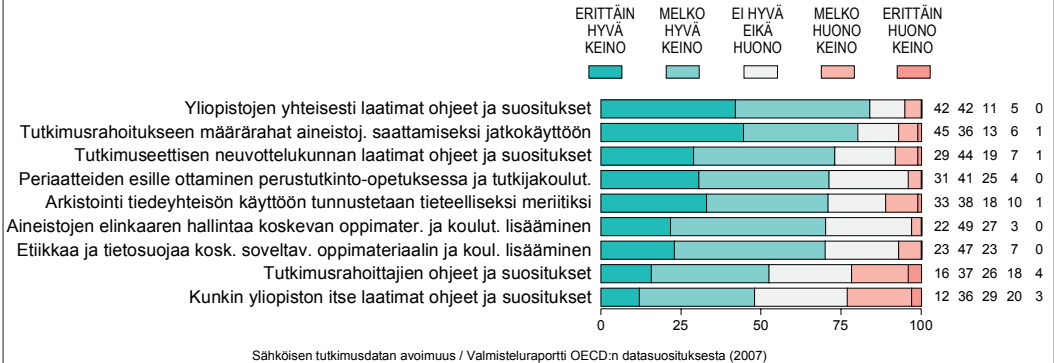
Kyselytietojen analysoinnin lopuksi kuvioissa 6.3–6.5 tarkastellaan KY-alan professoreiden näkemyksiä OECD:n datasuosituksen toimeenpanon mahdollisista keinoista ja kansallisen toimeenpanon vastuutahoista. Samat kysymykset esitettiin myös yliopistojen ja tutkimusrahoittajien vastaajille. Kuvioon 6.3 listatut keinot eivät ole toisensa poissulkevia, vaan niiden käyttö on mahdollista ja ehkä toivottavaakin rinnakkain.

Tutkijoiden mielestä kaksi keinoa nousee jonkin verran yli muiden. Erittäin hyvänä pidettiin yli 40 prosentissa vastauksia ainoastaan kahta keinoa: yliopistojen yhteisesti laatimia ohjeita ja suosituksia sekä tutkimusrahoitukseen sisällytettäviä määrärahoja aineistojen saattamiseksi jatkokäyttöön. Lisäksi yli kolmannes vastaajista piti näitä myös melko hyvinä keinoina, joten yhteenlaskettuna näitä piti hyvinä yli 80 prosenttia KY-alan professorivastaajista.

Seuraavassa viiden keinon ryppäessä jokainen yksittäinen toimenpide sai myös enemmistön kannatuksen. Järjestys oli seuraava: tutkimuseettisen neuvottelukunnan eli TENK:n ohjeet ja suositukset, periaatteiden esille ottaminen opetuksessa, arkistoinnista koituvan meritoitumisen esille nostaminen, aineistojen elinkaaren hallintaan liittyvä koulutus sekä tutkimuseettikkaa ja tietosuojaa koskevan koulutuksen lisääminen. Tutkijoiden mielestä näitä vähämerkityksisempiä olisivat tutkimusrahoittajien ohjeet ja suositukset sekä yksittäisten yliopistojen toimenpiteet.



**Kuvio 6.3. KUINKA HYVIÄ TAI HUONOJA ERI KEINOT OLISIVAT TUTKIMUSAINEISTOJEN OPEN ACCESS -PERIAATTEIDEN TOIMEENPANEMISEKSI (tekstejä osin lyhennetty, %).**



Kukaan yliopistovastaaja ei pitänyt yksittäisten yliopistojen laatimia ohjeita ja suosituksia erittäin hyvänä keinona edistää asiaa. Sitä vastoin yliopistovastaajat näyttivät katsovan toiveikkaasti erityisesti tutkimusrahoittajien suuntaan, sillä enemmistö vastaajista piti rahoittajien suosituksia erittäin hyvänä tapana viedä asiaa eteenpäin.

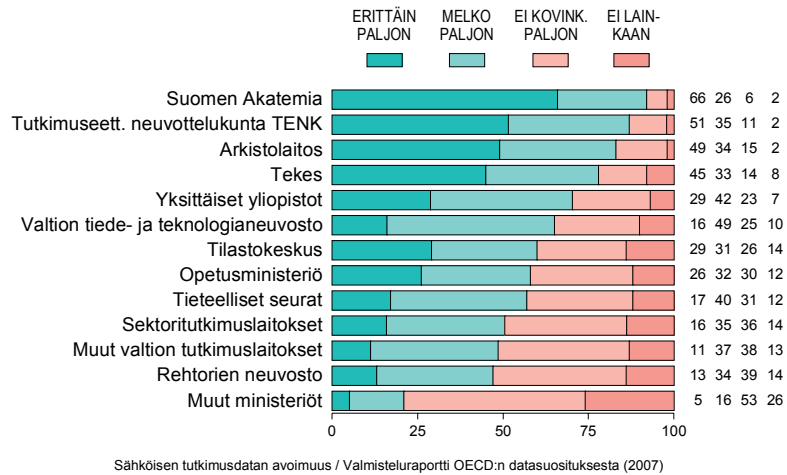
Tutkimusrahoittajat kannattivat paitsi oman toimintansa terävöittämistä asiassa, myös erityisesti yliopistojen yhteisesti laatimia ja tutkimuseettisen neuvottelukunnan laatimia ohjeita ja suosituksia. Lisäksi enemmistö tutkimusrahoittajavastaajista liputti koulutuksen puolesta asian edistämiseksi.

Lopuksi sekä tutkijavastaajia että yliopistojen ja tutkimusrahoittajien edustajia pyydettiin arvioimaan sitä, missä määrin eri tahojen tulisi olla mukana laatimassa ohjeita julkisrahoitteisen sähköisten tutkimusaineistojen käytön avoimuudelle. Aluksi vastaajat arvioivat kunkin tahon roolia neliportaisella skaalalla, joka vaihteli vaihtoehdosta ”erittäin paljon” vaihtoehtoon ”ei lainkaan”. Tämän jälkeen heitä pyydettiin vielä nimeämään listatuista kolmestatoista tahosta mielestään kolme tärkeintä.

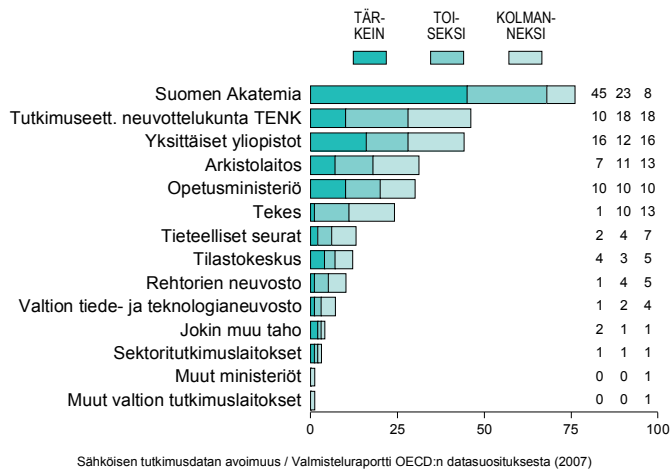
KY-alan professoreiden vastauksissa ylivoimaisesti eniten kannatusta sai Suomen Akatemia, jonka valitsi ohjeiden laatimisessa kolmen keskeisimmän tahon joukkoon kolme neljästä vastaajasta (ks. kuvio 6.4). Samalla kaksi kolmesta vastaajasta oli sitä mieltä, että Suomen Akatemian tulisi olla mukana asiassa erittäin paljon. Toiseksi suosituimmaksi tahoksi osoittautui sekä kokonaisjakaumissa että kolmen tärkeimmän tahon vertailuissa tutkimuseettinen neuvottelukunta.

Kärkitahojen suosion ohella on merkillepantavaa, että selvästi yli puolet listalla olleista tahoista keräsi enemmistön kannatuksen. Kaikkiaan asia vaatii siis tutkijoidenkin mielestä hyvin monien tahojen aktiivista yhteistyötä. Sama perusviesti oli luettavissa myös yliopistojen ja tutkimusrahoittajien vastauksista.

**Kuvio 6.4. MISSÄ MÄÄRIN ERI TAHOJEN TULISI OLLA MUKANA LAATIMASSA OHJEITA JULKISRAHOITTEISTEN SÄHKÖISTEN TUTKIMUSAINEISTOJEN KÄYTÖN AVOIMUDELLE (%).**



**Kuvio 6.5. MISSÄ MÄÄRIN ERI TAHOJEN TULISI OLLA MUKANA LAATIMASSA OHJEITA: KOLME TÄRKEINTÄ VASTUUTAHOA (%).**



## 7 Näkökulmia OECD:n datasuosituksen toimeenpanoon

Suomalaiset yliopistot ja muut tutkimusorganisaatiot sekä tutkijat ja tutkimusrahoittajat näyttävät siis periaatteessa suhtautuvan myönteisesti julkisrahoitteisen sähköisen tutkimusdatan avoimuuden lisäämiseen. Yhteiskunnan sekä kansallisen ja kansainvälisen tiedeyhteisön kokonaisuus on, että tutkimusaineistojen elinkaari muotoutuu mahdollisimman tarkoituksenmukaiseksi. Näin turvataan julkisten tutkimusinvestointien tehokkuus ja yleinen vaikuttavuus sekä vahvistetaan tutkimusjärjestelmän tuottavuutta.

Eri tieteenalat ja julkisvaroin toteutettavat tutkimusalat toimivat toisistaan poikkeavissa ympäristöissä ja ne tutkivat hyvin erilaatuisia asioita. Tutkimuksen kohteet ja päämäärät, tietojen luottamuksellisuus, tulosten kaupallinen hyödynnettävyys, tutkimuksen rahoitusrakenteet sekä tutkimusaineistojen käyttöön tarvittava välineistö ja hyödynnettävyys jatkotutkimuksissa vaihtelevat paljon.

Myös kansallinen lainsäädäntö koskee eri tavoin eri alojen tutkimustoimintaa. Lisäksi valtioiden tutkimusta koskevat lait ja säännökset voivat poiketa toisistaan paljonkin. Tästä syystä on ymmärrettävää, että OECD:n datasuosituksessa tutkimusaineistojen avoimuutta pyritään lisäämään yleisin periaattein rajautuen julkisrahoitteisiin tutkimusdatoihin.

Päättöluvussa koostetaan edeltävissä luvuissa esiteltyjä tuloksia ja toimintamalleja. Niitä pelitään OECD:n datajulistuksen ja datasuosituksen periaatteisiin. Tätä kautta jäsennetään ja konkretisoidaan datasuosituksen aihealueita. Toivottavasti datasuosituksen toimeenpanoa koskeva suomalainen keskustelu ja kansalliset jatkotoimet pääsevät näin mahdollisimman varhain käytännön tasolle.

### 7.1 Yhteenveto tutkimusdatan elinkaaren keskeisistä tekijöistä

Raportin toisessa luvussa sähköisen tutkimusdatan elinkaari jäsennettiin kymmeneen vaiheeseen (kaavio 2.1). Elinkaarimalli soveltuu erityisesti yliopistoissa tehtävään tutkimukseen, mutta se lienee käyttökelpoinen muunkinlaisten tutkimushankkeiden datan elinkaaren kuvaamiseen. Seuraavaan on tiivistetty keskeisten vaiheiden olennaisimpia aiheita sähköisten tutkimusaineistojen avoimuuden kannalta.

#### *Tutkimussuunnitelma*

Tutkimuksen yleinen suunnitelmallisuus ja huolellinen valmistelu edistävät tutkimusaineistojen avoimuutta ja pitkää elinkaarta. Tutkimukselta pitää edellyttää huolellisesti laadittua kirjallista suunnitelmaa. Siinä tulee ainakin lyhyesti yksilöidä myös seuraavat tutkimusaineistoihin liittyvät asiat, jos tutkimus niitä edellyttää:

- tutkimuskohteita koskevat informoinnit sekä suostumus-, lupa- ja sopimusasiakirjat, mikäli aineisto kohdistuu henkilöihin
- tutkimusaineistoon liittyvät omistus-, hallinta- ja tekijänoikeudet, ainakin jos ne poikkeavat tutkijaryhmän tutkimusorganisaation normaalikäytännöistä

- suunnitelma tutkimusaineiston ensikäyttövaiheen aikaisesta säilyttämisestä, käyttötavoista ja käyttöoikeuksista
- suunnitelma tutkimusaineiston jatkokäytön mahdollistamisesta ensikäytön jälkeen.

### *Eettinen arviointi*

Myös tutkimuksen eettinen arviointi tukee eri tavoin tutkimusaineistojen elinkaarta aineistojen ensikäytöstä (primary use) niiden jatkokäyttöön (secondary use). Joillekin tieteenaloille eettinen ennakoarviointi on jo institutionalisoitunut. Esimerkiksi sairaanhoitopiirien eettinen arviointi kohdistuu tutkimussuunnitelmaan, jonka pitää liitteineen sisältää yksilöidyt aineistotiedot, tiedonkeruustrumentit, tietoja tutkimusrahoituksesta sekä esimerkiksi informointi- ja tutkimussuostumusasiakirjat. Myös tutkijoiden ja tutkimusryhmien omaehtoinen tutkimuseettinen arviointi voi tukea merkittävästi tutkimusdatan avointa saatavuutta ja elinkaarta hankkeiden suunnitteluvaiheessa.

### *Tutkimusrahoittajien tukitoimet*

Tutkimusrahoittajat voivat tukea tutkimusdatan avoimuutta suosittamalla tai edellyttämällä ja myös rahoittamalla sitä, että hankkeissa kertyvät tutkimusaineistot luovutetaan tiedeyhteisön avoimeen jatkokäyttöön ensikäytön päätyttyä. Rahoittajien mahdollisia tukivälineitä ovat muun muassa

- yleisten tutkimuseettisten periaatteiden laatiminen ja esillä pito
- oman, rahoitettavia hankkeita sitovan datapolitiikan laatiminen
- tutkimusaineistojen arkistoinnin vaatiman työn rahoittaminen osana hankerahoitusta (aineiston valmisteleminen arkistoitavaksi).

### *Jatkokäyttömahdollisuuksien turvaaminen datan keruuvaiheessa*

Keruvaihe on ratkaisevin tutkimusdatan elinkaarelle ja avoimelle saatavuudelle, sen jälkeen moniin aineiston jatkokäytön kannalta olennaisiin asioihin ei voida enää vaikuttaa. Kokoamisen aikana tehdyt tai tekemättä jääneet sopimukset, sitoumukset ja informoinnit, itse aineiston koonnin onnistuminen sekä keruun tulosten ja keruuvaiheen toimintojen kirjaaminen vaikuttavat suoraan siihen, missä määrin aineistoa voidaan käyttää erilaisiin tarkoituksiin ensi- ja jatkokäyttövaiheissa.

### *Ensikäyttö*

Tutkimusaineiston ensikäyttö tarkoittaa alkuperäisessä tutkimussuunnitelmassa ilmaistujen käyttötarkoitusten mukaista käyttöä. Ensikäytön kesto on usein vaikea määritellä tarkasti tutkimushanketta käynnistettäessä. Tämän ei kuitenkaan pidä johtaa siihen, että sitä ei määritellä lainkaan. Ensikäyttövaiheessa tutkimusaineisto tulee dokumentoida riittävän ajantasaisesti ja huolehtia sen asianmukaisesta säilyttämisestä.

Tutkimusdatan avoimen saatavuuden kannalta on olennaista tietää, kuka tai ketkä päättävät aineiston jatkokäyttöön luovuttamisesta sekä milloin ja miten jatkokäyttö on mahdollista. Epäselvät vastuusuhteet sekä ensikäytön pitkittyminen ja laajeneminen voivat johtaa tutkimusdatan alikäyttöön.

### *Datan arkistointi ja julkaiseminen*

Ilman suunnitelmallista säilyttämistä sähköisten tutkimusaineistojen käytettävyys saattaa heikentyä ratkaisevasti jo muutamassa vuodessa. Tätä pidemmän ajan kuluessa aineistot voivat jopa tuhoutua kokonaan ilman asianmukaisia säilyttämistoimia. Sähköisten aineistojen elinkaaren hallinta edellyttää päätöksiä säilytettävistä aineistoista sekä pitkäaikaissäilytystä eli arkistointia jollakin tarkoituksenmukaisella tavalla. Tähän liittyvät muun muassa dokumentointiin ja pitkäaikaissäilyttämiseen soveltuvien standardien noudattaminen.

