

Julkaisutietojen 2011 keruu yliopistoista

Selvitys TUTKI-ohjausryhmälle

Sisällysluettelo

1	Johdanto	2
2	Tiedonkeruun eteneminen	3
2.1	Tietojen siirto	3
2.2	Tietojen tarkastus ja korjaus	5
2.2.1	Viitetietojen oikeellisuus	5
2.2.2	Yliopistojen sisäiset duplikaatit	8
2.2.3	Ristiriitaiset julkaisutyypiluokitukset	8
2.3	Julkaisun tieteenala	10
2.4	Julkaisufoorumiluokituksen yhdistäminen julkaisuihin	10
2.5	Tietojen julkaiseminen	13
3	Yhteenveto ja suositukset	13
	Liite 1 Opetusministeriön julkaisutyypiluokittelu 2010	15
	Liite 2 Tilastokeskuksen tieteenalaluokitus	16
	Liite 3. Yhteenveto yliopistojen tiedonkeruukyselyn vastauksista julkaisujen osalta	17
1.	Määrittelyjen yksiselitteisyys	17
2.	Tietojen kerääminen tietojärjestelmistä	18
3.	Tietojen vertailtavuus korkeakoulujen välillä	19
4.	Tietojen käyttö korkeakoulussa	19
5.	Esimerkkilaskentoja	20

1 Johdanto

RAKETTI-hankkeen tutkimushallinnon osa-aluetta tarkastelevan TUTKI-osahankkeen ohjausryhmä on asettanut yhdeksi tavoitteeksi korkeakoulujen perusrahoituksen kriteerinä käytettävien tietojen laadun varmistuksen. Erityisesti tieteellisistä julkaisuista kerättävänä tiedon laatu ja yhdenmukaisuus koetaan tärkeänä, sillä niiden painoarvo yliopistojen rahoitusmallissa vuosille 2013–16 on merkittävä: 13 prosenttia koko perusrahoituksesta.

Vuosien 2013–14 rahoitusmallissa julkaisuista huomioidaan tieteelliset julkaisut, jotka on jaoteltu kansainvälisiin referee-julkaisuihin ja muihin tieteellisiin julkaisuihin (kotimaiset referoidut ja ei-referoidut ja ulkomaiset ei-referoidut). Vuodesta 2015 alkaen julkaisujen luokittelussa on suunniteltu otettavan käyttöön julkaisuforumiluokitus, jossa julkaisukanavat on jaoteltu kolmelle tasolle. Rahoitus jaetaan kansainvälisten referee-julkaisujen sijaan julkaisuforumin tasojen 2 ja 3 julkaisujen määrän perusteella sekä muiden tieteellisten julkaisujen sijaan julkaisuforumin tason 1 julkaisujen määrän perusteella.

TUTKI-ohjausryhmä antoi tehtäväksi selvityksen tuottamisen julkaisutiedonkeruun onnistumisesta ja tiedonkeruuprosessista vuotta 2011 koskevassa yliopistojen tiedonkeruussa. Selvityksen ovat laatineet

ylitarkastaja Jukka Haapamäki, Opetus- ja kulttuuriministeriö
vanhempi sovellusasiantuntija Yrjö Leino, CSC - Tieteen tietotekniikan keskus Oy
korkeakouluharjoittelija Hanne Mikkonen, Opetus- ja kulttuuriministeriö
koordinaattori Hanna-Mari Puuska, CSC - Tieteen tietotekniikan keskus Oy
suunnittelija Janne Pölonen, Tieteellisten seurain valtuuskunta

Vuoteen 2010 asti yliopistojen ovat raportoineet Opetus- ja kulttuuriministeriön (OKM) vuosittaisessa tiedonkeruussa ainoastaan lukumäärätiedot julkaisuista. Vuodesta 2011 alkaen yliopistojen on tullut ilmoittaa kustakin julkaisusta myös metatiedot (mm. julkaisun nimi, tekijä, sarjan/emojulkaisun nimi, julkaisuvuosi, julkaisutyyppi, tieteenala jne.). Ammattikorkeakoulujen tulee toimittaa julkaisujensa vastaavat tiedot julkaisuista vuodesta 2012 alkaen.

Tässä selvityksessä kuvataan yliopistojen vuoden 2011 julkaisutiedonkeruussa raportoitujen tietojen laatua ja kattavuutta. Selvityksessä on tarkoitus

- 1) arvioida yliopistoilta kerättävien julkaisutietojen oikeellisuutta ja luotettavuutta
- 2) selvittää julkaisutietojen kirjaamiseen ja luokitteluun liittyviä ongelmakohtia
- 3) tarkentaa yliopistoille annettavaa ohjeistusta julkaisutietojen keruuseen

Samankaltainen selvitys tehtiin syksyllä 2011 vuoden 2010 julkaisutiedoista (JURE I -työryhmän raportti <http://raketti.csc.fi/paattyneet/jure/tyoryhmat/JURE1-tyoryhma>), mutta tuolloin käytettävissä oli ainoastaan kahdeksan yliopiston julkaisutiedot. Tässä raportissa tarkastellaan ensimmäistä kertaa kaikkien suomalaisten yliopistojen julkaisuja siten, että tiedot on kerätty osana valtakunnallista tiedonkeruuta yhdenmukaisten ohjeistuksen mukaisesti.

2 Tiedonkeruun eteneminen

2.1 Tietojen siirto

Yliopistot toimittavat julkaisutiedot vuosittain Opetus- ja kulttuuriministeriöön. Ohjeet kerättävistä sisältökentistä sekä tietojen muodosta on koottu OKM:n tiedonkeruukäsikirjaan, joka sisältää tarkat käsitelmääritellyt niille tiedoille, jotka yliopistoilta edellytetään OKM:n valtakunnallisessa tiedonkeruussa.

Kukin yliopisto toimitti vuoden 2011 julkaisujensa tiedot OKM:ään maaliskuun 2012 puolivälissä sähköpostitse Excel-tiedostoina. Alun perin tarkoitus oli, että yliopistot siirtäisivät tiedot suoraan Kotajärjestelmän kautta muiden OKM:n tiedonkeruiden tapaan, mutta julkaisutietojen lataus tietovarastoon ja automaattinen virheentarkistus ei ollut tiedonkeruun ajankohtaan mennessä vielä valmis. Yliopistoilta saaduissa julkaisutiedoissa oli useita muodollisia virheitä (puuttuvia tietoja, ei-koodistojen mukaisia arvoja). Muodollisista virheistä huolimatta julkaisutietoja pystyttiin käyttämään yliopistojen vuoden 2013 rahoitusta määriteltäessä uuden rahoitusmallin mukaisesti. Rahoituslaskelmissa käytettyjä tietoja käsiteltiin siten, että niistä poistettiin yhden yliopiston sisäiset duplikaatit (yliopiston aineistossa mainittu kaksi kertaa sama julkaisu) sekä yliopistojen välisten yhteisjulkaisujen julkaisutyypit yhdenmukaistettiin.

OKM:lle maaliskuussa toimitetuissa julkaisutiedoissa oli kuitenkin sen verran muodollisia virheitä, että ne lähetettiin yliopistoille vielä tarkistettaviksi. Lisäksi vaadittavien viitetietojen määrää lievennettiin hieman tiedonkeruukäsikirjan vaatimuksista. Lopulliset tiedonkeruun muodolliset tarkistukset on kuvattu taulukossa 1. Yliopistoja pyydettiin esimerkiksi ilmoittamaan, jos teknisessä tarkastuksessa jokin julkaisu on merkitty perusteettomasti sisäiseksi duplikaatiksi. Lisäksi yliopistoja pyydettiin lisäämään niille julkaisuille julkaisufoorumiluokitus, joilta se on tiedossa.

Yliopistot lähettivät tarkistetut tiedot vielä OKM:ön joka latasi tiedot yliopistojen puolesta Kotajärjestelmän kautta Vipunen tietovarastoon kesäkuussa. Järjestelmän automaattinen tarkistus tuottaa virhelistan, joka on yliopistojen nähtävissä ja jota ne voivat vielä korjata. Kun tiedot on syötetty Vipunen-tietokantaan, niihin lisätään julkaisufoorumiluokitukset. Lopullinen julkaisujen ja julkaisufoorumiluokistusten yhdistäminen tehdään keskitetysti CSC:ssä.

Taulukko 1. Julkaisutiedoille tehtävät tarkistukset.