Joissakin tapauksissa on parasta, että aineiston koonnut organisaatio säilyttää tutkimusaineiston itse ja huolehtii mahdollisuuksista sen jatkokäyttöön. Näin myöhemmillä käyttäjillä on mahdollisuus saada aineistoista lisätietoa suoraan sen keränneiltä tahoilta.

Pitkäaikaissäilytyksen ja jatkokäytön kannalta selvästi huonompi vaihtoehto on, että vastuu säilytyksestä ja jatkokäyttöön luovuttamisesta jää yksittäisille tutkijoille. Tämän Suomen yliopistoissa ja muissakin tutkimusorganisaatioissa vallitsevan nykykäytännön vähittäinen muuttaminen soveltuisi OECD:n datasuosituksen kansallisen toimeenpanon yhdeksi perusmissioksi.

Tutkimusdatan avoimen saatavuuden, elinkaaren turvaamisen ja aineistoinvestointien kokonaistehokkuuden kannalta on varmintaa, että datan tallentavat jatkokäyttöä varten alan erityisosaajat. Tähän tarvitaan joko tutkimus- ja tiedonkeruuorganisaatioiden sisällä toimivia erityisasiantuntijoita tai keskitettyjä arkistointiratkaisuja tarjoavia palveluyksiköitä, joihin voidaan ulkoistaa datan pitkäaikaissäilyttäminen ja jatkokäyttöön jakelu.

Usein on tarkoituksenmukaista turvautua erityisosaamisen keskittämiseen, koska silloin tutkimusdata dokumentoidaan varmimmin kansallisten tai kansainvälisten standardien mukaan ja aineisto voidaan saattaa tehokkaasti myös laajaan kansainväliseen jatkokäyttöön.

Hyvin oleellista on myös se, että data-arkistossa tutkimusdatan kuvailutiedot *julkaistaan* ja sille annetaan jatkokäyttöä helpottavat viitetiedot ja olennaiset oheistiedot. Tietoarkistot vastaavat myös aineistotietojen sekä jatkokäytössä kertyvien julkaisujen viitetietojen kokoamisesta tietokantoihin. Näin tutkijat voivat perehtyä julkaisujen lisäksi myös aineistojen kautta alansa aikaisempaan tutkimukseen.

### *Jatkokäytön tuki*

Tutkimusdatan laaja käytettävyys edellyttää, että tutkimusaineisto on tietosisällöltään ja muulta laadultaan tarkoituksenmukainen ja riittävästi dokumentoitu. Mahdollisuus yhdistää sähköinen aineisto muihin tutkimusdatoihin on niin ikään tärkeä. Henkilöaineistoissa käyttäjien tulee kuitenkin perehtyä yhdistämistä rajoittaviin säännöksiin.

Nykyisin myös luottamuksellisia tietoja voidaan käsitellä riittävän turvallisesti erilaisissa online-sovelluksissa, kunhan tietojärjestelmät ja käyttöliittymät takaavat käyttäjien luotettavan rekisteröitymisen, käyttöoikeuksien määrittämisen ja tunnistamisen sekä aineistojen tietosisältöihin liittyvien tietosuojakysymysten hallinnan.

Luottamuksellisimpien sähköisten aineistojen saatavuutta tutkimuksen käyttöön voidaan tukea myös onsite-ratkaisuilla. Tällöin aineistoja käytetään paikan päällä aineiston sijaintiorganisaatioissa, joka vastaa aineiston tietojen luottamuksellisuuden säilymisestä sekä neuvoo aineistojen jatkokäyttäjiä.

## 7.2 Yhteenveto raporttia varten kootuista kyselyistä

### 7.2.1 Kulttuurin ja yhteiskunnan tutkijoiden suhtautuminen tutkimusdatan avoimuuden esteisiin ja haittoihin

Tätä selvitysraporttia varten toteutettiin marras-joulukuussa 2006 verkkokysely, johon osallistui 150 kulttuurin ja yhteiskunnan tutkimuksen eli KY-alan professoria eri yliopistoista. Kyselyyn vastasi noin kolmannes tavoitelluista henkilöistä. Saadun sähköpostipalautteen perusteella kyselyyn vastasivat aktiivisimmin ne professorit, joiden laitoksilla ja oppialoilla käytetään sähköisiä aineistoja. Vastaukset tuottivat seuraavia perushavaintoja, jotka ovat monin osin yleistettävissä muillekin tieteenaloille:

Vain hyvin harvoilla yliopistolaitoksilla on ohjeita tai toimivia käytäntöjä sähköisen tutkimusdatan säilyttämiselle ja jatkokäytölle. Tutkimusdatat jäävät lähinnä tutkijoiden itsensä haltuun. Kolmanneksella vastaajien laitoksista sähköiset aineistot jäävät sellaisinaan, mutta keskitetysti, tutkijan laitokseen tai yksikköön.

Datojen keskitetty arkistointi yliopistoissa tai niiden laitoksilla on harvinaista. Tutkimusaineistojen säilyttämistä ei ole (yliopistoissakaan) ohjeistettu riittävästi.

Yli puolet professoreista koki sähköisten aineistojen jatkokäytön tärkeäksi esteeksi sen, että aineistojen tietosisällöt ja tiedostot ovat puutteellisesti dokumentoituja. Noin joka toinen arvioi aineistojen avoimuudessa haitaksi sen, että tutkijoiden työaikaa menisi vanhojen aineistojen muokkaamiseen käyttökelpoiksi. Yhtä moni arvioi suureksi haitaksi jatkokäyttäjien neuvontaan mahdollisesti kuluvan työajan.

Lähes puolet vastaajajoukosta katsoi tärkeäksi tai melko tärkeäksi syyksi aineistojen jatkokäytön vähäisyydelle sen, että aineistojen omistussuhteista ja käytöstä ei ole sovittu aineistoja kerättäessä. Joka toinen vastaaja piti melko tai hyvin suurena haittana sitä, että jatkotutkimuksissa ei osata käyttää aineistoja oikein. Kaksi viidestä piti toisten kokoamien aineistojen soveltumattomuutta nykyisiin tutkimusongelmiin tärkeänä syynä sille, että tutkimusdatat ei käytetä jatkotutkimuksissa. Kolmannes piti lainsäädäntöön, esimerkiksi tekijänoikeuteen liittyviä ongelmia tärkeänä tai melko tärkeänä syynä sille, ettei tutkimusaineistoja käytetä jatkotutkimuksissa. Neljännestä vastaajista katsoi kilpailuedun menettämisen tutkimusaineistojen avoimuudessa suureksi tai melko suureksi haitaksi.

Kaksi kolmesta professorista piti merkittävänä ongelmana tapauksia, joissa tutkittaville ei ole kerrottu, että heiltä kerättävät tiedot arkistoidaan tutkimuksen päätyttyä koko tiedeyhteisön käyttöön. Lähes puolet vastaajista arvioi myös tietosuojariskien ja tutkimuseettisten riskien lisääntyvän aineistojen avoimuuden myötä.

### 7.2.2 Tutkimusdatan jatkokäytön nykytilanne yliopistoissa

Raportissa tarkasteltiin myös tutkimusdatan jatkokäytön nykytilannetta kulttuurin ja yhteiskunnan tutkimuksessa. Lisäksi esiteltiin joitakin avoimuutta tukevia hyviä käytäntöjä muilta tieteenaloilta. KY-alan professorikyselyn nojalla voitiin tehdä seuraavia suuntaa antavia havainnot:

Valtaosalla laitoksista useimmat viimeksi kuluneiden kymmenen vuoden aikana kertyneistä aineistoista eivät ole enää jatkokäytettävissä. Tästä huolimatta sähköisiä aineistoja käytetään myös ensikäytön jälkeen. Jatkokäyttäjiksi ovat tavallisimmin ensikäyttäjät itse tai laitoksen opiskelijat. Mutta vain joka kymmenes professori yhtyi varauksetta väitteeseen, että heidän laitoksellaan tai yksikössään tutkimusdataa luovutetaan pyydettyä samaa aihepiiriä tutkiville tutkijoille. Joka toinen vastaaja katsoi pitävän paikkansa, ettei sähköisiä aineistoja juurikaan hyödynnetä heidän laitoksellaan tutkimusten päätyttyä.

### 7.2.3 Kulttuurin ja yhteiskunnan tutkijoiden suhtautuminen tutkimusdatan jatkokäyttöön ja avoimuuden hyötyihin

Kulttuurin ja yhteiskunnan tutkimuksen alan professorit arvioivat tutkimusdatan avoimuuden hyödyt kauttaaltaan avoimuuden haittoja suuremmiksi. Suurimmaksi hyödyksi tunnustettiin monipuolisten tutkimusasetelmien mahdollistuminen. Muiksi hyvin merkittäviksi hyödyksi katsottiin päällekkäisen tiedonkeruun väheneminen, taloudelliset hyödyt aineistojen jatkokäytön tehostuessa, tiedeyhteisön sisäisen vuorovaikutuksen ja yhteistyön lisääntyminen sekä tasa-arvoisempi aineistojen saatavuus eri käyttäjäryhmille.

Joka neljäs kyselyyn vastannut professori katsoi, että selvä enemmistö oman alan tutkijoista kannattaa aineistojensa avoimuuden lisäämistä. Kolmannes oli ehdottoman myönteinen omien sähköisten aineistojen avoimelle jatkokäytölle ja yli 40 prosenttia suhtautui asiaan melko myönteisesti.

Kun vastaajille kuvattiin esimerkkitapaus mahdollisesta aineiston luovuttamisesta arkistoitavaksi ja jatkokäyttöön, kolme neljästä vastaajasta katsoi, että tutkimusryhmälle tulisi suosittaa aineiston luovuttamista tiedeyhteisön avoimeen käyttöön. Ylivoimainen enemmistö kyselyyn osallistuneista yliopistoista ja tutkimusrahoittajistakin olisi suosittanut kuvatussa tilanteessa tutkimusaineiston luovuttamista tiedeyhteisön avoimeen käyttöön.

### 7.2.4 Kooste avoimuuden järjestämistä ja toimintavastuita koskevista mielipiteistä

Kyselyyn vastanneet pitivät data-arkistoja ensisijaisena vaihtoehtona toteuttaa tutkimusaineistojen arkistointi ja jatkokäyttöön luovutus. Parhaimmaksi tavaksi arvioitiin valtakunnalliset tieteenalakohtaiset data-arkistot. KY-alan professorivastaajat pitivät huonoimpana ratkaisuna vaihtoehtoa, jossa tutkija itse hoitaa omien aineistojensa arkistoinnin ja jatkokäyttöön luovuttamisen.

Raportin kuudennessa luvussa koottiin suuntaa antavasti vastaajien näkemyksiä avoimuuden yleisperiaatteista ja järjestämismahdollisuuksista. Kävi ilmi, että *OECD:n datasuositus* on KY-alan professoreiden mielestä *enimmäkseen toimeenpantavissa* heidän omalla alallaan. Vain runsas neljännes vastaajista oli eri mieltä asiasta.

Valtion tutkimusorganisaatioiden alustavaa suhtautumista OECD:n datajulistukseen tiedusteltiin keväällä 2005 (liite 3). Vastauksissa heijastuivat sähköisten aineistojen alakohtaiset tyyppierot sekä vallitsevan lainsäädännön erilaiset vaikutukset organisaatioiden toimintaan.

Ajatusta sähköisten aineistojen avoimuuden lisäämisestä kannatettiin periaatetasolla. Toisaalta monilla aloilla on erilaisia syitä asettaa sähköiselle tutkimusdatalle käyttörajoituksia. Näkemyksissä korostui tarve myöhempien käyttäjien informointiin (jatkokäytön tuki) sekä avoimuuden lisäämisen aiheuttamien kustannusten kattamiseen. Asian katsottiin vaativan ala- ja organisaatiokohtaisia ratkaisuja.

Vuoden 2006 lopulla tavoitetut ja kyselyyn vastanneet yliopistojen edustajat vahvistivat tutkijakyselystä tehdyn havainnon, että yliopistolla ei yleensä ole erityisiä ohjeita tai sääntöjä sähköisten tutkimusaineistojen säilyttämiselle.

Professorikyselyssä kartoitettiin lisäksi näkemyksiä OECD:n datasuosituksen toimeenpanon mahdollisista keinoista ja kansallisen toimeenpanon vastuutahoista. Samoihin kysymyksiin vastasivat yliopistot ja tutkimusrahoittajat. Professoreiden mielestä *julkisrahoitteisen tutkimusdatan avoimuutta edistäisivät parhaiten yliopistojen yhteisesti laatimat ohjeet ja suositukset sekä tutkimusrahoitukseen sisällytettävät määrärahat aineistojen saattamiseksi jatkokäyttöön.*

Enemmistö kannatti myös seuraavia tukitoimia:

- tutkimuseettisen neuvottelukunnan eli TENK:n ohjeet ja suositukset
- periaatteiden esille ottaminen opetuksessa
- tutkimusdatan arkistoinnista koituvan meritoitumisen esille nostaminen
- aineistojen elinkaareen liittyvä koulutus sekä tutkimusetiikkaa ja tietosuojaa koskevan koulutuksen lisääminen.

Yliopistojen ja tutkimusrahoittajien vastaukset painottuivat hieman eri tavoin. Yksikään yliopistojen edustajista ei pitänyt yksittäisten yliopistojen laatimia ohjeita ja suosituksia erittäin hyvänä keinona edistää asiaa. He ilmaisivat tärkeämmäksi tutkimusrahoittajien roolin. Enemmistö piti rahoittajien suosituksia erittäin hyvänä keinona edistää asiaa. Lisäksi yliopistovastauksissa kannatettiin yliopistojen yhteisesti laatimia ja tutkimuseettisen neuvottelukunnan laatimia ohjeita ja suosituksia. Enemmistö tutkimusrahoittajavastaajista kannatti puolestaan koulutuksen vahvistamista.

Eri vastaajaryhmiltä tiedusteltiin myös, missä määrin tiedeorganisaatioiden tulisi olla mukana laatimassa ohjeita julkisrahoitteisten sähköisten tutkimusaineistojen käytön avoimuudelle. KY-alan *tutkijoiden vastauksissa kärkisijoille ylsivät Suomen Akatemia ja tutkimuseettinen neuvottelukunta.* Yliopistojen ja tutkimusrahoittajien edustajien vastauksissa ei löytynyt näin yhtenäistä linjaa, vaan vastaukset vaihtelivat vastaajan tieteenalan mukaan. Avoimuuden edistäminen vaatiikin jatkossa alakohtaisia toimenpiteitä.

### 7.3 Yhteenveto aineistojen säilyttämiskäytännöistä yliopistoissa ja tutkimusorganisaatioissa

Tutkimusdatan säilyttämistä koskeva ja avoimuutta tukeva ohjeistus on yliopistoissa harvinaista. Julkisrahoitteisen tutkimusdatan säilyttämiseen ja jatkokäyttöön liittyvä ohjeistus sisältyy, jos sisältyy, arkistonmuodostussuunnitelmiin, laatukäsikirjoihin tai muihin laitostason sääntöihin tai ohjeisiin.



Sopimusoikeuden ja arkistolain perusteella yliopistojen arkistojen tulisi säilyttää julkisin varoin toteutettujen tutkimushankkeiden asiakirjojen lisäksi julkisin varoin hankitut tutkimusaineistot, ellei muusta ole tutkimusta aloitettaessa erikseen tutkimussopimuksissa tai muutoin kirjallisesti sovittu.

Käytännössä tutkimusaineistojen käsittely ja säilyttämisvastuu on jäänyt laitostasolle, ja sielläkin lähinnä tutkimusryhmien ja yksittäisten tutkijoiden tehtäväksi. Harvoja poikkeuksia lukuun ottamatta tutkimusdatan pitkäaikaissäilytyksestä ja jatkokäyttöön luovuttamisesta vastaavat yleensä tutkijat itse. Vain joillakin suurimmilla ja pitkäkestoisilla tutkimushankkeilla on ollut mahdollisuus organisoida jatkokäyttöön luovutus asianmukaisesti. Tällöinkään aineistojen pitkäaikaissäilytyksestä ja dokumentoinnista ei välttämättä ole huolehdittu riittävästi.

Suomen Akatemia suosittaa rahoittamiensa yhteiskuntatieteellisten tutkimushankkeiden tutkimusaineistojen arkistoinnista Yhteiskuntatieteelliseen tietoaarkistoon. Suosituksesta huolimatta yhteiskuntatieteellistenkin tutkimusaineistojen kohtalosta ei tavallisesti sovita ennen eikä jälkeen tutkimuksen.

Sähköisiin tutkimusaineistoihin liittyvät omistus-, hallinta- ja tekijänoikeudet sekä tietosuojakysymykset vaikuttavat merkittävästi tutkimusdatan avoimuuteen. Aihepiiri kaipaa kiipeästi sekä siihen kohdistuvan tutkimuksen että alaa jäsentävän koulutuksen vahvistamista.

## 7.4 OECD:n datasuosituksen periaatelista

OECD:n datasuosituksen periaatelista on yleisluonteinen. Yksittäisten tieteenalojen aineistojen ominaispiirteet ja avoimuuden lisääminen sisältävät niin paljon erityiskysymyksiä, ettei niitä ole yritetty kirjoittaa auki listattujen periaatteiden alle.

Suosituksen kansallista toimeenpanoa koskevassa jatkokeskustelussa on tärkeää pohtia itse päätavoitetta avoimuuden lisäämiseksi ja keinoja sen saavuttamiseksi. Jotkin päämäärät ja keinot edistävät useiden avoimuutta tukevien periaatteiden toteutumista, toiset taas vain joitakin yksittäisiä periaatteita. Periaatteiden limittyminen ei ole ongelma, vaan välttämätön edellytys tehokkaalle avoimuuspolitiikalle.

Datasuosituksessa listataan yhteensä kolmetoista julkisrahoitteisen tutkimusdatan avoimuutta tukevaa periaatetta (ks. liite 1). Englannin termit on käännetty siten, että vastintermeissä on otettu huomioon periaatteen soveltamisesta mainitut näkökohdat.

Suositus listaa seuraavat tutkimusdatan saatavuutta tukevat periaatteet:

- A Avoimuus (Openness)
- B Joustavuus (Flexibility)
- C Läpinäkyvyys (Transparency)
- D Lainmukaisuus (Legal Conformity)
- E Immateriaalisten oikeuksien suojaaminen ja kunnioittaminen (Protection of intellectual property)
- F Muodolliset vastuut (Formal responsibility)
- G Professionalismi (Professionalism)
- H Yhteensopivuus (Interoperability)
- I Laadunvarmistus (Quality)

- J Tietoturvallisuus ja tietosuojat (Security)
- K Tehokkuus (Efficiency)
- L Tilivelvollisuus/vastuullisuus (Accountability)
- M Pitkäaikainen säilyvyys ja käytettävyys (Sustainability)

*Avoimuudella* datasuositus viittaa tutkimusdatan tasavertaiseen ja helppoon käytettävyyteen. Avoimuuden takaa ajantasaista tietoa tarjoava verkkoympäristö. Internet mahdollistaa kustannustehokkaan ja edullisen tutkimusaineistojen jakelun tiedeyhteisölle.

*Joustavuudella* suositus viittaa etenkin vaihteleviin toimintaympäristöihin, joissa eri maat pyrkivät edistämään julkisrahoitteisten sähköisten tutkimusaineistojen avointa saatavuutta. Lainsäädäntö ja muut säännökset poikkeavat eri maissa niin paljon toisistaan, että datasuosituksen toimeenpanossa on oltava runsaasti joustonvaraa kansallisesti sovittaviin ratkaisuihin. Joustavuus liittyy myös alakohtaisten ratkaisujen tarpeeseen.

*Läpinäkyvyyden* periaatteella suosituksessa tarkoitetaan puolestaan sitä, että tutkimusdata tulee pyrkiä dokumentoimaan ja asettamaan esille tietokantoihin siten, että se on riittävän vaihtomasti kansainvälisen tiedeyhteisön tiedossa ja käytettävissä. Läpinäkyvyyttä tukevat toimet jäsennetään suosituksessa neljään osatekijään:

- 1) Tiedontuottajaorganisaatioiden eli tutkimusdataa kokoavien tahojen tulee tarjota erityisesti Internetissä riittävästi tietoa aineistoistaan ja niiden käyttöehdoista.
- 2) Tutkimusorganisaatioiden ja kansallisten tutkimusrahoittajien tulee viestiä aktiivisesti tutkimusaineistojen käyttäjille harjoittamastaan datapolitiikasta eli ohjeista ja säännöistä, jotka edistävät tutkimusaineistojen avoimuutta. (Jos politiikkoja ei ole, ne tulee muodostaa. *Lisäys SB*)
- 3) Tutkimusyhteisöjen tulee aktiivisesti luoda ja hyödyntää standardeja mahdollisimman käyttökelpoisten tutkimusdatojen kokoamiseksi ja niiden dokumentoimiseksi.
- 4) Tutkimusdatan keruu- ja arkistointiorganisaatioiden tulisi tehdä yhteistyötä datan hallintajärjestelmien ja saatavuusehtojen kehittämiseksi.

*Lainmukaisuudella* suositus viittaa sekä lainmukaiseen toimintaan yleensä että aineistoihin liittyvien oikeuksien ja intressien asianmukaiseen huomioon ottamiseen. Datasuosituksessa luetellaan poikkeukset, joiden kohdalla julkisrahoitteisen tutkimusdatan avoimuudesta voi olla perustelua tinkiä. Näitä ovat muiden muassa kansallinen turvallisuus, yksityisyyden suoja ja tietosuojakäytännöt, liikesalaisuudet ja eräät tekijänoikeuksiin liittyvät seikat. *Immateriaalisten oikeuksien suojaaminen ja kunnioittaminen* ovat periaatelistalla omana kohtana.

*Muodollisten vastuiden määrittely* liittyy laajasti pyrkimykseen sopia riittävän tarkasti siitä, millaisia oikeuksia ja velvollisuuksia eri tahoilla on tutkimusdataan. Liian usein avoimen saatavuuden kannalta olennaiset ehdot jäävät täyttymättä siksi, että aineistoa kerätessä tai sitä käytettäessä vastuista ja oikeuksista on sovittu vain epävirallisesti. Kirjalliset sopimukset sekä avoimuutta tukevien kirjallisten sääntöjen ja ohjeiden laatiminen edesauttavat osaltaan datan

saatavuutta. Myös tutkimusdatan pitkäaikais säilytys ja jatkokäyttöön toimittaminen sekä näihin liittyvät vastuut tulee suunnitella ja toteuttaa riittävän tarkasti.

Sähköisten tutkimusaineistojen elinkaareen liittyvällä *professionalismin* periaatteella OECD:n datasuositus viittaa tutkimustoiminnan ammattimaisuuden ja erityisesti ammatti-etiikan toteutumiseen tutkimusprosessin ja datan elinkaaren eri vaiheissa. Periaate liittyy tutkimusaineistojen kokoamista ja käyttöä koskeviin tutkimuseettisiin kysymyksiin ja tiedeyhteisön jäsenten väliseen luottamukseen aineistojen professionaalisessa käytössä. Ensikäyttäjien tulee voida luottaa jatkokäyttäjiin, ja päinvastoin. Professionalismin periaate korostaa tutkimustoiminnan yleistä suunnitelmallisuutta sekä tutkimusdatan elinkaaren hallinnan ammattimaisuutta (erityisosaamista) avoimuuden tukena.

Professionalismiin liittyvät laajasti myös *yhteensopivuuden* periaatetta tukevat seikat. Tietotekniset formaatit ja standardit tulee kehittää mahdollisimman laajasti yhteensopiviksi. Jokseenkin samaa asiaa ajetaan erilaisilla tutkimustoimintaan ja aineistojen pitkään elinkaareen liittyvillä *laatusstandardeilla*. *Laatu* on mainittu datasuosituksessa itsenäisenä, kokoavana periaatteena.

Periaatelistan viisi viimeistä kohtaa ovat nekin luonteeltaan tiedeyhteisön tunnistamia, hyvien aineistonkeruun ja -käsittelyn käytäntöjen ytimeen kuuluvia asioita. *Tietoturvallisuuden ja tietosuojaan* hallinta on välttämätön edellytys sähköisen tutkimusdatan pitkälle elinkaarelle ja asianmukaiselle käytölle. Tutkimusdataa tulee käyttää ja säilyttää kaikissa olosuhteissa turvalisesti ja luottamuksellisesti.

Tietosuoja on kuitenkin pyrittävä toteuttamaan rinnan yleisen avoimuusperiaatteen kanssa, koska vain tällä tavoin tutkimukseen suunnatut voimavarat voidaan hyödyntää riittävän *tehokkaasti*. Tutkimusten aineistotietojen tulee olla kattavasti ja avoimesti tiedeyhteisön saatavilla, jotta tiedon keruussa ja analysoinnissa vältetään päällekkäisyyksiltä. Päällekkäisyyksien ehkäisyä tukevat erityisesti myös datan dokumentointiin, säilyttämiseen ja jakeluun erikoistuneet organisaatiot, joiden tietopalvelut on suunniteltu tukemaan tätäkin periaatetta.

Tehokkuusperiaatetta edistää eri tavoin *tilivelvollisuuden tai -vastuullisuuden* periaate. Perusajatuksena on, että tutkimusdatan avoimuus lisää julkisten tutkimusinvestointien tehokkuutta ja tuottavuutta. Tämän mittaamiseksi kehoitetaan kehittämään arviointijärjestelmiä.

Sähköisen tutkimusdatan *pitkäaikainen säilyvyys ja käytettävyys* (sustainability) edellyttävät pitkäjänteistä toimintaa ja tätä varten luotujen datainfrastruktuurien kestävä rahoituspohjaa. Vain riittävän rahoituksen turvin pystytään löytämään erityyppisille tutkimusaineistoille soveltuvat pitkäjänteiset arkistointiratkaisut.

Periaatelistan lisäksi datasuosituksessa kiinnitetään kokonaisvaltaisesti huomiota *tarkoituksenmukaisten saatavuusjärjestelyjen suunnitteluun*. Kaikkien intressitahojen tulee tukea tutkimusdatan pitkää elinkaarta. Todennäköisimmin tämä toteutuu silloin, kun datan keruuseen, käyttöön ja säilytykseen liittyvät asiat suunnitellaan ja sovitaan tarpeeksi yksityiskohtaisesti jo tutkimushankkeiden alkuvaiheessa. Tämä puolestaan realisoituu todennäköisimmin silloin, kun sekä tutkimusorganisaatiot että tutkimusrahoittajat ohjaavat ja resursoivat tarkoituksenmukaisesti julkisrahoitteisen tutkimusdatan pitkää elinkaarta.

## 7.5 Miten tästä eteenpäin?

Nykytilanne ei ole hyvä, mutta monia asioita on mahdollista ja tarpeellistakin kehittää pitkällä aikajänteellä. Raportissa kuvatut näkemykset (professoreiden, yliopistojen ja muiden valtion tutkimusorganisaatioiden sekä tutkimusrahoittajien edustajien) ovat yleisellä tasolla positiivisia, mutta eivät varauksettomia. Osaltaan aiheen hankaluutta kuvastanee sekin, ettei OECD ole toistaiseksi itseään aktiivisesti tiedottanut datasuosituksesta.

### 7.5.1 Toimeenpanon hyötyjä

Aina ei ole itsestään selvää, että uuden tutkimusaineiston keruu on tarpeellista. Kokoamalla nykyistä huomattavasti kattavammin tietoa olemassa olevista aineistoresursseista tiedeyhteisö saisi vankemman tietopohjan uusien tietotarpeiden tunnistamiseen. Avoimuus vähentää turhaa tiedonkeruuta.

Datasuosituksen toimeenpanon hyödyt ovat suuremmat kuin haitat. Kaaviossa 7.1 on täsmennetty odotettavissa olevia hyötyjä. Avoimuudessa on kyse paljon muustakin kuin rahasta.

Hyödyn laatu	Hyöty / selite
Taloudellinen	Tutkimusaineistojen käytön tehostaminen
Tieteellinen	Tieteen avoimuus, tulosten ajantasaisuus, toistettavuus ja kontrolloituavuus, vuorovaikutuksen lisääntyminen datan yhteiskäytössä
Tutkimuseettinen	Mm. tutkimuskohteiden huomioon ottaminen datan keruussa ja käytössä
Oikeudellinen	Lainsäädännön huomioon ottaminen datan keruussa, käytössä ja säilyttämisessä
Demokraattinen	Tutkittavien ja tutkijoiden sekä tutkimusorganisaatioiden ja -rahoittajien aseman ja vastuiden (oikeuksien, velvollisuuksien ja vaikuttamismahdollisuuksien) täsmentäminen
Tasa-arvoinen	Tutkimusaineistojen tasavertaisten käyttömahdollisuuksien edistäminen
Yhteiskunnallinen	Yhteiskunnan tietohuollon laadun ja kustannustehokkuuden kohentuminen
Poliittinen, spesifi	OECD:n suositus velvoittaa jäsenmaita suosituksen huomioon ottamiseen

Kaavio 7.1  
OECD:n datasuosituksen toimeenpanon hyötynäkökulmia

## 7.5.2 Open Access: julkaisu verrattuna tutkimusdataan

Olenainen argumentti tutkimusdatan avoimuuden puolesta on *tutkimusympäristöjen muuttaminen sähköiseksi*. Kyse on perustavaa laatua olevasta uudelleenrakentumisesta, joka on jo muuttanut keskeisesti tieteellisten tutkimuksen käytäntöjä.

Datajulistuksessa tavoiteltu avoimuuden periaate on toteutettavissa pitkälti Internet-palvelujen avulla. Siksi myös tutkimusten aineistotietojen pitää olla yhä helpommin ja täsmällisemmin saatavilla verkosta. Tutkimusdataa on myös saatava avoimeen ja *informoituun* käyttöön turvallisesti verkon kautta. Tämä on keskeinen kysymys kehitettäessä houkuttelevia ja kilpailukykyisiä tutkimusympäristöjä.

Sähköisiin julkaisuihin painottuvat Open Access -tavoitteet eivät ole kaikilta osin suoraan sovellettavissa sähköisiin tutkimusaineistoihin. Tutkimusdatan avoimuus on tavallisesti vain tiedeyhteisölle rajattua. Immateriaalioikeuksien lisäksi tutkimusdatassa tarkasteltavana ovat usein myös tietosuojakäytännöt. Keskeiset tavoitteet julkaisujen ja datan avoimuuden taustalla ovat kuitenkin samat: tieteellisen tiedon avoin saatavuus ja käytettävyyden edistäminen tieteen kehittämistä ja tasavertaistavat tiedon käytön mahdollisuuksia. Kulttuuriset ja asenteelliset esteet avoimuuden lisäämiseksi ovat ainakin osin samanlaisia (ks. esim. Björk 2004).

Kaavioon 7.2 on pelkistetty joitakin sähköisten tutkimusjulkaisujen ja -datojen välisiä eroja. Tutkimusdatojen eri aineistotyyppistä koskee suuri joukko erityiskysymyksiä, joiden ratkaiseminen edellyttää ala- ja tilannekohtaista harkintaa.