Kaikille julkaisutyypeille A-E pakolliset tiedot:	
Julkaisutyyppi	Julkaisutyyppin täytyy löytyä tiedonkeruukäsikirjan julkaisutyyppiluokittelusta.
Julkaisun tieteenala	Tieteenalojen täytyy olla tieteenalaluokitus 2010 mukaisia.
Julkaisun koulutusala	Koulutusalojen täytyy olla yliopiston koulutusvastuun mukaisia olemassa olevia koulutusaloja.
Organisaation tekijät	Organisaation tekijöiden täytyy olla ilmoitettuna käsikirjan ohjeen mukaisesti.
Julkaisun tekijät	Julkaisun tekijöiden täytyy olla ilmoitettuna käsikirjan ohjeen mukaisesti.
Julkaisun tekijöiden lukumäärä	Ilmoitetaan, jos tekijöitä on yli 20.
Kansainvälinen	Arvon täytyy olla koodiston mukainen; 0 tai 1.

yhteisjulkaisu	
Julkaisun nimi	Kentässä täytyy olla jotain.
Julkaisuvuosi	Julkaisuvuoden täytyy olla 2011.
Julkaisun kieli	Tieto ei ole pakollinen, mutta jos tieto on merkitty niin sen täytyy olla Tilastokeskuk- sen "kielet 2003" -koodiston mukainen.
ISSN	Mikäli tieto ilmoitettu, koodin täytyy olla tiedonkeruukäsikirjan ohjeistuksen mukai- nen, "2 kertaa 4 merkkiä väliviivalla ilman ylimääräisiä merkkejä, esim. 0090-8258".
Julkaisumaa	Julkaisumaan täytyy olla Tilastokeskuksen koodiston mukainen.
Avoin saatavuus	Arvon täytyy olla koodiston mukainen; 0, 1, 2 tai 9.
Aikakauslehtiartikkeleille pakolliset tiedot (julkaisutyypit A1, A2 ja B1)	
Lehden / sarjan nimi	Kentässä täytyy olla jotain, <u>tai</u>
ISSN	Koodin täytyy olla tiedonkeruukäsikirjan ohjeistuksen mukainen, "2 kertaa 4 merkkiä väliviivalla ilman ylimääräisiä merkkejä, esim. 0090-8258".
Kokoomateosartikkeleille pakolliset tiedot (julkaisutyypit A3 ja B2)	
ISBN	Kentässä täytyy olla jotain, mikäli mahdollista ISBN-koodi merkitään väliviivojen kans- sa, <u>tai</u>
Kustantaja	Kentässä täytyy olla jotain, julkaisuforumilistan valmistuttua kustantajan nimen voi tarkistaa keskitetysti ylläpidetystä listasta, mikäli kustantaja on sisällytetty julkaisu- foorumin listalle.
Konferenssiartikkeleille pakolliset tiedot (julkaisutyypit A4 ja B3)	
Lehden / sarjan nimi	Kentässä täytyy olla jotain, kenttään syötetään konferenssijulkaisujen kohdalla konfe- renssin nimi, esim. International Symposium on Symbolic and Algebraic Computation, <u>tai</u>
ISSN	Koodin täytyy olla tiedonkeruukäsikirjan ohjeistuksen mukainen, "2 kertaa 4 merkkiä väliviivalla ilman ylimääräisiä merkkejä, esim. 0090-8258", <u>tai</u>
Emojulkaisun nimi	Kentässä täytyy olla jotain.
Monografioille pakolliset tiedot (julkaisutyyppi C1)	
ISBN	Kentässä täytyy olla jotain, <u>tai</u>
Kustantaja	Kentässä täytyy olla jotain, julkaisuforumilistan valmistuttua kustantajan nimen voi tarkistaa keskitetysti ylläpidetystä listasta, mikäli kustantaja on sisällytetty julkaisu- foorumin listalle.
Toimitetuille teoksille pakolliset tiedot (julkaisutyyppi C2)	
Lehden / sarjan nimi	Kentässä täytyy olla jotain, <u>tai</u>
ISSN	Koodin täytyy olla tiedonkeruukäsikirjan ohjeistuksen mukainen, "2 kertaa 4 merkkiä väliviivalla ilman ylimääräisiä merkkejä, esim. 0090-8258", <u>tai</u>
ISBN	Kentässä täytyy olla jotain, <u>tai</u>
Kustantaja	Kentässä täytyy olla jotain, julkaisuforumilistan valmistuttua kustantajan nimen voi tarkistaa keskitetysti ylläpidetystä listasta, mikäli kustantaja on sisällytetty julkaisu- foorumin listalle.

Taulukko 2. Vuoden 2011 julkaisutiedonkeruussa ilmoitettujen julkaisujen määrä yliopistoittain. Vertailuna vuoden 2010 tiedonkeruussa ilmoitetut julkaisumäärät. (huom. vuoden 2011 julkaisumäärät perustuvat yliopistojen maaliskuussa ilmoittamiin tarkistamattomiin tietoihin).

	2011		2010	
	julkaisujen lkm*	% kaikkien yliopistojen julkaisuista	julkaisujen lkm*	% kaikkien yliopistojen julkaisuista
Aalto-yliopisto	3 663	10,5 %	3 579	10,1 %
Helsingin yliopisto	10 184	29,3 %	10 508	29,6 %
Itä-Suomen yliopisto	2 870	8,3 %	2 714	7,6 %
Jyväskylän yliopisto	2 798	8,1 %	2 571	7,2 %
Kuvataideakatemia			38	0,1 %
Lapin yliopisto	560	1,6 %	538	1,5 %
Lappeenrannan teknillinen yliopisto	934	2,7 %	784	2,2 %
Oulun yliopisto	2 836	8,2 %	3 012	8,5 %
Sibelius-Akatemia	97	0,3 %	134	0,4 %
Svenska handelshogskolan	267	0,8 %	317	0,9 %
Tampereen teknillinen yliopisto	1 797	5,2 %	1 815	5,1 %
Tampereen yliopisto	2 464	7,1 %	2 213	6,2 %
Teatterikorkeakoulu	89	0,3 %	63	0,2 %
Turun yliopisto	4 020	11,6 %	5 090	14,3 %
Vaasan yliopisto	564	1,6 %	477	1,3 %
Åbo Akademi	1 579	4,5 %	1 647	4,6 %
Kaikki yhteensä	34 722	29,3 %	35 500	100,0 %

*) OKM:n julkaisutyypiluokat A-E

a) Yliopistojen väliset yhteisjulkaisut yhteissummassa useaan kertaan

2.2 Tietojen tarkastus ja korjaus

2.2.1 Viitetietojen oikeellisuus

Käsin tehty tarkastus osoitti, että useista yliopistojen maaliskuussa ministeriöön lähettämistä julkaisuista joko puuttui tietoja tai ne olivat virheellisiä. Puutteet koskivat koulutus- ja tieteenalojen merkintää sekä julkaisujen metatietoja. Taulukossa 3 on puutteellisten ja virheellisten julkaisutietojen määrä.

Taulukko 3. Puutteellisia tai virheellisiä tietoja sisältävien julkaisujen määrä.

Virhetyyppi	Julkaisujen määrä	% kaikista julkaisuista	Yliopistojen määrä, joilla puutteellisia tietoja
Tieteenala ei koodinmukainen	601	2,1 %	7
Koulutusala ei koulutusvastuun mukainen	379	1,3 %	1
Koulutusala tieto puuttuu	3	0,01 %	2

ISSN-koodi vääränmuotoinen	48	0,2 %	8
Julkaisuvuosi muu kuin 2011	156	0,5 %	5
Julkaisumaa puuttuu tai ei koodiston mukainen	886	3,1 %	14
Open access -tieto puuttuu	2 805	9,8 %	2*
Julkaisuja yhteensä (A-C) ^a	28 515	100 %	15

a) Yliopistojen väliset yhteisjulkaisut yhteissummassa useaan kertaan

*) lisäksi kolmessa yliopistossa tieto puuttui yhdeltä julkaisulta

Lukumääräisesti useimmin puuttui tieto julkaisumaasta sekä siitä, onko kyseessä open access -julkaisu. Suurin osa yliopistoista oli kuitenkin kirjannut open access -tiedot kaikille julkaisuilleen, sillä se puuttui vain kahden yliopiston julkaisuilta. Sen sijaan puuttuvia tai virheellisiä julkaisumaatietoja oli yhtä lukuun ottamatta kaikkien yliopistojen julkaisutiedoissa. Tarkastusvaiheessa julkaisu merkittiin kansainväliseksi, mikäli sen julkaisumaa on merkitty ja se on muu kuin Suomi. Mikäli julkaisumaata ei oltu merkitty, julkaisun kansainvälisyys määritettiin julkaisun kustannuspaikan perusteella (koskee julkaisutyyppiä A3, B2, C1, C2, D2, D5, E2). Mikäli julkaisun kansainvälisyyttä ei ole ollut mahdollista todentaa kummallakaan tavalla, julkaisu määritettiin kotimaiseksi.

Selkeästi vääränmuotoisia ISSN-koodeja esiintyi kahdeksan yliopiston julkaisutiedoissa, mutta niiden lukumäärä oli vähäinen, sillä niitä oli vain 48 julkaisussa. Lisäksi useissa julkaisuissa ISSN-koodi oli merkitty hiukan väärin, esim. ISSN-koodin väliviivan jälkeen koodissa oli turha välilyönti tai koodin alkuun oli lisätty merkkijono "ISSN". Vääränmuotoisia tieteenalakodeja sen sijaan oli selvästi useampia: yhteensä 601 julkaisussa. Useimmat väärät tieteenalakoodit olivat vuoden 2007 tieteenalakodeja.

Aineistossa oli 156 sellaista julkaisua, joissa julkaisuvuosi oli muu kuin 2011. Tarkastusvaiheessa poistettiin ne julkaisut, joiden julkaisuvuosi ei ollut 2011.

Taulukko 4. Tyypillisimmät virheet ja puutteet eri julkaisutietokentissä.

Kenttä	
Organisaatio	Koska kukin yliopisto toimitti erillisenä tiedostona oman aineistonsa, organisaation määrittäminen ei tuottanut ongelmia.
Julkaisutyyppi	Kaikki yliopistot käyttivät opetus- ja kulttuuriministeriön julkaisutyyppilistan koodeja. Julkaisu-tyypiluokituksissa oli kuitenkin ristiriitaisia tulkintoja (ks. Taulukko 4).
Tieteenala	Vuoden 2011 julkaisujen tieteenalat (1-6 kpl) oli pyydetty ilmoittamaan Tilastokeskuksen vuoden 2010 tieteenalaluokituksen koodien mukaisesti, ja tätä ohjetta oli noudatettu hyvin.
Julkaisun nimi	Julkaisujen nimissä esiintyi runsaasti vaihtelua pien- ja suuraakkosten käytössä. Länsimaisesta perusaakkostosta poikkeavat merkit (esim. kyrilliset kirjaimet, kiinalaiset merkistöt, matemaattiset symbolit) aiheuttavat ongelmia, kun tiedostoja siirretään järjestelmästä toiseen. Julkaisuista, joilla on pitkä nimi, oli joskus kirjattu vain nimen alkuosa. Joskus julkaisun nimen lisäksi kenttään oli kirjattu myös lehden nimi.
Tekijöiden nimet (oma organisaatio, kaikki	Osaan julkaisuista oli kirjattu vain oman organisaation tekijät, vaikka tämän lisäksi oli pyydetty myös täydellistä tekijälistaa. Jos sama julkaisu esiintyi useamman yliopiston listoilla, tekijöiden järjestys saattoi vaihdella täydellisessä tekijälistassa. Tekijöiden nimissä esiintyi jonkin verran

tekijät)	kirjoitusvirheitä. Sisäisten duplikaattien perusteella oman organisaation tekijöiden ilmoittaminen oli hyvin vaihtelevaa. Useissa sisäisissä duplikaateissa oman organisaation tekijät oli ilmoitettu täysin eri tavalla.
Tekijöiden lukumäärä	Yleensä ilmoitettu, mutta yhteisjulkaisujen ja sisäisten duplikaattien tietoja vertailemalla oli nähtävissä, että tekijöiden lukumäärää ei aina ollut kirjattu oikein. Kentässä paljon myös merkitään "0" (nolla).
Kansainvälinen yhteisjulkaisu	Tiedonkeruuhjeissa "kansainvälinen yhteisjulkaisu" oli määritelty julkaisuksi, jonka tekijöistä vähintään yksi on muun kuin suomalaisen organisaation palveluksessa oleva henkilö, tai jos joku tekijöistä on sekä suomalaisen että ulkomaisen organisaation palveluksessa, hän on merkinnyt julkaisuun affiliaatiokseen myös ulkomaisen organisaation. Tästä huolimatta pistokokein löytyi helposti julkaisuja, jotka oli ilmoitettu kansainvälisiksi yhteisjulkaisuksi, vaikka niiden alkuperäisissä tekijätiedoissa esiintyi vain yksi tai useampia suomalaisia organisaatioita.
Julkaisun kieli	Julkaisukielen kirjaamisessa oli noudatettu tunnollisesti Tilastokeskuksen kaksimerkkistä kielikoodia. Tieto julkaisukielestä puuttui vain yksittäisiltä julkaisuilta, samoin joissain julkaisuissa oli selkeästi väärä kielikoodi.
Lehden (sarjan) nimi	Lehtien niminä oli käytetty enimmäkseen lehtien täydellisiä nimiä, mutta myös jonkin verran lyhenteitä. Joskus nimistä puuttui sanoja (esim. englannin kielen määräinen artikkeli "the") tai niihin oli lisätty ylimääräisiä sanoja (vuosiluku tai määrite "Online"). Joissain tapauksissa oli virheellisesti kirjattu emojulkaisun nimi lehden nimeksi.
Lehden (sarjan) ISSN	ISSN-tunnusta ei läheskään aina ollut ilmoitettu, vaikka se olisi ollut saatavilla. Tyypillisiä ongelmia olivat mm.: tunnus oli sekoitettu ISBN-tunnukseen, alkuun oli lisätty merkkijono "ISSN" tai tunnuksessa oli kirjoitusvirhe (ylimääräisiä, puuttuvia tai vaihtuneita merkkejä). Osa virheistä johtui muutaman lehden käyttämästä virheellisestä ISSN-tunnuksesta.
Lehden (sarjan) osa ja numero (volume, issue)	Lehden osa ja numero oli jätetty usein ilmoittamatta. Numero oli toisinaan sekoitettu joidenkin lehtien käyttämään artikkelikohtaiseen numeroon. Useamman yliopiston kirjaamien yhteisjulkaisujen tietoja vertailemalla näkyy, että osa ja numero kirjautuvat usein väriin kenttiin.
Sivunumerot	Sivunumerot oli ilmoitettu kohtalaisen tunnollisesti niiden julkaisutyypin osalta, joista sivunumeroita oli pyydetty. Formaatti vaihteli paljon. Alku- ja loppusivujen erottamiseen oli käytetty yhtä tai kahta väliviivaa. Toisinaan numeroiden edessä käytettiin lyhenteitä "s.", "ss.", "p." tai "pp.". Sivunumerokenttään oli kirjattu myös joissain tapauksissa artikkelinnumero, emojulkaisun nimi tai ISSN. Ilmeisesti kirjoitustyön säästämiseksi esiintyi käytäntöä, jossa loppusivun numerosta jätetään pois se osa, joka toistuu samana alkusivun numerossa; siis sivut 123-125 merkitään lyhyemmin 123-5. Jonkun verran oli kirjattu pelkästään sivumääriä alku- ja loppusivujen sijasta, myös sellaisissa tapauksissa, joissa sivunumerot olisivat olleet olemassa.
Artikkelinnumero	Artikkelinumerolle oli varattu oma kenttä, mutta siitä huolimatta tämä tieto oli hyvin usein sijoitettu lehden numerolle tai sivunumerolla tarkoitettuun kenttään.
Vuosi	Julkaisuvuodeksi oli ohjeissa määritelty se vuosi, jolloin julkaisu ilmestyy ensimmäisen kerran muodossa, jossa sillä on täydelliset viitetiedot. Yksittäisissä pistokokeissa osoittautui, että yliopistot olivat ilmoittaneet myös julkaisuja, joiden ilmestymisvuosi tämän määritelmän mukaan olisi 2010 tai 2012.
ISBN	ISBN-tunnus oli joskus sekoitettu ISSN-tunnukseen. Käytäntö oli vaihteleva sen suhteen, merkitäänkö tunnus väliviivoin vai pelkästään numero- (ja kirjain-)merkein. Samalla julkaisulla voi olla useita ISBN-numeroita (esim. kova- ja pehmeäkantisella).
Kustantaja	Eriyisesti sarjoilta kustantaja oli usein jätetty ilmoittamatta. Kirjoitusasut vaihtelivat huomattavasti, mikä vaikeuttaa julkaisufoorumiluokituksen kirjaamista.