Tutkimusjulkaisu	Tutkimusdata
Tulokset ja tieto jalostettu	Tieto jalostamatta tuloksiksi
Käyttö edellyttää perusohjelmistoja ja -laitteita ja niiden tuntemusta	Käyttö edellyttää usein erityislaitteita ja -ohjelmistoja sekä niiden tuntemusta
Itsensä selittävä	Edellyttää arkistoitamattomana lisätietoja, dokumentointia
Ei sisällä salassa pidettäviä tietoja	Voi sisältää arkaluonteisia ja salassa pidettäviä tietoja
Käyttö ei edellytä lupaa	Käyttö usein luvanvaraista
Omistus- ja tekijänoikeudet useimmiten selvillä	Omistus- ja tekijänoikeudet usein epäselviä
Tiedeyhteisön avoimessa käytössä maksuttomasti ja maksullisena	Useita avoimuuden asteita (täysin avoimesta suljettuihin)
Ymmärretään tieteelliseksi tuotokseksi (mainitaan CV:ssä)	Ei nykyisin pidetä tieteellisenä tuotoksena / meriittinä edes silloin kun data julkaistu (ei yleensä mainita CV:ssä)
Valmis muiden käytettäväksi	Käyttö edellyttää prosessointia

Kaavio 7.2

Tutkimusjulkaisu, tutkimusdata ja Open Access: pelkistetty eroavuuskaavio

### 7.5.3 Laaja-alainen yhteistyö datasuosituksen toimeenpanon tueksi

Julkisrahoitteisten sähköisten tutkimusaineistojen avoimen saatavuuden edistäminen edellyttää mielestämme Suomessa laaja-alaista ja pitkäjänteistä yhteistyötä jo yksinomaan siksi, että opetusministeriön ohella OECD-toiminnassa ovat mukana monet muut viranomaiset. Yhteistyön tulisi tähdätä kansallisiin suosituksiin ja toimintamalleihin tutkimusdatan saatavuuden ja avoimen käytön edistämiseksi.

Kansallisten ohjeiden, suositusten ja datapolitiikkojen kehittämisessä pitää ottaa huomioon olemassa olevat kansainväliset mallit ja toimintastandardit. Yhteistyöhön tulisi saada mukaan julkisrahoitteista tutkimusta harjoittavat ja tutkimusdataa keräävät tutkimusorganisaatiot ja keskeiset tutkimusrahoittajat ja tiedeorganisaatiot eri aloilta. Opetusministeriön näkökulmasta keskeisiä toimijoita asiassa ovat muiden muassa Suomen Akatemia, yliopistot ja rehtorien neuvosto sekä tutkimuseettinen neuvottelukunta ja tiedonjulkistamisen neuvottelukunta.

Sähköisten aineistojen avoimuuden edistäminen voidaan toteuttaa esimerkiksi laaja-alaisena keskusteluforumina. Ainakin seuraavat seikat voisivat olla foorumin asialistalla:

1. OECD:n datasuosituksen toimeenpanoa koskeva kansallinen keskustelu ja yleisten toimintamallien hahmottaminen
2. Tutkimusaineistojen elinkaarta tukevien tutkimusaineistojen keruun suunnittelu- ja sopusäkäytäntöjen kehittäminen
3. Tutkimusaineistoihin liittyvien oikeuksien ja vastuiden selkiyttäminen
4. Tutkimusaineistojen elinkaaren hallinnan kehittäminen koulutuksen avulla
5. Tutkimusrahoittajien ja aineistonkerääjien rohkaiseminen datapolitiikkojen (data policy) luontiin
6. Keskustelu ja esitykset tutkimusaineistojen jatkokäyttöä edistävien suositusten ja rahoituskehtojen käyttöönoton mahdollisuuksista

Tutkimusdatan avoimuutta tukevien pelisääntöjen, suositusten ja toimintatapojen luonti ja vaikiinnuttaminen olisi Suomessa iso tiede- ja tutkimuspoliittinen harppaus, joka kehittäisi merkittävästi kansallisen tutkimusympäristömme laatua.

Periaatteelliset linjaukset ja strategisten tavoitteiden asettaminen ovat koko tiedeyhteisön ja sen taustaorganisaatioiden yhteinen asia. OECD-suosituksen kansallinen toimeenpano vaatii siksi laaja-alaista strategia- ja suunnittelutyötä. Käytännön ratkaisut toimintaympäristöjen, aineistojen säilytyksen, teknisten kysymysten ja lainsäädännöllisten seikkojen osalta vaativat tieteenalakohtaista keskustelua ja selvitystyötä.

Mahdollisuuksia OECD-suosituksen kansalliseen toteuttamiseen on monia. Alkuvaiheessa suunnittelua varten voitaisiin koota eri intressitahoja edustava valmistelutyöryhmä, joka työstäisi hankkeelle tavoite- ja toimintasuunnitelman. Työn aloittamisen tueksi on jo nyt käytössä muissa maissa laadittuja, aivan hiljattain ilmestyneitä taustaraportteja (esim. RIN 2007).

Joihinkin asioihin voidaan vaikuttaa melko nopeasti. Tiedeyhteisön vallitsevia toimintatapoja ja kulttuurisia käytäntöjä muuttavat keinot ovat pitkälti tutkimusrahoittajien käsissä. Esimerkiksi rahoitushakemuksissa edellytettävä tutkimusdataa ja sen pitkäaikaissäilyttämisestä koskeva suunnitelma ohjaisi käytännössä huolellisempaan datan keruuseen, käsittelyyn ja säilyttämiseen. Nyt vallitsevassa tilanteessa arvokasta tutkimusdataa vanhenee koko ajan tietotek-

nisesti käyttökelvottomaksi. Aineistojen elinkaaren suunnitelman edellyttäminen osaksi tutkimussuunnitelmia ei maksaisi mitään.

Lopuksi haluamme korostaa, että raportti keskittyi ihmistieteisiin. Professorikyselyn tuloksissa yhdeksi keinoksi datan avoimuuden toteutumisessa mainittiin asiaa koskevat sisällölliset ja tekniset ohjeet ja koulutus. Yksi mahdollisuus olisi laatia KY-alojen pilottina sähköisen datan käsittelyyn yksityiskohtaisesti opastava käsikirja, joka konkreettisesti tukisi tutkimusdatan elinkaaren hallintaa. Esimerkiksi Isossa-Britanniassa UK Data Archive (UKDA) laati aluksi yhteiskuntatieteisiin perusteellisen datan käsittelyoppaan, jota hyödyntäen UKDA on nyttemmin ryhtynyt laatimaan vastaavia myös muille tieteenaloille. Yhteiskuntatieteellinen tietoaarkisto olisi Suomessa pilotin luonteva toteuttaja.

## LIITE 1.

## OECD:n Pariisin julistus 2004 ja Datasuositus 2007 (OECD Principles and Guidelines for Access to Research Data from Public Funding)

### DECLARATION ON ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

adopted on 30 January 2004 in Paris

The governments (1) of Australia, Austria, Belgium, Canada, China, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Russian Federation, the Slovak Republic, the Republic of South Africa, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States

Recognising that an optimum international exchange of data, information and knowledge contributes decisively to the advancement of scientific research and innovation;

Recognising that open access to, and unrestricted use of, data promotes scientific progress and facilitates the training of researchers;

Recognising that open access will maximise the value derived from public investments in data collection efforts;

Recognising that the substantial increase in computing capacity enables vast quantities of digital research data from public funding to be put to use for multiple research purposes by many research institutes of the global science system, thereby substantially increasing the scope and scale of research;

Recognising the substantial benefits that science, the economy and society at large could gain from the opportunities that expanded use of digital data resources have to offer, and recognising the risk that undue restrictions on access to and use of research data from public funding could diminish the quality and efficiency of scientific research and innovation;

Recognising that optimum availability of research data from public funding for developing countries will enhance their participation in the global science system, thereby contributing to their social and economic development;



Recognising that the disclosure of research data from public funding may be constrained by domestic law on national security, the protection of privacy of citizens and the protection of intellectual property rights and trade secrets that may require additional safeguards;

Recognising that on some of the aspects of the accessibility of research data from public funding, additional measures have been taken or will be introduced in OECD countries and that disparities in national regulations could hamper the optimum use of publicly funded data on the national and international scales;

Considering the beneficial impact of the establishment of OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980, 1985 and 1998) and the OECD Guidelines for the Security of Information Systems and Networks (1992, 1997 and 2002) on international policies for access to digital data;

*DECLARE THEIR COMMITMENT TO:*

Work towards the establishment of access regimes for digital research data from public funding in accordance with the following objectives and principles:

**Openness:** balancing the interests of open access to data to increase the quality and efficiency of research and innovation with the need for restriction of access in some instances to protect social, scientific and economic interests.

**Transparency:** making information on data-producing organisations, documentation on the data they produce and specifications of conditions attached to the use of these data, available and accessible internationally.

**Legal conformity:** paying due attention, in the design of access regimes for digital research data, to national legal requirements concerning national security, privacy and trade secrets.

**Formal responsibility:** promoting explicit, formal institutional rules on the responsibilities of the various parties involved in data-related activities pertaining to authorship, producer credits, ownership, usage restrictions, financial arrangements, ethical rules, licensing terms, and liability.

**Professionalism:** building institutional rules for the management of digital research data based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.

**Protection of intellectual property:** describing ways to obtain open access under the different legal regimes of copyright or other intellectual property law applicable to databases as well as trade secrets.

**Interoperability:** paying due attention to the relevant international standard requirements for use in multiple ways, in co-operation with other international organisations.

**Quality and security:** describing good practices for methods, techniques and instruments employed in the collection, dissemination and accessible archiving of data to enable quality control by peer review and other means of safeguarding authenticity, originality, integrity, security and establishing liability.

**Efficiency:** promoting further cost effectiveness within the global science system by describing good practices in data management and specialised support services.

**Accountability:** evaluating the performance of data access regimes to maximise the support for open access among the scientific community and society at large.

Seek transparency in regulations and policies related to information, computer and communications services affecting international flows of data for research, and reducing unnecessary barriers to the international exchange of these data;

Take the necessary steps to strengthen existing instruments and - where appropriate - create within the framework of international and national law, new mechanisms and practices supporting international collaboration in access to digital research data;

Support OECD initiatives to promote the development and harmonisation of approaches by governments adhering to this Declaration aimed at maximising the accessibility of digital research data;

Consider the possible implications for other countries, including developing countries and economies in transition, when dealing with issues of access to digital research data.

#### *INVITE THE OECD:*

To develop a set of OECD guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data from public funding, to be endorsed by the OECD Council at a later stage.

---

(1) Including the European Community

## LIITE 2.

## Kuvaus selvityshankkeen aineistonkeruusta vuonna 2006

Pariisin julistuksen jälkeen OECD:n datasuosituksen valmistelussa toteutettiin kaksi kansallista tutkimus- ja tiedeorganisaatioiden konsultointikierrosta. Ensimmäiselle osallistui vuonna 2005 toistakymmentä valtion tutkimuslaitosta (ks. liite 3) ja suppealle toiselle kierrokselle vuonna 2006 osallistuivat VTT, KTL ja FSD.

Näiden konsultointien tuloksia on käytetty tässä raportissa tausta-aineistona. Eniten hyödynnetty selvitysaineisto koottiin vuoden 2006 marras-joulukuussa verkkokyselynä kulttuurin ja yhteiskunnan tutkimuksen alan yliopistoprofessoreilta. Samassa yhteydessä toteutettiin pienimuotoiset rinnakkaiskyselyt yliopistojen ja tutkimusrahoittajien alustavista kannoista OECD:n datasuositusluonnokseen. Tiedonkeruut olivat sisällöllisesti yhteneviä.

Professorikysely suunnattiin Yhteiskuntatieteellisen tietoarkiston (FSD) toimintapiiriin sijoitettaville tieteenaloille: yhteiskuntatieteisiin, kasvatustieteisiin ja humanistisiin tieteisiin. Tieteenalat selviävät luvussa 3.1.2 esitetystä taulukosta 1. Professorikysely pyrittiin toteuttamaan kokonaistutkimuksena siten, että kaikkien mainittujen alojen yliopistolaitosten verkkosivuilta koottiin keruuhetkellä professoreina toimivien sähköpostiosoitteet. Verkkokysely toteutettiin 9.11.2006 automatisoituna yksilölähetyksenä siten, että palvelin lähetti jokaiselle nimetylle vastaanottajalle seuraavan henkilökohtaisen viestin:

”Arvoisa NN,

OECD:n tiede- ja teknologiaministerit hyväksyivät tammikuussa 2004 pidetyssä ministerikokouksessa julistuksen, jolla tuetaan julkisrahoitteisten sähköisten tutkimusaineistojen käytön avoimuutta. Julkilausemalla ja sen pohjalta tehdyllä suosituksella pyritään lisäämään sähköisten tutkimusaineistojen avointa ja kustannustehokasta käyttöä.

OECD:n suosituksessa sähköisillä aineistoilla ei tarkoiteta julkaisuja tai raportteja, vaan elektronisia tutkimusaineistoja, joita tutkimuksessa analysoidaan. Niistä voidaan puhua esimerkiksi pohjaaineistoina, perusaineistoina tai havaintoaineistoina. Ne voivat olla sähköisiä numeerisia aineistoja, havaintomatriiseja, tekstiaineistoja, kuva-aineistoja tai digitaalisia audio- tai audiovisuaalisia aineistoja.

Opetusministeriön toimeksiannosta Yhteiskuntatieteellinen tietoarkisto kartoittaa, miten tutkimusorganisaatioissa on nyt järjestetty sähköisten tutkimusaineistojen säilyttäminen. Lisäksi selvitetään, missä määrin ja millä tavoin tutkimusaineistot ovat tiedeyhteisön jatkokäytössä alkuperäisten tutkimusten valmistuttua. Oheinen kysely on osa selvitystä ja se on suunnattu humanististen, yhteiskuntatieteellisten ja käyttäytymistieteellisten alojen professoreille.

Opetusministeriö pitää erittäin tärkeänä, että mahdollisimman moni vastaa kyselyyn kattavan aineiston saamiseksi. Toivomme, että vastaatte oheiseen verkkokyselyyn

<https://elomake.uta.fi/lomakkeet/1475/lomake.html>

kahden viikon kuluessa, keskiviikkoon 22.11.2006 mennessä. Kyselyn tulokset julkaistaan alkutalvella 2007 Yhteiskuntatieteellisen tietoarkiston julkaisusarjassa. Tulokset raportoidaan niin, että yksittäiset vastaajat eivät ole niistä tunnistettavissa. Vastauksista koottava aineisto arkistoidaan Yhteiskuntatieteelliseen tietoarkistoon.

Suosittelimme, että ennen kysymyksiin vastaamista tutustutte OECD:n suositukseen

[http://www.fsd.uta.fi/julkaisut/open\\_access/OECDsuositus.pdf](http://www.fsd.uta.fi/julkaisut/open_access/OECDsuositus.pdf)

ja siitä tiivistettyyn suomenkieliseen muistioon

[http://www.fsd.uta.fi/julkaisut/open\\_access/OECDtiivistelma.pdf](http://www.fsd.uta.fi/julkaisut/open_access/OECDtiivistelma.pdf)

Molempiin on linkki myös verkkokyselyn yhteydessä. Osallistumisenne on tärkeää, jotta saamme kattavasti tietoa suosituksen toimeenpanon käytännön edellytyksistä.

Pahoittelemme, että aikataulumme ei mahdollistanut lomakkeen ja linkkien kääntämistä ruotsiksi.

Yhteistyöstä etukäteen kiittäen,

Sami Borg

Yhteiskuntatieteellisen tietoarkiston johtaja

[sami.borg@uta.fi](mailto:sami.borg@uta.fi) ”

Muistutusviesti vastaamiseen lähetettiin marraskuun 28. päivä, jolloin viimeiseksi vastausajankohdaksi määriteltiin joulukuun 5. päivä. Tuohon mennessä yhteensä 150 tavoitelluista 528 professorista vastasi kyselyyn. Vastausprosentti jäi verraten matalaksi (28,4 %).

Sähköpostilla ja verkkokyselynä toteutetulla tiedonkeruulla oli etunsa ja pulmansa. Tavanomaisella postilla lähetettyyn kirjeeseen olisi saatettu saada enemmän vastauksia mutta tiedonkeruukustannukset olisivat olleet paljon suuremmat kuin verkkokyselyssä.

Verkkokyselyllä myös oheismateriaali saatiin välitetyksi kätevästi vastaajille. Datasuositusluonnos oli pitkä ja kunnollinen perehtyminen siihen vei kieltämättä vastaajilta aikaa. Siksi verkkokyselyyn oheistettiin linkki suomenkieliseen tiivistelmään luonnoksesta. Itse verkkokysely ei ollut kovin pitkä.