Pysyvä verkko-osoite	Osa yliopistoista oli kirjannut pysyväksi verkko-osoitteeksi vain merkkijonon "http://dx.doi.org/", joka ei sellaisenaan vielä ole täydellinen osoite julkaisulle.
EVO-julkaisu	EVO-julkaisuja voivat olla vain terveys- ja lääketieteen alan julkaisut. Tästä huolimatta EVO-julkaisuksi oli kirjattu monien muidenkin tieteenalojen julkaisuja.

2.2.2 Yliopistojen sisäiset duplikaatit

Aineistoista löytyi yhteensä 163 yliopiston sisäistä duplikaattia, eli sellaista julkaisua, joka on tallennettu kahteen kertaan yhden yliopiston julkaisurekisteriin. Tämä on 1,2 % kaikista yliopistojen ilmoittamista julkaisuista. Yliopistoittain osuus vaihteli 0-6 prosentin välillä. Duplikaateista 27 oli sellaisia, joissa sama julkaisu oli kirjattu kahteen kertaan, mutta niille oli kirjattu eri julkaisutyyppejä.

Duplikaattien tunnistamiseksi vuoden 2010 pilottiaineistolla kehitettiin laskenta-algoritmi. Ohjeesta huolimatta aineisto tuli kuitenkin eri yliopistoista eri muodossa esimerkiksi siten, että osa oli jättänyt jonkun kentän pois ja osa oli lisännyt joitakin kenttiä. Siksi tunnistusalgoritmia ei sellaisenaan voinut käyttää ja duplikaatit tunnistettiin manuaalisesti. Tunnistetut duplikaatit lähetettiin tarkistettaviksi yliopistoille.

2.2.3 Ristiriitaiset julkaisutyypiluokitukset

Kukin yliopisto kirjaa julkaisulle yhden julkaisutyypin. Sama julkaisu voi kuitenkin esiintyä useassa yliopistossa siten, että eri yliopistojen kirjaamat julkaisutyypiluokitukset poikkeavat toisistaan. Mikäli kahden tai useamman yliopiston välinen sama julkaisu oli merkitty eri yliopistojen toimesta eri julkaisutyypille, on julkaisujen julkaisutyypit yhtenäistetty. Yhtenäistäminen tapahtui siten, että julkaisutyypiksi on merkitty aakkosista ensin oleva ja numerona pienempi (esim. B2->A3 ja A2->A1).

Koko aineistossa oli 293 sellaista vähintään kahden yliopiston välistä julkaisua, joille oli annettu eri julkaisutyypiluokka eri yliopistoissa (Taulukko 5). Näistä kahdeksan julkaisua oli kirjattu kolmeen eri julkaisutyypiin. Tyypillisimmät tapaukset (yhteensä 101 kpl) olivat artikkeleita, jotka oli luokiteltu toisessa yliopistossa vertaisarvioituksi ja toisessa vertaisarvioimattomaksi artikkeliksi (A/B). 56 tieteellisessä lehdessä julkaistua artikkelia oli kirjattu sekä alkuperäis- että katsausartikkeliksi (A1/A2). Sellaisia julkaisuja, jotka oli toisessa yliopistossa kirjattu tieteelliseksi (A-C) ja toisessa ammatti- tai suurelle yleisölle suunnatuksi (C-D) oli yhteensä 68 kappaletta.

Taulukko 5. Yliopistojen väliset yhteisjulkaisut, jotka on luokiteltu eri julkaisutyypiluokkiin.

	A2	A3	A4	B1	B2	B3	C1	C2	D1	D2	D3	D4	D5	E1	E2	Yht.
A1 Alkuperäisartikkeli tiet. aikakauslehdessä	56	4	12	18				1	11							102
A2 Katsausartikkeli tiet. aikakauslehdessä		1		13					4							18
A3 Kirjan t. muun kokoomateoksen osa			7		39	2	1	2		5				1		57

A4 Artikkelit konferenssijulkaisussa	1	28	1	1	1	1	1	33								
B1 Kirjoitus tiet. aikakauslehdessä	1	12	1	5	19											
B2 Kirjan t. muun kokoomateoksen osa	2	1	12	1	2	3	21									
B3 Vertaisarvioimaton artikkeli konferenssijulkaisussa				2	2											
C1 Kustannettu tiet. erillisteos		5	5	1	11											
C2 Toimitettu kirja, kokoomateos, konferenssijulkaisu t. lehden erikoisnumero				3	5	8										
D1 Artikkelit ammattilehdessä				2	12	14										
D2 Artikkelit ammatillisessa käsi- t. opaskirjassa, ammatillisessa tietojärjestelmässä t. oppikirja-aineisto					2	1	3									
D3 Artikkelit ammatillisessa konferenssijulkaisussa						0										
D4 Julkaistu kehittämis- t. tutkimusraportti taikka selvitys						2	2									
D5 Oppikirja, ammatillinen käsi- t. opaskirja taikka sanakirja						1	1									
E1 Yleistajuinen artikkeli, sanomalehtiartikkeli						1	1									
E2 Yleistajuinen monografia					1	1										
Yhteensä	56	5	19	31	40	33	1	10	27	17	3	13	6	23	9	293

Lisäksi aineistossa oli useita sellaisia julkaisuja, jotka oli luokiteltu eri julkaisutyypiluokkaan kuin toiset samassa lehdessä tai sarjassa julkaistut artikkelit. Tarkastelu sellaisista lehdistä, joiden ISSN-tunnus on kirjattu, osoittaa, että useissa lehdissä tai sarjoissa julkaistuja artikkeleita on luokiteltu erityyppisiksi sekä lehti- (A1, A2, B1), kokoomateos- (A3, B2) että konferenssiartikkeleiksi (A4, B3) sekä muutamissa tapauksissa jopa monografioiksi (C1). Lisäksi useissa lehdissä julkaistuja artikkeleita on luokiteltu sekä tieteellisiksi (A-B) että ei-tieteellisiksi (D-E). Aineistossa esimerkiksi on 1 809 sellaista julkaisua, jotka on luokiteltu julkaisuksi tieteellisellä julkaisukanavalla, vaikka toisissa julkaisuissa kyseinen julkaisukanava on tulkittu ammatillisiksi tai yleistajuisiksi. Tämä on noin kuusi prosenttia kaikista ilmoitetuista tieteelli-

sistä julkaisuista. Eräissä kotimaisessa lehdessä julkaistuja artikkeleita on merkitty sekä lehti-, konferenssi-, kokoomateosartikkeleiksi, monografioiksi, toimitetuiksi teoksiksi että tutkimusraporteiksi.

Taulukko 6. Samoissa lehdissä tai sarjoissa julkaistujen eri julkaisutyyppiluokkiin luokiteltujen julkaisujen lukumäärä.

	Tieteellinen lehti-/konferenssi-artikkeli	Tieteellinen lehti-/kokoomateos-artikkeli	Kokoomateos- / konferenssi-artikkeli	Tieteellinen /ei-tieteellinen
Lehtiä tai sarjoja	55	46	36	218
Julkaisujen määrä näissä lehdissä tai sarjoissa	240/203	494/171	161/222	1809/952

2.3 Julkaisun tieteenala

Kaikki yliopistot olivat ilmoittaneet kaikille julkaisuilleen vähintään yhden tieteenalan (Tilastokeskuksen tieteenalaluokitus). Julkaisulle valitaan tieteenala eri kriteerein eri yliopistoissa. Osa yliopistoista tekee luokituksen julkaisun tieteenalan perusteella, osa julkaisun tekijän edustaman tieteenalan mukaan ja osassa tieteenalaksi on määritelty suoraan tekijöiden laitosten edustamat tieteenalat. Osa yliopistoista tekee luokituksen julkaisun tieteenalan perusteella, osa julkaisun tekijän edustaman tieteenalan mukaan ja osassa tieteenalaksi on määritelty suoraan tekijöiden laitosten edustamat tieteenalat. Julkaisuista 72 prosentille on merkitty vain yksi tieteenala, mutta osuus vaihtelee yliopistojen välillä: Aalto 18%, jne.

Sisäisten duplikaattien perusteella havaittiin, että samaa julkaisua oli merkitty eri tieteenaloille yhden yliopiston sisällä.