Muutamana kymmenen spontaanin sähköpostipalautteen perusteella syitä vastaamattomuuteen oli useita. Kyselyn ajoittuminen lukukauden verraten kiireiseen aikaan on yksi taustatekijä. Moni vastaaja ei myöskään katsonut laitoksellaan olevan lainkaan tai juuri ollenkaan sähköisiä tutkimusaineistoja, tai kertoi, ettei hän itse käytä tällaisia aineistoja tutkimuksessaan tai opetuksessaan. Moni ei siis katsonut asian kuuluvan itselleen.

Myös datasuosituksen tosiasiallinen keskeneräisyys saattoi vähentää vastaajien määrää. Kyse ei ollut ”virallisesta” lausuntokierroksesta vaan kommenttipyyntöistä luonnokseen. Jotakin vastaajia lienee vaivannut myös se, että osa kysymyksistä koski laitoksen yleistä tilannetta. Varsinkin vähän aikaa viroissaan toimineet eivät ehkä katsoneet tietävänsä asioista riittävästi vastataksaan kyselyyn.

Voitaneen olettaa, että vastanneet 150 professoria edustavat kuitenkin suuntaa antavasti sitä tutkijajoukkoa ja niiden KY-alan yliopistolaitosten tilannetta, joille sähköisten tutkimusaineistojen säilyttämisen kysymykset ovat ajankohtaisia. Voi tietenkin olla, että vastanneet suh-

tautuvat tutkimusaineistojen avoimuuteen myönteisemmin kuin vastaamatta jättäneet, mutta välttämättä asia ei ole näin.

Yhteensä 20 yliopistolle ja noin 40 tutkimusrahoittajalle suunnatut postikyselyt toteutettiin niin ikään marras-joulukuussa 2006. Konsultoitaviksi tutkimusrahoittajiksi valittiin myös yksityisiä säätiöitä, koska niitä katsottiin tarpeelliseksi ainakin informoida tulevasta datasuosituksesta tämän selvitystyön yhteydessä. Yliopistoissa kirje osoitettiin tutkimuksesta vastaavalle vararehtorille, jota pyydettiin vastaamaan lomakkeeseen itse tai välittämään se vastattavaksi soveltuvaksi katsomalleen henkilölle yliopistossa.

Rahoittajatahoista yhdeksän vastasi kyselyyn, jolloin osallistumisaktiivisuus osoittautui samansuuruiseksi kuin KY-alan professorikyselyssä. Yliopistoista lomakkeen palautti kahdeksan eli 40 prosenttia. Vastajamäärien jäätyä näin pieniksi yliopisto- ja rahoittajakyselyn tuloksiin viitataan selvityksessä vain niukasti. Kaikkiaan näyttää ilmeiseltä, että monilla tahoilla ei ollut halukkuutta kommentoida keskeneräistä ja moniaineksista asiaa vielä tässä vaiheessa. Yliopistojen ja tutkimusrahoittajien konsultointiin ja integrointiin mukaan tutkimusdatan avoimuuden edistämiseen on kuitenkin hyvät mahdollisuudet datasuosituksen toimeenpanon tulevissa vaiheissa.

## LIITE 3.

## Tietoja valtion tutkimuslaitosten alustavasta suhtautumisesta OECD:n Pariisin julistukseen ja tutkimusaineistojen avoimesta saatavuudesta 2005

Keväällä 2005 Yhteiskuntatieteellinen tietoaarkisto tiedusteli lähes kolmeltakymmeneltä valtion tutkimusorganisaatiolta niiden *alustavaa* suhtautumista OECD:n Pariisin julistukseen. Tiedustelu toteutettiin verkkokyselynä ja sen tuloksia hyödynnettiin silloisessa OECD:n data-suosituksen kansainvälisessä valmistelutyössä.

Koska kyselyn vastaukset kuvannevat melko hyvin myös nykytilannetta, on silloiset alustavat kannanotot tarkoituksenmukaista liittää tähän selvitysraporttiin. Tiedot *eivät* ole tutkimusorganisaatioiden virallisia kannanottoja asiaan. Kyselyyn vastasi neljätoista tutkimusorganisaatiota. Kukin organisaatio valitsi itse vastaajansa.

### Kyselylomakkeen kysymykset:

#### Organisaation linja tutkimusaineistojen Open Access -politiikassa

Onko organisaatiollanne virallisesti päätettyä toimintapolitiikkaa tai toimintalinjaa, joka tukee julkisrahoitteisten tutkimusaineistojen mahdollisimman avointa jatkokäyttöä? Jos on, millainen se on sisällöltään? Jos ei ole, miksi ei?

Jos organisaatiollanne on tutkimusaineistojen avoimuutta tukevia päätöksiä tai toimintalinjoja, millä tavoin ne ovat syntyneet? Kuvaile tiiviisti.

Mille seuraavista aloista organisaationne keräävät tutkimusaineistot sijoittuvat?

Merkitkää kaikki kyseeseen tulevat alat (sarake A).

Arvioikaa lisäksi sitä (sarake B), minkä alojen julkisrahoitteisten aineistojen avointa käyttömahdollisuutta olisi mielestänne erityisesti lisättävä Suomessa (vastaajan oma mielipide, useampi kuin yksi vastaus sallittu).

	A	B
	Aineistot sijoittuvat	Tulisi lisätä
Luonnontieteet	—	—
Tekniikan ja teknologioiden kehittäminen	—	—
Terveystieteet	—	—
Yhteiskuntatieteet	—	—
Humanistiset tieteet	—	—
Jokin muu, määrittele mikä	—	—

## Mielipiteet OECD:n julkilausumasta

Ovatko OECD:n julkilausumassa ja sen perusteluissa luonnostellut periaatteet toteuttamiskelpoisia organisaatiossanne? Jos eivät ole, mitä ovat keskeisimmät esteet?

Voivatko luonnostellut periaatteet toimia OECD:n sovellusohjeiden (OECD Guidelines) pohjana?

Mitä lisänäkökohtia asian jatkovalmistelussa tulisi ottaa huomioon?

Onko mielessänne vielä joitakin muita asiaan liittyviä kommentteja?

---

## Kyselyn tuloksia

### organisaatio: Tilastokeskus

vastaaja: hallintojohtaja

Virallinen toimintalinja:

Tilastokeskus ei kerää tietoaineistoja varsinaisesti tutkimustarkoituksiin kuin joissakin tapauksissa maksullisena toimintana ulkopuolisten toimeksiannosta. Aineistot kerätään pääsääntöisesti tilastotarkoituksiin, jolloin aineiston jatkokäyttö määräytyy tilastointia koskevan kansallisen ja EU-lainsäädännön mukaan. Tilastolaki edellyttää käyttöluvan myöntämistä tilastoaineistoihin jokaisessa yksittäistapauksessa erikseen. Tietojen luovuttamiseksi sovelletaan JulkL:n 28 §:n mukaisia ehtoja. Tilastokeskuksen käytännön toimintalinjat on määritelty ohjeessa käyttöluvien myöntämisestä Tilastokeskuksen yksikkötason aineistoon. Julkisuuslain periaatteisiin viitaten ohjeessa todetaan:

”Lupaa harkittaessa on myös huolehdittava siitä, että tieteellisen tutkimuksen vapaus turvataan. Käytännössä tämä tarkoittaa sitä, että harkinnassa lähtökohtana on käyttöluvan myöntäminen. Mahdolliset kielteiset päätökset tulee selkeästi perustella”.

Toteuttamiskelpoisuus:

Kuten edellä on todettu Tilastokeskuksen aineistot on pääosin kerätty tilastotarkoituksiin ja niiden käyttö määräytyy tilastolainsäädännön ja tilastoalalla käytännössä sovellettavien kansainvälisten periaatteiden mukaan. Niiltä osin kuin periaatteet eivät ole ristiriidassa tilastoalan lainsäädännön ja periaatteiden kanssa, ne ovat sovellettavissa.

Sovellusohjeiden pohjana: Niiltä osin kuin periaatteet rajautuvat tutkimustarkoituksiin alunperinkin kerättyihin aineistoihin, varmaan hyvinkin voivat. Jos kysymys on alunperin muihin tarkoituksiin kerätyistä tiedoista (mm. tilastotarkoitus), ensisijainen käyttötarkoitus määrittelee viime kädessä myös käsittely- ja käyttöperiaatteet. Periaatteissa on kyllä viitattu rajoituksiin näiden aineistojen osalta, mutta eron voisi tuoda esille vielä selkeämminkin väärin käsitysten välttämiseksi.

---

## organisaatio: Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus, Stakes

vastaaja: lakimies

Virallinen toimintalinja: Stakesilla ei ole virallisesti päätettyä toimintapolitiikkaa.

Stakesin tutkimusta luonnehtivia piirteitä on, että tutkimus on yksi Stakesille sitä koskevasa laissa (1073/1992) asetetuista tehtävistä (2 §) ja että merkittävä osa tutkimuksesta perustuu Stakesin näiden tehtävien suorittamiseksi lain nojalla ylläpitämien terveydenhuollon valtakunnallisten henkilörekisterien (556/1989) tietoihin sekä Stakesin tilastotoimesta annetun lain (409/2001) nojalla kerättäviin tietoihin. Kaikki Stakesin edellä mainittujen lakien nojalla keräämät sosiaali- ja terveydenhuollon tiedot ovat salassa pidettäviä. Tutkimusaineistoja kerätään jossakin määrin myös kyselyillä ja haastatteluilla tutkittavien suostumuksella. Erityisesti rekisteritutkimukselle tyypillinen piirre on, että sitä tehdään Stakeslain 4 §:ään perustuvien yhteistyösopimusten perusteella. Ulkopuolisille tutkijoille Stakes luovuttaa rekistereihin kerättyjä tietoja po. erityislakien sekä julkisuuslain nojalla joko Stakesin tai sosiaali- ja terveysministeriön viranomaisluvalla. Tunnisteellisena tietoa voidaan luovuttaa vain terveydenhuollon tiedoista, sosiaalihuollon tietoja luovutetaan vain tunnisteettomana.

Omien päättyneiden tutkimusten aineistojen osalta Stakes voi antaa julkisuuslain 28 §:n mukaisena viranomaisena luvan arkistosuunnitelmansa nojalla arkistoidun tutkimusaineiston käyttöön. Lupa myönnetään mikäli tieteellisen tutkimuksen yms. lainsäädännöstä nousevat kriteerit täyttyvät. Käytännössä kielteisiä päätöksiä ei tehdä. Keskenäisten tutkimusaineistojen osalta päätös voi olla kielteinen tai myönteinen. Julkisuuslaki suojaa tieteellisen tutkimuksen suunnitelmaa ja perusaineistoa riippumatta siitä, sisältääkö se salassa pidettävää tietoa (24 §1. mom. 21-kohta) siinä tapauksessa että tiedon antamisesta ei aiheudu haittaa tutkimuksen suorittamiselle, sen hyödyntämiselle tai asianmukaiselle arvioimiselle, tutkijalle tai tutkimuksen toimeksiantajalle. Tutkittavien suostumuksella saatuja aineistoja ei voida luovuttaa edelleen ulkopuolisille (JulkL 24 § 1.mom. 16-kohta), ellei lupaa siihen ole pyydetty aineistoa kerätessä, mikä ei ole käytäntönä. Yhteistyösopimuksilla tehtävien tutkimusten aineistoja ei voida luovuttaa edelleen muutoin kuin kaikkien sopijapuolten antamalla viranomaisluvalla. Käytännössä kysymys on harvoin ollut esillä.

Toteuttamiskelpoisuus:

Keskeinen este yleiselle avoimuudelle perustuu sekä yleisesti että erityisesti Stakesin sosiaali- ja terveydenhuoltoa koskevien tietojen osalta tietojen luonteeseen arkaluontoisina ja salassa pidettävinä. Periaatteissa esitetty läpinäkyvyys toteutuu nykyisten käytäntöjen mukaan henkilötietolain edellyttämän rekisteriselosteen kautta, joka on julkinen asiakirja ja joka laaditaan kaikista henkilötietoja sisältävistä tutkimusaineistoista. Stakesin lakisääteisten rekisterien samoin kuin päättyneiden tutkimusten tutkimusaineistojen hyödyntäminen on jo nykyisellään mahdollista viranomaisluvalla. Käytäntöjen muuttamista ei salassa pidettävien tietojen tietosuoja vaarantamatta ole mahdollista tehdä eikä voida pitää perusteltuna. Lainsäädäntö suojaa keskenäisten tutkimusten tutkimusidea ja tutkimuksen perusaineistoa ja sitä vastaava suoja on keskenäisten tutkimusaineistojen osalta perusteltu.



Sekä kansallinen henkilötietolaki että myös EY:n tietosuojadirektiivi lähtevät siitä, että henkilötietoja voidaan käyttää vain siihen tarkoitukseen, jota varten ne on kerätty. Poikkeuksena on tieteellinen tutkimus, mutta tällöin edellytetään, että sitä koskevat kriteerit tarkistetaan kussakin tapauksessa erikseen myönnettävän luvan yhteydessä.

Kysymys tunnisteettomien rekisteritietojen ja tutkimusaineistojen nykyistä laajemmasta käytöstä tutkimuksessa liittyy siihen, että aineistoa kerätään verovaroin. Tietojen luovutuksen perusteena on valtion maksuperustelaki (150/1992) sekä esimerkiksi Stakesin osalta STM:n päätös Stakesin maksullisista suoritteista (1628/1995). Tietojen luovutuksia tulisi maksukysymyksen osalta tätä taustaa vasten arvioida kansallisena kysymyksenä.

Sovellusohjeiden pohjana: Periaatteet ovat sinänsä kannatettavia, mutta eivät sovellu salassa pidettävien tietojen kuten sosiaali- ja terveydenhuollon tunnisteellisten tutkimusaineistojen käyttöön

Lisänäkökohtia: Periaatteiden ei tulisi johtaa tulokseen, jossa keskeiset tutkimuksissa käytettäviä tietoja tuottavat viranomaiset joutuvat käytännössä vastaamaan myös esimerkiksi yksityisen sektorin aineistokuluista.

### **organisaatio: Teknologian kehittämiskeskus**

vastaaja: johtaja

Virallinen toimintalinja: Tekes rahoittaa tutkimusta, mutta ei tilaustutkimustyyppisesti vaan osarahoitusperiaatteella, eikä siten kerää rahoitettujen tutkimusten tutkimusaineistoja itselleen. Tutkimusaineistot jäävät niiden muodostajille käytettäväksi kulloinkin sovitulla tavalla periaatteena on rahoitettujen hankkeiden tutkimustulosten julkisuus kun puhutaan yliopistojen ja tutkimuslaitosten tekemiä hankkeista, yrityshankkeiden tutkimusaineistot puolestaan ovat luottamuksellisia ja yritysten omaisuutta. Tekes julkaisee tutkimusaineistoista tehtyjä raportteja, etenkin teknologiaohjelmien raportteja ja loppuraportteja, mutta nämä eivät kuulune kyselyynne piiriin.

Toteuttamiskelpoisuus:

Tutkimusaineistojen hankkiminen Tekesiin on vain harvoin eteen tuleva tilanne, joka koskee lähinnä vaikuttavuusarviointia tai vastaavan tyyppistä ostopalveluhankintaa, ja silloinkin vain poikkeustapauksissa. Tietoaaineistojen julkisuuden esteenä on niihin sisältyvät liiketoimintasalaisuudet.

## organisaatio: Valtion teknillinen tutkimuskeskus VTT

vastaaja: kehityspäällikkö

Virallinen toimintalinja: VTT:llä ei ole yhtenäisesti ja virallisesti päätettyä linjaa julkisrahoitteisten tutkimusaineistojen avoimuudesta. Syy on varmasti se, että mistään ongelmista ei ole aiheutunut tarvetta tehdä yhteisiä linjauksia ja organisaatiomme toimijat ovat pystyneet normaalein käytännöin tilanteet ratkaisemaan. Periaatteessa julkisesti rahoitetun tutkimuksen tulosaineistot ovat VTT:ssä julkisia ja aineistojen käyttöön saaminen on ulkopuolisille mahdollista, tosin se voi olla mutkikasta. Teknologisen tutkimuksen tulosaineistot ovat hyvin usein sellaisessa muodossa jota ulkopuolinen ei juuri pysty hyödyntämään. Tämä tekee monissa tapauksissa avoimuuskysymyksestä teoreettisen. VTT:n kansallisesta tehtävästä suomalaisen elinkeinoelämän tukijana seuraa eräitä erikoistilanteita, joissa avoimuudella on rajoituksia. Näistä alempana tarkemmin.