2.4 Julkaisuforumiluokituksen yhdistäminen julkaisuihin

Julkaisufoorumi-hankkeen (2010–2012) toteuttama, vuonna 2011 valmistunut tasoluokitus kattaa yli 20 000 tieteellistä julkaisukanavaa: n. 19 500 lehteä ja sarjaa sekä n. 1 200 kirjakustantajaa. Luokituksen on tarkoitus toimia yliopistojen tieteellisen julkaisutuotannon laadun indikaattorina opetus- ja kulttuuriministeriön käyttämässä yliopistojen valtiorahoitusmallissa vuodesta 2015 lähtien. Tämä edellyttää, että Julkaisufoorumi-luokitus voidaan liittää luotettavasti yliopistojen julkaisutietoihin.

Lehti- ja sarjaluokituksen tai kirjakustantajaluokituksen soveltaminen tieteellisten julkaisujen julkaisufoorumi-luokan määrittämiseen riippuu julkaisutyyppistä. Pääsääntö on, että lehti- ja konferenssiartikkelien tasoluokan määrittää lehti- ja sarjaluettelossa oleva lehden tai konferenssin Julkaisufoorumi-luokka. Sama koskee toimitettuja lehden erikoisnumeroita tai konferenssijulkaisuja. Kirjakustantajaluokitusta käytetään ainoastaan määrittämään kokoomateosten ja erillisteosten, tai toimitettujen kirjojen tasoluokka. Lehti- ja sarjaluettelossa on luokiteltu myös yksittäisiä kirjasarjoja. Jos lehti- ja sarjaluettelossa luokitellun kirjasarjan ja kirjakustantajan tasoluokka poikkeavat toisistaan, voidaan kokoomateosartikke-

lin, erillisteoksen tai toimitetun kirjan Julkaisufoorumi-luokka määrittää korkeamman tasoluokan mukaan.

Taulukko 7. Julkaisufoorumi-luokituksen soveltaminen julkaisutyypeittäin.

Julkaisutyyppi (tieteelliset julkaisut A-C)	Lehti- ja sarja-luettelo	Kirjakustantaja-luettelo
A1 Alkuperäisartikkeli tieteellisessä aikakauslehdessä	X	
A2 Katsausartikkeli tieteellisessä aikakauslehdessä	X	
A3 Kirjan tai muun kokoomateoksen osa	X	X
A4 Artikkelit konferenssijulkaisussa	X	
B1 Kirjoitus tieteellisessä aikakauslehdessä	X	
B2 Kirjan tai muun kokoomateoksen osa	X	X
B3 Vertaisarvioimatonta artikkelit konferenssijulkaisussa	X	
C1 Kustannettu tieteellinen erillisteos	X	X
C2 Toimitettu kirja, kokoomateos, konferenssijulkaisu tai lehden erikoisnumero	X	X

Tärkeimpiä Julkaisufoorumi-luokan määrittämisen mahdollistavia tunnistajia ovat lehtien ja sarjojen osalta nimeke ja ISSN-tunnus, ja kirjakustantajien osalta nimi ja ISBN-tunnus (ISBN-tunnuksen alkuosa muodostaa kustantajakoodin). Koska käsitellään suuria julkaisumääriä, on pyrittävä käyttämään mekaanista tunnistamista ja välttämään manuaalista tunnistamista. Mekaaninen tunnistaminen voidaan suorittaa ensisijaisesti ISSN- ja ISBN-tunnusten perusteella, jotka virheettömästi kirjattuina ovat yksiselitteisiä tunnistetietoja. Jos ISSN- tai ISBN-tunnus puuttuu, tai on merkitty virheellisesti, voidaan käyttää lehden, sarjan, konferenssin tai kirjakustantajan nimeä. Nimi on kuitenkin mahdollista kirjata julkaisutietoihin hieman eri tavoin, mikä vaikeuttaa mekaanista tunnistamista ja voi edellyttää manuaalisia tarkistuksia.

Taulukko 8. Julkaisufoorumitunnisteiden puuttuminen julkaisutyypeittäin (perustuu yliopistojen maaliskuussa ilmoittamiin tietoihin)

Julkaisutyyppi:	lkm	% kaikista julkaisuista	yo-lkm
Lehtiartikkelit (julkaisutyyppit A1, A2, B1) <i>Sekä lehden/sarjan nimi että ISSN-koodi puuttuvat</i>	6	0,02 %	5
Kokoomateosartikkelit (julkaisutyyppit A3, B2) <i>Sekä kirjakustantajan nimi että ISBN-koodi puuttuvat</i>	45	0,2 %	8
Konferenssiartikkelit (julkaisutyyppit A4, B3) <i>Sekä lehden/sarjan nimi, ISSN-koodi että emojulkaisun nimi puuttuvat</i>	405	1,4 %	3
Erillisteokset (C1) <i>Sekä kirjakustantajan nimi että ISBN-koodi puuttuvat</i>	4	0,01 %	3
Toimitetut teokset (C2) <i>Sekä Lehden/sarjan nimi että ISSN-tunnus tai sekä</i>	4	0,01 %	3

kirjakustantajan nimi että ISBN-koodi puuttuvat

Parhaimmat edellytykset yliopistojen vuoden 2011 julkaisutiedot tarjoavat tieteellisten lehtiartikkelien (julkaisutyypit A1, A2 ja B1) Julkaisufoorumi-luokan tunnistamiseen. Sekä ISSN-tunnus että nimeke puuttuu vain 6 julkaisun osalta. Vertaisarvioituista artikkeleista 95 %, ja vertaisarvioimattomista artikkeleista 91 %, sisältää ISSN-tunnuksen, minkä lisäksi käytännössä kaikista julkaisuista on mainittu lehden tai sarjan nimeke. Lehden/sarjan nimekettä joudutaan käyttämään Julkaisufoorumi-luokan tunnistena 939 lehtiartikkelin kohdalla (5,6 % julkaisuista).

Kokoomateosartikkelien ja erillisteosten (julkaisutyypit A3, B2 ja C1) osalta vuoden 2011 julkaisuaineiston tunnistetiedot ovat myös erittäin kattavat. Sekä ISBN-tunnus että kirjakustantajan nimi puuttuu 49 julkaisun osalta (0,2 % kaikista julkaisuista). Tieteellisistä erillisteoksista 92 %, vertaisarvioituista kokoomateosartikkeleista 81 % ja vertaisarvioimattomista kokoomateosartikkeleista 86 % sisältää ISBN-tunnuksen. Kirjakustantajan nimeä, joka on mainittu 92-97 % julkaisuista, joudutaan käyttämään tunnistena kaikkiaan 841 erillisteoksen ja kokoomateosartikkelin osalta (16,5 % kaikista julkaisuista). Tunnistus voi osittain tapahtua myös kirjasarjan ISSN-tunnuksen perusteella, joka on mainittu 28 % tapauksista.

Ongelmallisimmaksi Julkaisufoorumi-luokan tunnistaminen muodostuu vuoden 2011 julkaisuaineistoon kuuluvien konferenssijulkaisujen osalta. Sekä ISSN-tunnus, lehden ja sarjan (=konferenssin) nimi, että emojulkaisun nimi puuttuu 405 konferenssijulkaisusta (1,5 % julkaisuista). Vain harvalla konferenssarjalla on ISSN-tunnus, mikä näkyy myös julkaisutiedoissa: vain 17 % vertaisarvioituista ja 14 % vertaisarvioimattomista konferenssijulkaisuista on merkitty ISSN-tunnus. Konferenssin nimikin on käytettävissä vain 47-49 % tapauksista. Lehden/sarjan nimeä joudutaan käyttämään tunnistena 4 851 konferenssiartikkelin osalta (83 % julkaisuista). Mekaanisen tunnistamisen helpottamiseksi onkin selvitettävä, voidaanko konferenssien nimien vakiintuneita kirjainlyhenteitä jatkossa käyttää ISSN-tunnuksen kaltaisena tunnistetietona.

Voidaan olettaa, että eri julkaisutyypin Julkaisufoorumi-luokan tunnistamista koskevat tunnistetiedot esiintyvät julkaisutiedoissa samalla tavalla myös toimitettujen lehtien erikoisnumeroiden, konferenssijulkaisujen, kokoomateosten ja kirjojen (julkaisutyyppi C2) kohdalla.