Toimintalinjan synty: VTT:n sisäiset toimintatavat aineistojen avoimuudessa ovat syntyneet käytännön tilanteiden ohjaamina. On mahdollista, että tutkimusyksiköillämme on laatujärjestelmässään tästä jotain ohjelmaintoja, mutta sen yksityiskohtainen selvittäminen ei liene tässä yhteydessä tarpeellista.

Toteuttamiskelpoisuus:

Luonnostellut periaatteet ovat hyvin yleisellä tasolla toteuttamiskelpoisia VTT:ssäkin, kun otetaan huomioon niissä mainitut legitiimit rajoitukset sekä erilaiset innovaatioiden suojaamisesta ja esim. julkisten rahoittajien omista ehdoista tulevat rajoitukset. Kuitenkin periaatteiden soveltuminen teknologiseen ja luonnontieteelliseenkin tutkimukseen on monin osin vaikea hahmottaa varsinkin käytännön tilanteita ajatellen. On myös selvää, että kansainvälisen avoimuuden kehittyminen, niin kannatettavaa kuin se onkin, esim. mainittujen standardien ja Code of conduct-säännösten myötävaikutuksella vaatii edelleen sekä aikaa että panostusta.

Tutkimusaineistojen avoimuutta ajatellen sisältyy VTT:n julkisrahoitteiseenkin tutkimustointintaan tilanteita, joissa avoimuus on mahdollista toteuttaa vain viivästetysti. Kun ollaan suojaamassa jokin tutkimuksen tuloksena syntynyt innovaatio IPR-prosessin avulla, ei avoimuutta voida toteuttaa ennenkuin suojaus on saatu valmiiksi. Avoimuutta ei myöskään voida noudattaa tapauksissa, jossa aineiston sisällön avulla ulkopuolinen asiantuntija pystyy päätelemään jonkin innovaation, osaamisen tai muun kaupallisesti hyödynnettävissä olevan know-hown ytimen, sisällön tai sen osan. Näissäkin yhteyksissä avoimuus voi toteutua, mutta harkinnan mukaan viivästettynä.

Esteitä voi liittyä esim. user-friendly, web-based access -mahdollisuuksien toteuttamiseen. On vaikea kuvitella tutkimusprojektien rakentavan omalla kustannuksellaan yksinomaan avoimuutta varten erityisiä järjestelmiä, alustoja tai palveluja ellei se erikseen tutkimuksen rahoittajan kanssa ole otettu huomioon ja sovittu. Voi olla tilanteita, joissa julkinen rahoittajakin mieluummin investoi tutkimustuloksiin kuin avoimuusjärjestelyihin.

Sovellusohjeiden pohjana: Periaatteessa voivat, mutta on syytä varautua siihen, että käytännössä saattaa tulla vaikeitakin tilanteita. Antamalla ohjeistus hyvin yleisen tasoisena saadaan tulkinnanvaraa vaarantamatta itse päätarkoitusta. Yleisiä ohjeita voinee tällöin myös täydentää koskemaan yksityiskohtaisemmin juuri sellaisia esim. yhteiskuntatieteellisen tai terveystieteellisen tutkimuksen tuottamia aineistoja, tietokantoja ym. joissa avoimuuden kehittäminen saattaa helpoimmin parantaa nykytilaa.

Lisänäkökohtia: Jonkinlainen tarkastelu olisi syytä tehdä myös pitkälle kehitetyn avoimuuden kustannusvaikutuksista: miten paljon muutoinkin tiukoissa kehyksissä olevasta julkisesta tutkimusrahoituksesta ollaan valmiita panostamaan avautumiseen.

Avoimuuden kehittämisessä on punnittava myös turvallisuusriskit, jotka etenkin kansainvälisessä toiminnassa saattavat nykyään ja varsinkin digitaalisessa ympäristössä olla yllättäviä (esim. rikollisuus ja terrorismi).

Muita kommentteja: Kun tämän aiheen motivointi koko tutkimusyhteisölle tulee ajankohitaiseksi, siihen on syytä paneutua huomattavasti perusteellisemmin ja varata tila kunnolliselle keskustelulle. Todennäköisesti vasta sitten nähdään kattavammin eri tyyppiseen tutkimukseen liittyvät käytännön ongelmakohdat.

---

### **organisaatio: MTT (Maa- ja elintarviketalouden tutkimuskeskus)**

vastaaja: tutkimusjohtaja

Virallinen toimintalinja: Ei ole linjausta Open access -politiikassa. Ei ole nähty tarpeelliseksi. Käytännössä ollut melko liberaali linja tutkimusaineistojen luovutuksessa tutkimuskäyttöön. Useimmiten kyse vastavuoroisuudesta. Aineiston luovutuksesta sovitaan aineistokohtaisesti. Halutaan turvata tutkijan oikeus julkaista ensimmäiseksi keräämästään aineistosta. Useimmiten kysymys on myös aineiston oikeasta tulkinnasta. Joskus myös tietosuojasta.

Toimintalinjan synty: Ei ole päätöksiä. Usein aineiston hallinnasta sovitaan tutkimusosapuolten välisessä sopimuksessa.

---

### **organisaatio: Museovirasto**

vastaaja: kehittämisspäällikkö

Virallinen toimintalinja: Museovirastolla ei ole virallisesti päätettyä erillistä Open Access -toimintapolitiikkaa sähköisten tietoaineistojen käytölle. Tällä hetkellä aineistojen käyttöä sääntelevät ohjeet sisältyvät osana Museoviraston suoritteiden ja palvelujen hinnoittelu- ja jakelukäytäntöihin ja niistä annettuihin ohjeisiin.

Toimintalinjan synty: Museoviraston sähköiseen muotoon tallennetuista aineistoista keskeiset ovat kulttuuriympäristön tietokannat sekä museokokoelmien julkaisujärjestelmä. Avoimuutta tukevat toimintalinjat ovat:

- 1) Julkisessa verkossa aineistoa on hyvin rajoitetusti; linjauksena määrän lisääminen.
- 2) Aineistoja luovutetaan tutkimus- ja viranomaiskäyttöön irrotuskustannuksin (tuntiveloitteinen palvelu).
- 3) Kaupallisiin tarkoituksiin käytettäessä aineistot hinnoitellaan liiketaloudellisin perustein.

Toteuttamiskelpoisuus:

Open Access -periaatteet ovat kannatettavia. Ratkaisut Open Access -periaatteen käyttöön otosta pitäisi kuitenkin tehdä keskitetysti koko valtionhallinnon tasolla ja saattaa niiden käyttö sopusointuun viime vuosina voimistuneiden valtionhallinnon toimijoille asetettujen tuottotavoitteiden kanssa.

Sovellusohjeiden pohjana: Kyllä. Tämä edellyttäisi, että OECD:n tasolla sovittaisiin osana prosessia kansainvälisistä standardeista ja hyvistä käytännöistä.

Lisänäkökohtia: Julistuksen liitteessä kaksi esitetyt määritelmät tutkimusaineistolla ovat puutteelliset. Etenkin karttojen ja valokuvien sisältyminen tutkimusaineistoon jää epäselväksi.

## **organisaatio: Geologian tutkimuskeskus**

vastaaja: tietohallintojohtaja

Virallinen toimintalinja:

GTK:n aineistojen saatavuutta säätelevät toisaalta maksuperustelaki ja toisaalta GTK:n johdon päätökset.

Aineisto- ja hinnoittelupolitiikkaa ollaan parhaillaan selkiyttämässä ja uudistamassa.

GTK:n tutkimusaineistot voidaan jaotella kolmeen pääryhmään:

1. Julkaisut, kartat ja julkiset raportit; näiden osalta on tehty päätös, että ovat verkkovälittisinä maksuttomia. Kaikki uudet verkkopalvelun piirissä, vanhempia numeeristetaan jatkuvasti.
2. Paikkatietoaineistot; näiden osalta tehty päätös jakelusta hyvin kohtuullisesti hinnoiteltuina tuotteina. Erikoispalvelut (räätälöidyt tuotteet) tilauksesta vahvistetun työhinnaston mukaisesti.
3. Primääritiedostot (työ- ja mittaustietokannat yms.; pääosin sisäisiä tietojärjestelmiä)

Toimintalinjan synty: Paikkatietoaineistojen osalta on käyty viimeisten vuosien aikana jatkuvaa keskustelua oikeasta hinnoittelun tasosta ja -periaatteista. Keskustelu on osa koko Suomes-  
sa käytävää keskustelua aiheesta. Viime aikoina PATINEn laatima kansallinen paikkatietostra-

tegia ja tulossa oleva INSPIRE direktiivi ovat selkiyttäneet kenttää ja johtaneet kohti avointa aineistopolitiikkaa.

GTK:n asema on kuitenkin poikkeuksellinen monella tavoin; malmiesiintymien osalta laitoksella on tehtävä valvoa valtion etua; jotkut aineistot ovat olleet erittäin kalliita kerätä ja niistä kiinnostuneet tahot ovat lähinnä yhtiöitä.

Toteuttamiskelpoisuus:

Yleiset periaatteet ovat kannatettavia ja toteuttamiskelpoisia keskipitkällä aikavälillä.

Pääosa tulee GTK:n tapauksessa toteutumaan INSPIRE direktiivin implementoinnin yhteydessä.

Sovellusohjeiden pohjana:

Varmaankin, mutta sovellusohjeiden lienee oltava hyvin yleisluontoisia, jotta vältetään ristiriitailta niiden ja kansallisten suositusten/lainsäädännön kanssa.

Lisänäkökohtia: Organisaatiomme keskeisimmät aineistot ovat ns. paikkatietoaineistoja. Aiheeseen liittyen on valmisteilla INSPIRE direktiivi; lisäksi asioita pohtii PATINE, joka myös antaa kansallisia suosituksia.

Jatkovalmistelussa kannattaisi ehkä olla yhteydessä PATINEen, jotta valmistelutyötä ei tehdä koordinoimattomasti.

Muita kommentteja: On tärkeää yrittää edistää tutkimusaineistojen käytettävyyttä yleensä. Tämä asia on kuitenkin kokemukseni mukaan HYVIN monitahoinen; asiaa on vaikea kattavasti valottaa tämän tyyppiseen kyselyyn vastaamalla.

Tässä ollaan vähän samantyyppisellä alueella kuin standardoinnissa tai lainsäädännön harmonisoinnissa. Tarpeellista ja hyödyllistä, käytännössä usein ERITTÄIN hidasta.

---

## **organisaatio: Suomen ympäristökeskus**

vastaaja: osastonjohtaja

Virallinen toimintalinja:

Suomen ympäristökeskuksessa noudatetaan tiedon ja tietosuoritteiden avoimuuskysymyksissä ympäristöministeriön päätöstä, joka koskee tietosuoritteiden luovuttamista, käyttöoikeuksia ja hinnoittelua sekä päätöstä direktiivin 2003/4/EY edellyttämät tehtävät ympäristötiedon saatavuuden, aktiivisen ja järjestelmällisen levittämisen ja laadun varmistamiseksi (9.2.2005 YM23/400/2004).

Päätökset tukevat ympäristötiedon verraten avointa jatkokäyttöä. Yleinen suuntaus hallinnonalalla on siirtyä mahdollisimman avoimeen tietojen jatkokäyttöön tarjoamalla tietoja käyttöön mm. internetin kautta www-palveluina.

Toimintalinjan synty:

Tavoitteena on ollut julkisin varoin tuotettujen aineistojen mahdollisimman vapaa ja avoin käyttö. Päätökset ovat ympäristöministeriön päätöksiä ja ne on valmisteltu normaaliin tapaan työryhmätyönä.

Toteuttamiskelpoisuus: Data-tiedon osalta SYKEssä noudatetaan ympäristöministeriön yleistä data-politiikkaa, joka tulee suurella todennäköisyydellä olemaan lähellä OECD:n linjaa.

---

### **organisaatio: SC - Tieteellinen Laskenta Oy (Tieteen tietotekniikan keskus CSC)**

vastaaja: kehityspäällikkö

Virallinen toimintalinja: CSC:llä ei ole virallisesti päätettyä toimintapolitiikkaa tai -linjaa julkisrahoitteisten tutkimusaineistojen avoimesta jatkokäytöstä.

Tähän on useitakin syitä. Merkittävin syy on varmaan se, että CSC ei ole tutkimuslaitos, eikä CSC tämän takia tuota kovin isoja tutkimusaineistoja. Kaikki CSC:ssä tuotettu tutkimusaineisto on tuotettu yhdessä jonkin suomalaisen korkeakoulun tai partnerin kanssa ja mikäli asiasta on yhteisesti sovittu, on aineisto saatavilla verkon kautta.

Toteuttamiskelpoisuus:

Kyllä OECD:n julkilausumassa esitetyt periaatteet ovat toteuttamiskelpoisia myös CSC:ssä. CSC:llä on myös tietyssä määrin tehtävänä levittää muiden tutkimusryhmien tietoaineistoja avoimesti verkkoon. Tämä toteutuu esimerkiksi biokemiassa, jossa muualla tuotettua tutkimusaineistoa talletetaan CSC:n tietokantoihin, jotka ovat mahdollisuuksien mukaan kaikkien avoimesti käytettävissä.

Lisänäkökohtia: Tutkimusta rahoittavien tahojen tulisi ehdottomasti vaatia jo tutkimushakemuksessa edellyttää suunnitelmaa siitä, miten tutkimustulokset on tarkoitus viedä avoimesti verkkoon.

Muita kommentteja: Erittäin hyvä asia, joka vaatii toteutuakseen niin poliittisia päätöksiä kuin tieteellisten toimintatapojen muuttamista.

---

## organisaatio: Kuluttajatutkimuskeskus

vastaaja: tutkimuspäällikkö

Viitaten saamaani kyselyyn julkisrahoitteisten tutkimusaineistojen dokumentoinnista pelisääntöineen ml. kannanotot OECD:n periaatteisiin totean Kuluttajatutkimuskeskuksen osalta seuraavan:

Kuluttajatutkimuskeskuksessa on kerätty 15 toimintavuoden aikana sekä määrällisiä että laadullisia aineistoja mm. omaa kuluttajapaneelia hyväksi käyttäen sekä väestöpohjaisten survey-tutkimusten avulla. Lisäksi käytämme mm. (ostamme) Tilastokeskuksen aineistoja, mm. kulutustutkimusta ja ajankäyttöä koskien. Edelleen olemme käyttäneet ostopalveluja (esim. Taloustutkimus).

Kuluttajatutkimuskeskuksessa on menossa tietohallintostrategian kehittämis- ja uudistamistyö, minkä osana laaditaan myös strategia ja suunnitelma tutkimusaineistojen (määrälliset, laadulliset, kuvamateriaalit) dokumentoimiseksi. Tutkimusaineistojen dokumentointiprojektin pitäisi valmistua kuluvan kevään aikana. Tämän jälkeen voimme palata varsinaiseen kyselyyn ja kannanottoihin aineistojen jatkoluovuttamisesta. Luovutettavia tutkimusaineistoja meillä on varmasti.

Yleisinä periaatteina OECD:n aloitteet ovat ihan ok. Miten lienevät sitten käytännössä erilaiset juridiset säädökset, maksatusasetukset ym.

---

## organisaatio: Kansanterveyslaitos

vastaaja: Kirjasto ja tietopalveluyksikön johtaja

Virallinen toimintalinja: Kansanterveyslaitoksen strategiassa vuodelta 2001 todetaan, että ”Tieto- ja näyteaineistot ovat Kansanterveyslaitoksen arvokkainta omaisuutta. Aineistojen hyvä hallinta on strateginen tavoitteemme ja edellytys sille, että ne voidaan asettaa koko tutkijayhteisön käyttöön. Huolehdimme siitä, että tietoa-aineistojemme käytettävyyttä turvataan pysyvästi työvälineiden uudistuessakin”.

Toisaalta KTL:n johtoryhmässä tehtiin helmikuussa 2005 periaatepäätös siitä, että kaikkein suurimmista tutkimusaineistoista ei tulla antamaan dataa avoimesti ulkopuolisille, koska datan jatkokäyttö edellyttää sellaista tietämystä ja aineiston syvällistä tuntemusta, jota on vain KTL:ssä.

Toimintalinjan synty: Ks. edellä. Lisäksi laitoksessamme valmistellaan parhaillaan sisäistä määräystä Kansanterveyslaitoksessa noudatettavista käytännöistä, harkittaessa aikaisemmin kerättyjen tutkimusaineistojen käyttöä uusien tutkimussuunnitelmien toteuttamiseen. Määräys-

luonnoksessa todetaan mm., että vanha tutkimustarkoitukseen kerätty, henkilötietoja sisältävä tietoaaineisto on käytettävissä vain tutkittavan tietoisessa suostumuksessa tai muussa hänelle annetussa informaatiossa esitettyyn käyttötarkoitukseen.

Toteuttamiskelpoisuus:

Periaatteet sinänsä ovat kannatettavia ja osittain toteuttamiskelpoisia.

Terveystieteisiin liittyvällä tutkimusdatalla on usein sensitiivinen luonne. Lisäksi perusteellisesta dokumentoinnista huolimatta ulkopuolinen tutkija ei yleensä tunne aineistoa riittävän hyvin voidakseen käyttää sitä asianmukaisesti. Tästä voi seurata virhetulkintoja ja perusteettomia tuloksia, joita tutkijamme joutuvat korjaamaan. Tästä syystä yksilötason aineistoja ei pääsääntöisesti luovuteta, vaan KTL priorisoi tutkimusyhteistyötä datan jatkokäytössä. Jo nyt KTL:n julkisessa verkkopalvelussa on aggregoituja KTL:n aineistoista muodostettuja tilastotietokantoja selattavissa ja analysoitavissa.

Tutkimuseettiset näkökohdat ovat tärkeitä ihmisistä ja ihmisiltä kerättävien aineistojen käsittelyssä. Tulee huolehtia siitä, että henkilön anonymiteetti säilyy ja, että tutkittavilta henkilöiltä kerättävälle datalle on asianmukaiset suostumukset olemassa. Datan jatkokäsittelyssä on aina otettava huomioon suostumusten laajuus, eli millaiseen käsittelyyn tutkittava alun perin on suostumuksensa antanut.

Sovellusohjeiden pohjana: Edellä kuvatuilla edellytyksillä kyllä. Eräs tutkijamme kommentoi-kin periaatteita siten, että ”maksimaalisen käytön” sijasta tulisi mieluummin puhua ”optimaalisesta käytöstä”, koska maksimaalinen käyttö ei aina -ainakaan edustamillamme aloilla - johda maksimaaliseen hyötyyn.

Lisänäkökohtia: Tieteenalakohtaisten suositusten laatimista voisi ehkä pohtia.

---

**organisaatio: Kotimaisten kielten tutkimuskeskus**  
(Research Institute for the Languages of Finland)

vastaaja: tutkimusosaston johtaja

Virallinen toimintalinja:

Kotuksen sähköisiä aineistoja ja korpuspalvelinta saa käyttää tutkimustarkoituksiin. Monissa aineistoissa tai osa-aineistoissa tämä on erillisenä ehtona. Käyttöluvan myöntää tietohuolto-osaston johtaja, ja siihen sisältyy korpuspalvelimen käyttöoikeus. -- Käyttöoikeus on siis suhteellisen tiukasti rajattu, mutta se laajenee merkittävästi, kun käyttöön otetaan erillinen www-pohjainen asiakaspalvelin, ilmeisesti v. 2006.

Tutkimushankkeissa kerättyjä video- ja äänitallenteita voi käyttää tutkimustarkoituksiin tietohuolto-osaston johtajan ja vastuuhenkilön luvalla. (Aineistoihin liittyvä kuvattujen tai nauhoitettujen henkilöiden luvat, jotka myös rajaavat käyttöoikeutta.)



Ei-sähköiset aineistot (sanalippuja, äänitteitä ym.) ovat vapaasti käytettävissä Kotuksen tiloissa, ja niistä voi tilata kopioita.

Toimintalinjan synty:

Toimintalinjoista on päätetty tutkimuskeskuksen (Kotuksen) johtoryhmässä 19.6.2001 tietohuolto-osaston johtajan tekemän muistion perusteella. (Yleisperiaate sähköisten aineistojen käyttöluvista on myös Kotuksen tutkimuspoliittisessa ohjelmassa vuodelta 2000, joka on myös johtoryhmän ja silloisen johtokunnan hyväksymä.)

Toimintaohjeet ovat luettavissa Kotuksen www-sivuilla: [www.kotus.fi/aineistot](http://www.kotus.fi/aineistot).

Kotuksen toiminta- ja taloussuunnitelmassa vuosiksi 2006-09 mainitaan uuden asiakaspalvelimen julkistaminen ja sähköisten aineistojen lisääminen. Suunnitelma liitetään valtion Netrasivustoon.

Toteuttamiskelpoisuus: Periaatteet ovat hyviä ja toteuttamiskelpoisia.

---

## **organisaatio: Työterveyslaitos**

vastaaja: erikoistutkija

Virallinen toimintalinja: Mitään erillistä päätöstä asiasta ei ole. Kuitenkin Työterveyslaitoksen toiminnassa on vuosikymmeniä ollut käytäntö, että tieteellisissä tutkimuksissa on varauduttu mahdollisiin jatkotutkimuksiin (ensisijaisesti luonnollisesti laitoksen omaan jatkokäyttöön, mutta myös yhteistyötutkimukseen muiden tutkimusorganisaatioiden kanssa sekä tietojen lainmukaiseen luovutukseen laitoksen ulkopuolisille tutkimusorganisaatioille ja tutkijoille). Tutkimusaineistojen käyttö on ollut sekä kotimaista että kansainvälistä.

Toimintalinjan synty: Tutkimuksen avoimuutta koskevat päätökset ovat paljolti pohjautuneet siihen, että Työterveyslaitos on WHO:n referenssilaitos, jossa on perinteisesti runsaasti kansainvälistä yhteistyötä. Samoin laitoksen tutkijat toimivat yhteistyössä sekä muiden tutkimuslaitosten että korkeakoulujen kanssa. Tutkimusaineistojen käytettävyyden tulee kuitenkin aina perustua lailliseen käyttöoikeuteen ja sen todentamiseen, jotta tutkimushenkilöiden ja asiakkaiden luottamus voidaan säilyttää. Tutkimuslupia myönnettäessä otetaan huomioon myös julkisuuslain tutkijan oikeuksia koskevat säännökset, joten lupa pyritään aina myöntämään ellei siihen ole estettä.

Toteuttamiskelpoisuus:

OECD:n julkilausumassa ei ole periaatteellisesti sellaista, mitä ei voisi toteuttaa (ja jonka mukaista menettelyä on toteutettukin) myös Työterveyslaitoksessa. Periaatteethan rakentuvat kuitenkin sekä kansainvälisen että kansallisen lainsäädännön pohjalle.

Työterveyslaitoksen tutkimusaineistot ovat valtaosaltaan salassa pidettäviä arkaluonteisia aineistoja, mikä asettaa omat rajoituksensa. Lisäksi aineistojen koko on tyypillisesti pienehkö.

Työterveyslaitos ei myöskään ole tilastoviranomainen, joka tuottaisi laajoja tilastoaineistoja julkisin varoin. Lisäksi osa laitoksen tutkimusaineistosta saadaan palvelutoiminnan aineistosta laitoksen omaan käyttöön asiakkaiden (yritysten) luvalla ja kohdehenkilöiden nimenomaisella suostumuksella.

Sovellusohjeiden pohjana: Periaatteet soveltuvat OECD:n Guidelines'in pohjaksi.

Lisänäkökohtia: Valmistelussa tulisi varmaan ottaa huomioon jossain määrin aineistojen erilaisuus. Suuret kansalliset tilasto- ja rekisteriaineistot ovat aivan toisen luokan kysymys omine ongelmiseen, kun taas tutkimuslaitosten pienet hyvin arkaluonteisia tietoja sisältävät aineistot muodostavat osin erilaisen kuvion eri tavoin ratkaistavia kysymyksiä, vaikka peruseriaatteet olisivat samat.

Muita kommentteja: OECD:n Guidelines tulisi saada hyvissä ajoin myös lausunnolle, koska sillä on merkitystä myös yksittäisten tutkimuslaitosten tutkimuslupien myöntäjille.

---

## **organisaatio: Metsäntutkimuslaitos**

vastaaja: tutkimusjohtaja

Virallinen toimintalinja: Tutkimusaineistojemme avoin käyttö tapahtuu yleensä yhteistutkimusten merkeissä. Tutkimuksen ajan aineistomme ja ohjelmistomme ovat kaikkien yhteistyösapuolten käytössä. Aineistoja, joiden ”ensimmäinen käsittely” on jo tehty, annetaan sopimuksesta ulkopuolisten käyttöön erityisesti opinnäytetöiden tekoon. Avoimuutta rajoittaa jossakin määrin se, että osa aineistoistamme sisältää yksityisyyden suojaamaa tietoa tai on asiakasrahoitteista, tuotteistettua toimintaamme. Metla on avannut pitkäaikaiset koesarjansa pohjoismaisella tasolla ja on myös pyrkinyt tekemään saman koko Euroopassa EU-hanke-esityksenä (ei ole toistaiseksi saanut rahoitusta).

Toimintalinjan synty: Metlalla ei ole toistaiseksi asiasta virallisia asiakirjoja.

Toteuttamiskelpoisuus: Julkilausumassa asiat on sanottu oikeassa muodossa. Yksityisyyden suoja on syytä tunnustaa. Aineistojen täysi avoimuus voi myös haitata organisaation kaupallista toimintaa, josta aiheutuvat kustannukset tulisi jollakin tavoin pyrkiä korvaamaan.

Sovellusohjeiden pohjana: Periaatteet antavat mielestäni melko hyvän pohjan sovellusohjeiden laadintaan.

Muita kommentteja: Tutkijan keräämän aineiston ”ensiyön oikeus” (esim. väitöskirja-aineisto) pitäisi pyrkiä jollakin tavoin säilyttämään.

---

## LIITE 4.

## Aineistonkuvailun tietoelementit

Tietoarkistokielellä ilmaisten aineiston tosiasialliseen käytännön arvoon vaikuttaa ratkaisevasti aineiston *kuvailu* eli *dokumentaatio*. Yhtä tärkeää on valita sähköiset tallenneformaattit sellaisiksi, että aineistot ovat käytettävissä mahdollisimman pitkään.

Tietoarkistoon arkistoidut aineistot kuvaillaan kansainvälisen kvantitatiivisten tutkimusaineistojen kuvailuun kehitetyn DDI-formaatin mukaisesti. Se sisältää yli 300 kenttää, joista Yhteiskuntatieteellisessä tietoarkistossa käytetään noin kolmasosa. Formaatin tarkan esittelyn sijasta tässä esitetään tietoelementit, jotka ovat hyvin olennaisia siitä riippumatta, arkistoidaanko tutkimusaineisto aineistonkerääjän omaan tutkimuslaitokseen vai varsinaiseen arkistoon. Tarkemmat tiedot löytyvät kirjasta Tutkimusetiikka – aineiston hankinta, käyttö ja säilytys (Kuula 2006, 223–226).

*Aineiston tekijä/t.* Aineiston tekijöitä ovat henkilöt, jotka ovat vastuussa aineiston varsinaisesta ajatuksellisen sisällön kehittämisestä.

*Aineiston aihepiiri ja sisältö.* Olennainen tieto aineistosta on sen alkuperäinen tarkoitus (so. tutkimus), aineiston luonne ja sisällön tärkeimmät aihealueet. Kyselyaineistoista voi kirjata muistiin myös keskeiset muuttujaryhmät.

*Aineistoyksikkö/havaintoyksikkö.* Kvalitatiivisesta aineistosta on lueteltava kaikki kerätyt aineistot. Ne voivat olla haastatteluja, vuorovaikutuksen tallenteita, päiväkirjoja kenttämuistiinpanoja, lehtileikkeitä jne. Aineistoyksikkökohtaiset tiedot kirjataan huolella. Perustiedot aineistoyksiköistä on hyvä olla kunkin yksikön yhteydessä, esimerkiksi haastattelulitteraatin alussa, ja myös erillisenä aineistolistauksena. Kvantitatiivisen aineiston havaintoyksikkö on kohde, josta tutkimuksessa pyritään tekemään empiirisiä havaintoja. Se voi olla yksilö, ryhmä, organisaatio, asiakysymys, tapahtuma, jäsenyys, kotitalous, kunta, seutukunta, maakunta, lääni, valtio, seurakunta, yritys, maatila, sanomalehti, vastaanottokäynti jne.

*Aineiston valinta tai otantamenetelmä.* Kvalitatiivisesta aineistosta kerrotaan esimerkiksi haastateltavien, lehtijuttujen tms. valintaperusteet ja keinot joita tutkittavien tavoittamiseksi ja aineiston kartuttamiseksi on käytetty. Kvantitatiivisesta tutkimuksesta kerrotaan otannan tyyppi ja rakenne, jota on käytetty valittaessa vastaajat edustamaan tutkimuksessa koko populaatiota.

*Aineiston aikaulottuvuus.* Kvalitatiivisesta aineistosta kerrotaan, onko aineistoa kerätty vain kerran vai toistuvasti samoilta tutkittavilta/tutkimuskohteesta. Kvantitatiivisen aineiston aikaulottuvuuden kertoo se, onko aineisto poikkileikkausaineisto, paneeliaineisto, pitkittäisaineisto vai kenties aikasarja-aineisto.

*Aineiston tekninen keruu ja tiedostojen formaatit.* Olennainen tieto aineistosta on sen tekninen keruutapa. Keruussa käytetyt välineet kerrotaan aineiston kuvailutiedoissa. Myös tiedot säily-

tettävän aineiston tiedostoformaateista ovat hyvin olennaisia, jotta tiedetään, täytyykö aineistoja aika ajoin konvertoida.

*Tiedostojen nimet ja selitteet.* Tutkimusaineiston sähköiset tiedostot täytyy nimetä selkeää loogiikkaa noudattaen. Jokaista aineistokokonaisuutta varten on luotava lisäksi erillinen dokumentti (esimerkiksi nimellä tiedostot.txt), jossa selitetään kunkin tiedoston nimi, formaatti ja sisältö.

*Asiakirjat/keruuvälineet aineiston sisällön tuottamisessa.* Aineiston sisältöä määrittävät olennaisesti aineistonkerääjille ja erityisesti tutkittaville annetut sisällölliset ohjeet tai aineistonkeruun instrumentit. Sellaisia ovat esimerkiksi kyselylomake, haastattelukysymykset, kirjoituskehotus tai -pyyntö, ryhmähaastattelun virittävät kuvat, tekstit tai tutkijan alkupuhe keskustelun viritämiseksi. Aineistoa kontekstoivia dokumentteja ovat myös haastattelijoille tai aineiston käsitteijöille (koodaajille tai litteroijille) annetut ohjeet.

Kuvailun merkitys ei rajoitu vain aineiston säilyttämiseen ja jatkokäyttöön. Aineiston ja sen keruun huolellinen toteuttaminen ja kuvaus ovat olennainen osa tutkimusprosessia. Tutkimusaineiston perustietojen kattava kuvaus on tärkeä luotettavuuden osoitus tutkimustulosten raportoinnissa.

## Kirjallisuus ja verkkolähteet

Arzberger, Peter & Schroeder, Peter & Beaulieu, Anne & Bowker, Geof & Casey, Kathleen & Laaksonen, Leif & Moorman, David & Uhliir, Paul & Wouters, Paul (2004a) An International Framework to Promote Access to Data. *Science* 303(5665), 1777.