Taulukko9. Tunnistetietojen esiintyminen julkaisutiedoissa julkaisutyypeittäin

Julkaisutyyppi	Julkaisujen lkm	Sarja	ISSN	Kustantaja	ISBN	Ei tietoa
A	22 212	81 %	70 %	31 %	0 %	1 %
A1	14 029	100 %	95 %	14 %	ei	0 %
A2	726	100 %	94 %	14 %	ei	0 %
A3	2 997	37 %	28 %	97 %	81 %	0 %
A4	4 460	47 %	17 %	42 %	ei	5 %
B	5 121	64 %	51 %	47 %	0 %	3 %

B1	2 137	100 %	91 %	19 %	ei	0 %
B2	1 629	29 %	28 %	93 %	86 %	0 %
B3	1 355	49 %	14 %	35 %	ei	12 %
C	1 182	40 %	34 %	92 %	83 %	0 %
C1	459	39 %	28 %	95 %	92 %	0 %
C2	723	41 %	38 %	90 %	78 %	0 %
Yhteensä	28 515	66 %	63 %	37 %	3 %	1 %

2.5 Tietojen julkaiseminen

Julkaisujen lukumäärätiedot julkaistaan opetushallinnon uudessa raportointipalvelussa Vipusessa <http://vipunen.csc.fi> , joka sisältää muutakin tilasto- ja indikaattoritietoa korkeakouluista. Vipuseen tulee mahdollisesti myös sellainen ominaisuus, jonka avulla esimerkiksi jonkin yliopiston tai tieteenalan julkaisujen lukumäärää klikkaamalla saa julkaisujen viitetiedot näkyviin. Myöhemmin on lisäksi kehitteillä viitetietojen selailuun julkaisuportaali, jota Kansalliskirjaston pilotoi vuoden 2011 julkaisutiedoilla vuoden 2012 loppupuolella.

3 Yhteenveto ja suositukset

Tarkastelu osoittaa, että julkaisuissa on paljon yhteismitallisuutta ja vertailukelpoisuutta heikentäviä puutteita ja virheitä. Vuosien 2013–14 rahoitusmallin kannalta ongelmallisia ovat puutteet julkaisumaa- ja julkaisutyypiluokituksessa. Julkaisutyypiluokituksen oikeellisuus on tärkeää, sillä sen perusteella esimerkiksi määritetään, huomioidaanko julkaisu rahoitusmallissa. Ei-tieteellisillä julkaisukanavilla julkaistuja ei ole mallissa mukana. Julkaisumaatieto on tärkeä, sillä sen perusteella erotellaan kansainväliset referee-julkaisut kotimaisista.

Vuonna 2015 käyttöön otetaan julkaisuforumiluokitus, jolloin julkaisumaatiedolla ei ole enää rahoitusmallin kannalta merkitystä. Koska julkaisukanava määritetään auktorisoidun listan avulla, julkaisun tieteellisyyden määrittelemine ei ole enää ilmoitetun julkaisutyypin varassa. Olennaista tällöin kuitenkin on, että erityisesti julkaisukanavaa koskevat tiedot (ISSN/ISBN ja lehden/sarjan/kustantajan nimi) olisivat oikein julkaisun yhdistämiseksi oikeaan julkaisukanavaan. Tieteellisistä lehdissä julkaistuista artikkeleista valtaosa (yli 90 %) näyttäisi sisältävän ISSN-tiedon, minkä avulla lehti tai sarja useimmiten voidaan identifioida. Konferenssijulkaisujen ja kirjojen osalta tunnistaminen on kuitenkin huomattavasti ongelmallisempaa.

Kehittämiskohteita:

1. Opetusministeriön ohjeistus vaatii ehkä joitakin pieniä viilauksia (esim. julkaisutyypiluokitus, julkaisuvuosi)
2. Yliopistojen tarkistusprosesseissa parannettavaa
3. Ajantasainen yhteinen julkaisukanavalista, joka sisältäisi myös ns. 0-listan ei-tieteellisistä julkaisukanavista ja jota yliopistot voisivat käyttää omissa julkaisurekistereissään (mahd. syöttökäyttöliittymässä)

- parantaisi laatua ja yhdenmukaistaisi tietoja
- helpottaisi julkaisujen yhdistämistä julkaisukanaviin ja julkaisufoorumitasoihin
- yliopistot voisivat hyödyntää myös omassa seurannassaan
- yliopistojen julkaisutietoja ja niiden omia julkaisukanavalistoja voisi hyödyntää -> kattavampi julkaisukanavalista

Toimenpide-ehdotukset:

- Kahta ensimmäistä käsitellään OKM:n järjestämässä seminaarissa yliopistojen julkaisutiedonkerääjille
- Julkaisukanavalistan päivittämistä ja ylläpitoa pohtimaan perustetaan työryhmä, jossa edustus JUFOsta, yliopistoista, OKM:stä ja CSC:stä.

Liite 1 Opetusministeriön julkaisutyypiluokittelu 2010

A Vertaisarvioidut tieteelliset artikkelit

- A1 Alkuperäisartikkeli tieteellisessä aikauslehdessä (Journal article-refereed, Original research)
- A2 Katsausartikkeli tieteellisessä aikauslehdessä (Review article, Literature review, Systematic review)
- A3 Kirjan tai muun kokoomateoksen osa (Book section, Chapters in research books)
- A4 Artikkelit konferenssijulkaisussa (Conference proceedings)

B Vertaisarvioimattomat tieteelliset kirjoitukset

- B1 Kirjoitus tieteellisessä aikauslehdessä (Unrefereed journal articles)
- B2 Kirjan tai muun kokoomateoksen osa (Book section)
- B3 Vertaisarvioimaton artikkeli konferenssijulkaisussa (Unrefereed conference proceedings)

C Tieteelliset kirjat (monografiat)

- C1 Kustannettu tieteellinen erillisteos(Book)
- C2 Toimitettu kirja, kokoomateos, konferenssijulkaisu tai lehden erikoisnumero (Book (editor))

D Ammattiyhteisölle suunnatut julkaisut

- D1 Artikkelit ammattilehdessä
- D2 Artikkelit ammatillisessa käsi- tai opaskirjassa, ammatillisessa tietojärjestelmässä tai oppikirja-aineisto
- D3 Artikkelit ammatillisessa konferenssijulkaisussa
- D4 Julkaistu kehittämis- tai tutkimusraportti taikka -selvitys
- D5 Oppikirja, ammatillinen käsi- tai opaskirja taikka sanakirja

E Suurelle yleisölle suunnatut julkaisut

- E1 Yleistajuinen artikkeli, sanomalehtiartikkeli
- E2 Yleistajuinen monografia

F Julkinen taiteellinen ja taideteollinen toiminta

- F1 Julkaistu itsenäinen taiteellinen teos
- F2 Julkinen taiteellinen teoksen osatoteutus
- F3 Julkinen taiteellinen esitys tai näyttely
- F4 Tuotantoon/hyödyntämiseen otettu malli tai suunnitelma

G Opinnäytteet

- G1 Ammattikorkeakoulututkinnon opinnäytetyö, kandidaatintyö
- G2 Pro gradu, diplomityö, ylempi amk-opinnäytetyö
- G3 Lisensiaatintyö
- G4 Monografiaväitöskirja
- G5 Artikkeliväitöskirja

H Patentit ja keksintöilmoitukset

- H1 Myönnetty patentti
- H2 Keksintöilmoitus

I Audiovisuaaliset aineistot ja tieto- ja viestintätekniset ohjelmat

- I1 Audiovisuaalinen aineisto
- I2 Tieto- ja viestintätekniset ohjelma tai ohjelmisto

Liite 2 Tilastokeskuksen tieteenalaluokitus

1 Luonnontieteet	4 Maatalous- ja metsätieteet
111 Matematiikka	411 Maatalous- ja metsätieteet
112 Tilastotiede	412 Eläintiede, maitotaloustiede
113 Tietojenkäsittely ja informaatiotieteet	413 Eläinlääketiede
114 Fysiikka	414 Maatalouden bioteknologia
115 Avaruustieteet ja tähtitiede	415 Muut maataloustieteet
116 Kemia	
117 Maantiede ja ympäristötieteet	5 Yhteiskuntatieteet
118 Biotieteet	511 Kansantaloustiede
119 Muut luonnontieteet	512 Liiketaloustiede
	513 Oikeustiede
2 Tekniikka	514 Sosiaalitieteet
211 Arkkitehtuuri	515 Psykologia
212 Rakennus ja yhdyskuntatekniikka	516 Kasvatustieteet
213 Sähkötekniikka, elektroniikka, tietojenkäsittelytekniikka	517 Valtio-oppi, hallintotiede
214 Kone- ja valmistustekniikka	518 Media- ja viestintätieteet
215 Teknillinen kemia, kemian prosessitekniikka	519 Yhteiskuntamaantiede, talusmaantiede
216 Materiaalitekniikka	520 Muut yhteiskuntatieteet
217 Lääketieteen tekniikka	
218 Ympäristötekniikka	6 Humanistiset tieteet
219 Ympäristön bioteknologia	611 Filosofia
220 Teollinen bioteknologia	612 Kielitieteet, kirjallisuus
221 Nanoteknologia	613 Taiteiden tutkimus
222 Muu tekniikka	614 Teologia
	615 Historia ja arkeologia
3 Lääke- ja terveystieteet	616 Muut humanistiset tieteet
311 Peruslääketieteet	
312 Kliiniset lääketieteet	9 Muut
313 Hammaslääketieteet	999 Muut
314 Terveystieteet	
315 Liikuntatiede	
316 Hoitotiede	
317 Farmasia	
318 Lääketieteen bioteknologia	
319 Muut lääketieteet	

Liite 3. Yhteenveto yliopistojen tiedonkeruukyselyn vastauksista julkaisujen osalta

Tähän liitteeseen on koottu yhteenveto yliopistojen julkaisutiedonkeruuta koskevista vastauksista OKM:n lokakuussa 2011 toteuttamaan kyselyyn, jossa selvitettiin yliopistojen ja ammattikorkeakoulujen näkemyksiä valtakunnallisesta tiedonkeruusta. Kyselyyn vastasi 14 yliopistoa.

1. Määrittelyjen yksiselitteisyys

Kysymys: Onko vuoden 2010 käsikirjassa olevien tietojen keruu mielestänne kaikin osin yksiselitteinen? Jos ei ole, kuvataa yleisellä tasolla, mitkä määrittelyt aiheuttavat eniten tulkintaa teidän korkeakoulusanne.

Ongelmia julkaisutietojen määrittelyissä kommentoi 11 yliopistoa.

Julkaisutyypiluokkien tulkinta on aiheuttanut ongelmia ainakin viidessä yliopistossa. Määritelmistä ei esimerkiksi selviä, kirjataanko oman sarjan julkaisut luokkaan B2 vai D4. Ensyklopediatekstit arvostetuilla kustantajilla voisi erään yliopiston vastauksen mukaan kirjata mieluummin luokkaan A3 kuin D2. Julkaisutyypiluokitus ongelmallinen, jos tietoja halutaan vertailla yliopiston sisällä aikaisempien vuosien kanssa. Yhdessä vastauksessa kritsoitiin luokan D4 raporttien osakokonaisuuksien tekijöiden kirjaamisen erilaista logiikkaa suhteessa muihin julkaisutyyppeihin. Koska raportin osien tieteellisyys on myöskin tulkinnanvarainen, yhden yliopiston vastauksessa ehdotetaan erillistä luokkaa D6 näille julkaisuille. Erään yliopiston mukaan konferenssijulkaisuista on vaikea saada selville, ovatko ne vertaisarvioituja (A4) vai eivät (B3).

Kahden yliopiston vastauksessa todetaan, että kansainvälisyyden määrittäminen julkaisumaan perusteella on ongelmallinen Suomessa julkaistujen kansainvälisten julkaisujen osalta. Tämä koskee erityisesti Suomessa järjestettäviä konferensseja ja niiden Proceedings -julkaisuja. Eräässä yliopistossa kansainvälisyyden kriteerinä on pidetty, että vähintään 20 % konferenssin osallistujista on ulkomaalaisia ja näitä on vähintään kahdesta eri maasta.

Yhden yliopiston vastauksessa todetaan, että kansainvälisen yhteisjulkaisun määrittäminen on ongelmallista, kun esimerkiksi kansallinen henkilö muuttuu kansainväliseksi, jos tämä on käyttänyt lisäksi ainakin yhtä kansainvälistä affiliaatiota. Tällöin tietojen määrittelyssä on turvaututtava käsityöhön.

Yhden yliopiston vastauksessa nostetaan esille, että bibliografiset tiedot artikkeleista kokoomateoksissa (A3) ja konferenssijulkaisuissa (A4) ovat melko samanlaiset, mutta tiedonkeruuohjeissa ei ole konferenssijulkaisuille kenttiä "Emojulkaisun nimi" ja "Emojulkaisun toimittajat", vaan "Lehden / sarjan nimi". Epäselvää on, jätetäänkö konferenssijulkaisun toimittajat merkitsemättä.

Julkaisuvuoden määritelmä koettiin ainakin kolmessa yliopistossa ongelmalliseksi niiden julkaisujen osalta, jotka ovat kyseisen vuoden julkaisuja, mutta joista kirjoittaja saa tiedon vasta seuraavana vuonna, kun tiedonkeruu on jo ohi. *(Tämä kohta oli osittain korjattu käsikirjan lopulliseen versioon, mutta ei kerrota täsmälleen, mitä näiden julkaisujen kanssa tehdään).*

Eräässä yliopistossa oli koettu ongelmallisiksi muun muassa tiedonkeruukäsikirjan luonnosversiossa esitetyt Kielikeskuksen henkilöstön julkaisujen kirjaaminen kaikille koulutusaloille ja julkaisun kunkin tekijän merkitseminen omaan kenttäänsä. Joissakin kohdissa epäselvää oli myös, viitattiinko kansainvälisellä julkaisulla ulkomailta julkaistuun vai kansainväliseen yhteisjulkaisuun. Nämä kohdat on kuitenkin korjattu käsikirjan lopulliseen versioon. Niin ikään eräässä vastauksessa mainittu ristiriita tieteellisiltä julkaisuilta A-C edellytettävän vertaisarvioinnin ja luokan B määritelmän (vertaisarvioimaton) välillä oli korjattu lopulliseen versioon.

Yhden yliopiston vastauksessa kommentoitiin julkaisutiedonkeruuseen liittyviä muutoksia: tieteenalakoodin uusi muotoilu ilman pilkkua, tekijätietojen muoto sekä erilliset kentät julkaisujen kirjoittajien lukumäärälle sekä artikkelinumerolle. Nämä edellyttävät muutoksia järjestelmään ja tuottavat lisä kustannuksia yliopistolle.

Taidealan julkaisujen kohdalla "julkaisu"-termi sekoittuu varsinaisiin julkaisuihin. Epäselvää on myös, mihin sijoitetaan entisen luokituksen mukainen "F4 Tuotantoon/hyödyntämiseen otettu malli tai suunnitelma".

2. Tietojen kerääminen tietojärjestelmistä

Kysymys: Mitkä tiedonkeruun tiedoista löytyvät sellaisenaan tietojärjestelmistä, miltä osin tieto voidaan johtaa tietojärjestelmien tiedoista, ja miltä osin tieto kerätään manuaalisesti? Kuvatkaa tarvittaessa, miltä osin tiedonkeruu on erityisen työläs.

Kahdeksan yliopistoa vastasi kysymykseen julkaisutiedonkeruun työläydestä ja julkaisutietojen keräämisestä. Vastanneissa yliopistoissa lähes kaikki pyydetyt tiedot saadaan suoraan omista järjestelmistä lukuun ottamatta kolmea yliopistoa, jossa tiedot kerätään manuaalisesti. Erään yliopiston vastauksessa kommentoidaan, että kaikki tiedot saadaan, mutta tietojärjestelmään joudutaan teettämään muutoksia, mikäli kaikki OKM:n tietojenmuutosvaatimukset aiotaan huomioida. Erään toisen yliopiston mukaan ainoastaan keksintötietoja ei tuoteta julkaisutietokannan kautta, vaan ne saadaan yliopiston tutkimuspalveluyksiköltä, joka kirjaa tiedot omiin järjestelmiinsä.

Yksi yliopisto kommentoi, että tietojen siirto raportointipohjaan on hankalaa ja tietojen saaminen annetulla aikataululla tuottaa myös vaikeuksia. Kahden yliopiston mukaan julkaisutietojen keruu on työlästä myös, koska esimerkiksi tietojen tarkastus vie aikaa. Erään yliopiston vastauksessa todetaan, että koska viitetietojen toimittaminen toteutetaan vuonna 2011 ensimmäistä kertaa, ei vielä ole tiedossa, kuinka täydellisinä tiedot voidaan toimittaa opettajien ja tutkijoiden ohjeistamisesta huolimatta.