Arzberger, Peter & Schroeder, Peter & Beaulieu, Anne & Bowker, Geof & Casey, Kathleen & Laaksonen, Leif & Moorman, David & Uhliir, Paul & Wouters, Paul (2004b) Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal* 3, 135–152.

Bo-Christer, Björk (2004) Open access to scientific publications – an analysis of the barriers to change? *Information Research* 9(2). Verkkójulkaisu: <http://informationr.net/ir/9-2/paper170.html> [Viitattu 10.5.2007]

Clubb, Jerome & Austin, Erik & Geda, Carolyn & Traugott, Michael (1985) Sharing Research Data in the Social Sciences. Teoksessa Fienberg, Stephen E. & Martin, Margaret E. & Straf, Miron L. (eds.) *Sharing Research Data*. Washington, D.C.: National Academy Press, 30–88.

Corti, Louise & Foster, Janet & Thompson, Paul (1995) Archiving qualitative research data. *Social Research Update*. Issue 10, Autumn 1995. Verkkójulkaisu: <http://www.soc.surrey.ac.uk/sru/SRU10.html> [Viitattu 1.2.2007]

Corti, Louise (2006a) Editorial. *Methodological Innovations Online* 1(2). Verkkójulkaisu: <http://siri.us.soc.plymouth.ac.uk/~andyp/viewarticle.php?id=33&layout=html> [Viitattu 11.12.2006]

Corti, Louise (2006b) ESDS Qualidata: accessing, exploring and using data. *Research methods festival*. Oxford, July 2006. Verkkójulkaisu: <http://www.ccsr.ac.uk/methods/festival/programme/rwa/> [Viitattu 31.1.2007]

DDI Alliance (2004) DDI Version 3 Conceptual Model (draft, June 2004). Verkkójulkaisu: <http://www.icpsr.umich.edu/DDI/committee-info/Concept-Model-WD.pdf> [Viitattu 28.7.2007]

Duncan, George T. (2004) Exploring the Tension Between Privacy and the Social Benefits of Governmental Databases. Paper presented at Security, Technology, and Privacy: Shaping a 21st Century Public Information Policy, 2003 April 24–25, Georgetown University Law Center, Washington, DC. Conference sponsored by Carnegie Mellon University, Georgetown University Law Center, The Carnegie Corporation of New York, and The Century Foundation. Verkkójulkaisu: <http://www.ccsr.ac.uk/methods/archive/AccessGrid/documents/ExploringtheTension.pdf> [Viitattu 11.12.2006]

Edelman, Tom & Mylly, Tuomas & Rajala, Katariina (1998) Immateriaalioikeudet tutkijan kannalta. Teoksessa Saarnilehto, Ari (toim.) *Tutkijan oikeudet ja velvollisuudet*. Helsinki: Werner Söderström lakitieto, 21–56.

ESRC Research Funding Guide (2005) Verkkójulkaisu: [http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Research\\_Funding\\_Guide\\_2005-Mar07\\_tcm6-9734.pdf](http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Research_Funding_Guide_2005-Mar07_tcm6-9734.pdf) [Viitattu 22.1.2007]

Fielding, Nigel (2000) The Shared Fate of Two Innovations in Qualitative Methodology: The Relationship of Qualitative Software and Secondary Analysis of Archived Qualitative Data [43 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3). Verkkójulkaisu: <http://www.qualitative-research.net/fqs-texte/3-00/3-00fielding-e.htm> [Viitattu 30.11.2006]

Gazzaniga, Michael S. & Rockinore, Daniel N. (2001) Data Are Most Useful When Openly Shared. *The Chronicle of Higher Education*, March 16/2001. Verkkójulkaisu: [http://www.cs.dartmouth.edu/~rockmore/Data\\_Chron.html](http://www.cs.dartmouth.edu/~rockmore/Data_Chron.html) [Viitattu 26.1.2007]

Gibson, James (1991) Cautious Reflections on a Data Archiving Policy for Political Science. *Political Science & Politics* 28(3):473–476.

Gleditsch, Nils & Metelits, Claire & Strand, Håvard (2003) Posting Your Data: Will You Be Scooped or Will You Be Famous? *International Studies Perspectives* 4(1), 89–97.

- Heeney, Catherine (2004) *The Role of Privacy and Confidentiality in the Work of National Statistical Institutes*. Unpublished Ph.D. thesis, submitted to the University of Manchester.
- Henkilötietolaki 22.4.1999/523.
- Hyvä tieteellinen käytäntö (2002). *Tutkimuseettinen neuvottelukunta*.
- Kansanperinteen arkisto. Musiikintutkimuksen laitos. Kotisivut: <http://www.uta.fi/laitokset/mustut/arkistosivut/> [Viitattu 21.12.2006]
- Kansanterveyslaitos (2005) Hyvä tutkimustapa Kansanterveyslaitoksessa. Kansanterveyslaitoksen julkaisuja B 2/2005. Verkkajulkaisu: [http://www.ktl.fi/attachments/suomi/julkaisut/julkaisusarja\\_b/2005/2005b2.pdf](http://www.ktl.fi/attachments/suomi/julkaisut/julkaisusarja_b/2005/2005b2.pdf) [Viitattu 11.5.2007]
- Keckman-Koivuniemi, Hannele & Kleemola, Mari (2006) *Opas kvantitatiivisten tutkimusaineistojen jatkokäyttöön*. Tampere: Yhteiskuntatieteellinen tietoaristo. Verkkajulkaisu: [http://www.fsd.uta.fi/julkaisut/julkaisusarja/FSDjs02\\_jatkokaytto.pdf](http://www.fsd.uta.fi/julkaisut/julkaisusarja/FSDjs02_jatkokaytto.pdf) [Viitattu 26.1.2007]
- Kiikeri, Mika & Ylikoski, Petri (2004) *Tiede tutkimuskohteena: filosofinen johdatus tieteen tutkimukseen*. Helsinki: Gaudeamus.
- King, Gary (1995) *Replication, Replication*. PS. *Political Science and Politics* 28(3), 443–499.
- King, Gary (2006) *Data Sharing and Replication*. Verkkajulkaisu: <http://gking.harvard.edu/replication.shtml> [Viitattu 12.12.2006]
- Kiuru, Essi & Koponen, Sanna (2001) *Sopimukset yliopistojen tutkimustyössä*. Oulu: Oulun yliopisto, Koulutus- ja tutkimuspalvelut.
- Kleemola, Maija (2000) *Tietosuojat ja tieteellinen tutkimus henkilötietojen kannalta*. *Tietosuojat* 12(4), 20–30.
- Klemettinen, Pasi (toim.) (2006) ”Ei se synny synnyttämättä”. *Selvitys digitointiprojektin vaiheista ja työprosesseista*. Suomalaisen kirjallisuuden seuran raportti 1/2006. Helsinki: SKS. Verkkajulkaisu: [www.finlit.fi/tutkimus/elias\\_loppuraportti.pdf](http://www.finlit.fi/tutkimus/elias_loppuraportti.pdf) [Viitattu 6.2.2006]
- Kotimaisten kielten tutkimuskeskus. Kotisivut: <http://www.kotus.fi/> [Viitattu 21.12.2006]
- Kulttuurien tutkimuksen laitoksen TYKL ja TKU arkistojen kotisivut (Turun yliopisto): <http://kultut-arkistot.utu.fi/index.html> [Viitattu 21.12.2006]
- Kuula, Arja (2007) *Laadullinen tutkimussuhde ja arkistoinnin etiikka*. *Tieteessä tapahtuu* 24(2), 42–45.
- Kuula, Arja (2006) *Tutkimusetiikka – aineistojen hankinta, käyttö ja säilytys*. Jyväskylä: Vastapaino.
- Launis, Veikko (2007) *Tutkimuksen eettinen ennakkoarviointi – mitä se on?* *Tieteessä tapahtuu* 24(1), 28–33.
- Leppänen, Markku (2006) *Miten tutkimusaineistojen säilytysarvo tulisi määritellä? Esitelmä Arkistoyhdistyksen syysseminaarissa 3.11.2006* Tieteiden talo, Helsinki.
- Lowrance, William (2006) *Access to Collections of Data and Materials for Health Research. A report to the Medical Research Council and the Wellcome Trust*. Verkkajulkaisu: [www.wellcome.ac.uk/assets/wtx030842.pdf](http://www.wellcome.ac.uk/assets/wtx030842.pdf) [Viitattu 19.1.2007]
- Mahlamäki, Tiina (2002a) *Perinnearkiston tehtävät digitoinnin aikakaudella. Säilyvyys, saatavuus ja prosessointi*. Teoksessa *Arkistot ja kulttuurit, toim. Arkistoyhdistys ry:n toimituskunta*, 151–163.
- Mahlamäki, Tiina (2002b) *Löytyykö se netistä? Digitaalisen tutkimusaineiston käytön erityispiirteitä*. *Elore* 9(2). Verkkajulkaisu: [http://cc.joensuu.fi/~loristi/2\\_02/mah202.html](http://cc.joensuu.fi/~loristi/2_02/mah202.html) [Viitattu 11.5.2007]
- Mahlamäki, Tiina (2002c) *Turun yliopiston kulttuurien tutkimuksen laitoksen arkistot*. *Agricolan Tietosanomata* 4/2002. Verkkajulkaisu: <http://www.utu.fi/agricola/tietosanomata/numero4-02/kultutlaitos.html> [Viitattu 11.5.2007]
- Mauthner, Natasha & Parry, Odette & Backett-Milburn, Kathryn (1998) *The data are out there, or are they? Implications for archiving and revisiting qualitative data*. *Sociology*, 32(4), 733–745.
- Medical Research Council (2007) *MRC Policy on Data Sharing and Preservation* Verkkajulkaisu: [www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/](http://www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/)

- DataSharing/PolicyonDataSharingandPreservati-  
on/index.htm [Viitattu 18.1.2007]
- Mäkelä, Klaus (2007) Lääketieteen etiikka ja yhteis-  
kuntatutkimuksen ennakkosääteily. *Tieteessä tapahtuu*  
24(1), 21–27.
- National Science Board (2005) Long-Lived Digital  
Data Collections: Enabling Research and Education  
in the 21st Century. National Science Foundation.  
Verkkojulkaisu: [http://www.nsf.gov/pubs/2005/  
nsb0540/](http://www.nsf.gov/pubs/2005/nsb0540/) [Viitattu 18.1.2007]
- Nieminen, Risto (2007) *Tiede 2.0. Akatiimi 2/2007*.
- Niiniluoto, Ilkka (2002) Jaettu vastuu kannattelee hy-  
vää tiedettä. *Tieteessä tapahtuu* 20(4). Verkkojulkaisu:  
<http://www.tsv.fi/ttapaht/024/niiniluoto402.pdf>  
[Viitattu 11.5.2007]
- Nolte, Michael and Keller, Janet (2001) Access to Re-  
stricted Data in a Controlled Environment: The Michi-  
gan Center on the Demography of Aging Data Enclave.  
Verkkojulkaisu: [http://micda.psc.isr.umich.edu/encla-  
ve/EnclaveJSM2001.pdf](http://micda.psc.isr.umich.edu/enclave/EnclaveJSM2001.pdf) [Viitattu 23.8.2007]
- NSF (2005) Grant policy manual. Verkkojulkaisu:  
[http://www.nsf.gov/pubs/manuals/gpm05\\_131/  
gpm05\\_131.pdf](http://www.nsf.gov/pubs/manuals/gpm05_131/gpm05_131.pdf) [Viitattu 18.1.2007]
- Pelkonen, Anna & Mäkelä, Mika & Haahtela, Tari  
(2004) Kliinisen tutkijan seikkailu byrokratian viida-  
kossa. *Duodecim* 120(23), 2815–2821.
- Ray, James & Valeriano, Brandon (2003) Barriers  
to Replication in Systematic Empirical Research on  
World Politics. Julkaisussa deMesquita, Bruce Bueno &  
Gleditsch, Nils Petter & James, Patrick & King, Gary &  
Metelits, Claire & Ray, James Lee & Russett, Bruce &  
Strand, Håvard & Valeriano, Brandon (2003) Sym-  
posium on Replication in International Studies Research.  
*International Studies Perspectives* 4(1), 72–107. Verk-  
kojulkaisu: [http://gking.harvard.edu/files/replvdc.  
pdf](http://gking.harvard.edu/files/replvdc.pdf) [Viitattu 24.1.2007]
- RIN (2007) Stewardship of digital research data: a fra-  
mework of principles and guidelines. Responsibilities  
of research institutions and funders, data managers,  
learned societies and publishers. Draft for consultation,  
April 2007. Verkkojulkaisu: [http://www.rin.ac.uk/fi-  
les/Research%20Data%20Principles%20and%20  
Guidelines%20-%20published%20draft%20for%20  
consultation.pdf](http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20-%20published%20draft%20for%20consultation.pdf) [Viitattu 7.5.2007]
- Saarnilehto, Ari & Edelman, Tom (1998) Tutkimusso-  
pimus. Helsinki: Werner Söderström lakitieto.
- Salokannel, Marjut (2003) Immateriaalioikeudet – tut-  
kimuksen turva vai este? *Tieteessä tapahtuu* 21(1), 27–  
29. Verkkojulkaisu: [http://www.tsv.fi/TTAPAHT/  
031/Salokannel.pdf](http://www.tsv.fi/TTAPAHT/031/Salokannel.pdf) [Viitattu 11.5.2007]
- Sieber, Joan (1991) Social Scientists' Concerns About  
Sharing Data. Teoksessa Sieber, Joan (ed.) *Sharing So-  
cial Science Data: Advantages and Challenges*. Lon-  
don: Sage.
- Tekijänoikeuslaki 8.7.1961/404
- Tekijänoikeusneuvoston lausunto 1999:15. Tekijän-  
oikeudet harjututkimuksen yhteydessä syntyneeseen  
aineistoon. Verkkojulkaisu: [http://www.minedu.fi/  
opm/asiantuntijat/tekijanoikeusneuvosto/tekijan-  
oikeusneuvosto.html](http://www.minedu.fi/opm/asiantuntijat/tekijanoikeusneuvosto/tekijan-<br/>oikeusneuvosto.html) [Viitattu 12.12.2006]
- Terveystieteiden ja -talon laitoksen laatuksi-  
kirja (2006). Kuopion yliopisto, versio 1.0. Koonnut  
Taskinen, Helena. Verkkojulkaisu: [http://www.uku.  
fi/tht/laatuksikirja/laatuksikirja1.0.pdf](http://www.uku.fi/tht/laatuksikirja/laatuksikirja1.0.pdf) [Viitattu  
11.5.2007]
- University of Columbia (2006) Responsible Conduct  
of Research. Verkkojulkaisu: [http://www.unl.edu/re-  
search/orr/rcr\\_modules/rcr\\_data/introduction/in-  
dex.html](http://www.unl.edu/research/orr/rcr_modules/rcr_data/introduction/in-<br/>dex.html) [Viitattu 11.12.2006]
- Weigand, Hans & Heeney, Cate (2003) Privacy protec-  
tion and communicative respect. Proceedings of the 8th  
International Working Conference on the Language-  
Action Perspective on Communication Modelling,  
LAP 2003 Tilburg, The Netherlands, July 1–2, 2003.  
Verkkojulkaisu: [http://infolab.uvt.nl/~weigand/  
weigandheeney.pdf](http://infolab.uvt.nl/~weigand/weigandheeney.pdf) [Viitattu 12.12.2005]
- Yhteiskuntatieteellinen tietoaarkisto. Kotisivut: [http://  
www.fsd.uta.fi/](http://www.fsd.uta.fi/) [Viitattu 11.5.2007]
- Yliopistolaki 27.6.1997/645



YHTEISKUNTATIETEELLINEN  
TIETOARKISTO

Käyntiosoite: Åkerlundinkatu 2 A, 5. krs  
Tampere  
Postiosoite: Yhteiskuntatieteellinen tietoaarkisto  
33014 Tampereen yliopisto  
Puhelin: (03) 3551 8519  
Faksi: (03) 3551 8520  
Sähköposti: [fsd@uta.fi](mailto:fsd@uta.fi)  
Kotisivu: <http://www.fsd.uta.fi>