Yhden yliopiston vastauksessa kommentoidaan julkaisun tieteenalan määrittämistä ja toivotaan, että nykyistä käytäntöä, jossa tieteenalatieto jyvitetään yksikkötiedon perusteella, voitaisiin jatkaa. Perusteluna on se, että tieteenalajakaumatiedon tallentaminen julkaisukohtaisesti merkitsisi akateemiselle henkilökunnalle lisääntyvää byrokratiaa ja työmäärää, minkä lisäksi järjestelmämuutoksista aiheutuu luonnollisesti ylimääräisiä kustannuksia.

3. Tietojen vertailtavuus korkeakoulujen välillä

Kysymys: Kuvatkaa käsityksenne siitä, minkä tietojen osalta vertailtavuus korkeakoulunne ja muiden korkeakoulujen välillä on erityisen ongelmallinen. Perustelkaa lyhyesti vastauksenne tietokohtaisesti.

Ongelmia vertailtavuudessa nähtiin lähinnä ensimmäisessä kysymyksessä mainituissa julkaisutyypiluokkien eriävissä tulkinnoissa. Yhden yliopiston vastauksessa nostetaan esille sellaisten julkaisujen tilastointi, jotka henkilö on tehnyt edellisen työpaikan rahoituksella, mutta jotka ovat ilmestyneet työsuhteen aikana nykyiseen yliopistoon. Kyseisessä yliopistossa nämä on jätetty pois OKM:lle raportoitavista tiedoista tilastoista, mutta joissakin yliopistoissa nämä on kirjattu OKM:n tilastoon. Taideyliopistot nostavat esille ongelmia vertailtavuudessa tiede- ja taideyliopistojen välillä sekä taidealan julkaisujen sijoittamisessa julkaisutyypiluokkiin.

4. Tietojen käyttö korkeakoulussa

Kysymys: Käytättekö tiedonkeruissa kerättäviä tietoja omiin sisäisiin käyttötarkoituksiinne? Jos, niin mitä tietoja? Mitä tietoja ette seuraa sisäisesti?

Seitsemän yliopistoa vastaa keräävänsä julkaisutietoja myös sisäiseen käyttöön. Erään yliopiston mukaan merkittävimpiä tietoja ovat julkaisuiden lukumäärät käyttäen ryhmittelevinä tekijöinä kansainvälisyyttä, julkaisutyyppejä ja yksikkötietoa. Eräs toinen yliopisto kerää julkaisutietoja myös kansainvälisiä akkreditoitajia varten. Erään taideyliopiston vastauksessa todetaan, että taiteellisen toiminnan tiedot ovat merkittävämpiä kuin julkaisutiedot.

Yhdessä yliopistossa julkaisutietoja käytetään osavuosikatsauksissa sekä indikaattorien laskennassa ja yksiköt hyödyntävät näitä tietoja vuosisuunnitelmia tehdessään. Julkaisutietojärjestelmässä on satunnaisesti vaihtuva tutkijaesittely, mikä lisää yliopiston tutkijoiden näkyvyyttä yliopiston ulkopuolelle. Lisäksi yliopiston henkilökunta käyttää tallennettuja tietojaan YPJ-neuvotteluissaan. Kyseisen yliopiston vastauksessa mainitaan myös, että kaikilla tiedoilla ei ole käyttöä yliopiston toiminnan suunnittelussa ja seurannassa (esim. julkaisujen tai artikkeleiden sivumäärät ja -numeroinnit).

Eräissä yliopistossa julkaisutiedoista tuotetaan eri kriteerein raportteja laitoksittain ja tutkijoittain mm. yliopiston sisäiseen tulosseurantaan, yliopiston julkiseen Tilastokirjaan, toiminnanohjausjärjestelmään sekä OKM:lle ja Tilastokeskukselle. Lisäksi tuotetaan tilastotietoja vuosikertomukseen ja henkilöstöjärjestelmään palkanlisä varten. Julkaisutietoja on käytetty myös vuonna 2010–2011 yliopiston kansainvä-

lisen arvioinnin yhteydessä perustetussa tutkimuksen tietojärjestelmässä. Kaksi yliopistoa vastaa, että julkaisuja käytetään sisäisen rahoitusmallin perusteena

5. Esimerkilaskentoja

Kuvatkaa mahdollisimman tarkasti sanallisesti, tai pseudokoodin/SQL-lausein, miten olette tuottaneet tiedot vuonna 2010. Laskentaprosessin lisäksi kuvatkaa, miten julkaisutyyppi ja tieteenala määritetään. Saatteko tiedot ja kaikki ominaisuudet tietojärjestelmästä? Miten varmistetaan julkaisujen sisältyminen tiedonkeruun piiriin? Onko olemassa tarkistusjärjestelmää joka estää saman julkaisun tallentamisen kahteen kertaan?

Julkaisujen laskentaprosessia kuvasi vastauksessaan 10 yliopistoa. Suurimmassa osassa vastanneita yliopistoja on käytössään julkaisurekisteri, jonne tutkijat tallentavat julkaisunsa. Kahdessa julkaisutiedot kerätään lomakkeella ja kirjasto tallentaa tiedot. Esimerkiksi kahdessa yliopistossa julkaisujen tiedot tallennetaan tutkimustietojärjestelmään ja siirretään sieltä yliopiston tietovarastoon vuosiraportointia varten.

Julkaisutyyppi syötetään joko suoraan tallennusvaiheessa ainakin viidessä yliopistossa. Kahdessa yliopistossa se johdetaan yliopiston omasta julkaisutyyppiluokittelusta esimerkiksi yhdistämällä julkaisun laji, alalaji ja referee-tieto.

Neljässä yliopistossa julkaisut jyvitetään kirjoittajien edustamille yksiköille ja näille määritellyille tieteenaloille tiettyjen prosenttiosuuskien mukaan. Yhdessä yliopistossa julkaisun tieteenalat määritetään ainoastaan ensimmäisen kirjoittajan yksikön tieteenalojen perusteella. Viidessä yliopistossa julkaisun kirjoittaja tai muu tallentaja määrittää julkaisulle tieteenalan syöttäessään julkaisun tietoja. Yliopistojen välillä on eroja siinä, kuinka moneen tieteenalaan julkaisu voi enintään kuulua.

Kaikissa kysymykseen vastanneissa yliopistoissa kirjasto tai muu tutkimushallinnon yksikkö tarkastaa tallennetut julkaisutiedot: puuttuvat tiedot täydennetään ja mahdolliset duplikaatit poistetaan. Useimmissa yliopistoissa kirjasto myös korjaa vääriä tietoja ja käy esimerkiksi läpi julkaisutyyppien oikeellisuuden. Tieto siitä, onko julkaisu vertaisarvioitu selvitetään ainakin kahdessa yliopistossa esimerkiksi ISI Web of Science -tietokannan JCR-lehtiluettelosta tai Ulrich'sin lehtiluettelosta. Ainakin toisessa näistä kirjastoon lähetettävissä kopioissa pyydetään kopiota sivusta, jossa näkyy referee-käytäntö.

Seitsemässä vastanneessa yliopistoissa julkaisutietojärjestelmässä on duplikaattien tunnistamiseksi jonkinlainen toiminto, joka huomauttaa, kun samannimistä julkaisua yrittää tallentaa toiseen kertaan, mutta ei kuitenkaan estä julkaisun tallentamista. Toiminto ei tunnista duplikaatteja, jos julkaisujen nimissä on esimerkiksi kirjoitusvirheitä. Kaikissa vastanneissa yliopistoissa tehdään kuitenkin päällekkäisten tallennusten tarkastusta myös manuaalisesti ja duplikaatteja poistetaan käsin ennen kuin ne raportoidaan OKM:lle. Esimerkiksi kahdessa yliopistossa julkaisutiedot kootaan excel-taulukkoon, josta siivotaan päällekkäiset julkaisut.

Sen varmistamista, että kaikki julkaisut sisältyvät tiedonkeruuseen, pidetään hankalana eikä varsinaista varmistusprosessia ole olemassa yhdessäkään yliopistossa. Useat yliopistot kuitenkin toimittavat laitosten julkaisulistat laitoksille tai yksiköille tarkistettavaksi ennen lopullista raportointia. Neljä yliopistoa kommentoi, että laaja tiedottaminen ja jatkuva muistuttaminen julkaisutietojen keruusta yliopiston sisällä varmistavat mahdollisimman suuren kattavuuden. Yhden yliopiston mukaan laitosten saamat tulospisteet julkaisuista vaikuttavat.