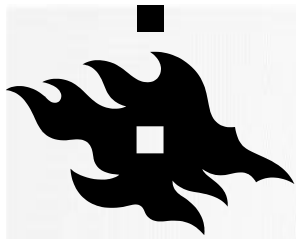




DATAVIITTAUSTEN MERKITYS AVOIMEN TIETEEN EDISTÄMISESSÄ

Susanna Nykyri, YTT
Tietoasiantuntija

Helsingin yliopiston kirjasto, Tutkimuksen palvelut
(Tutkimusdata, avoin julkaiseminen ja avoimen tieteen palvelut)



AVOIN DATA

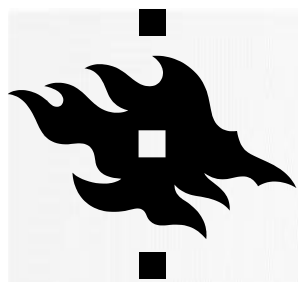
- Rahoittajatahot vaativat, tiedelehdet edellyttävät, tutkijat haluavat...
- “Open Science is changing every aspect of the scientific method to become more open, inclusive and interdisciplinary...Ensuring Europe is at the forefront of Open Science means promoting open access to scientific data and publications alongside the highest standards of research integrity.”
- Carlos Moedas (2016), Commissioner for Research, Science and Innovation

”As open as possible, as closed as necessary”

– European Commission

- ”Akademian edellyttää rahoittamiltaan hankkeilta sitoutumista julkaisujen avoimeen saatavuuteen.

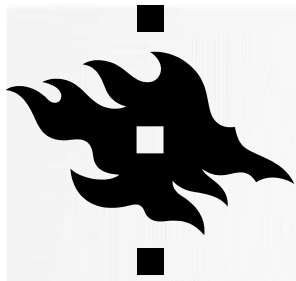
Lisäksi Akademia edellyttää, että sen rahoittamien hankkeiden vastuulliset tutkijat vastaavat siitä, että Akademian rahoittamien hankkeiden tutkimusaineistot tallennetaan ja avataan oman tieteenalan kannalta tärkeässä kansallisessa tai kansainvälisessä arkistossa tai tallennuspalvelussa. Aineistoilla voi kuitenkin perustelluista syistä olla erilaisia avoimuuden asteita vaihdellen kaikille täysin avoimesta salassa pidettävään.” (SA 2017)



METATIETO - AVOINTA JA TÄSMÄLLISTÄ SEKÄ YHTEENTOIMIVAA

Avoimen tieteen ja tutkimuksen käsikirjassa (ATT 2015) muistutetaan tutkijoita mm.

- **Huolehdi kaikkien aineistojesi ja julkaisujesi mahdollisimman laajasta kuvailusta ja metatietojen avoimuudesta.**
- Aikaan ja paikkaan liittyvät tiedot ovat myös tärkeitä, ilmoita ne ohjeiden mukaan. Aineistojen säilytys- tai hakupalvelut usein ohjaavat metatietojen täyttämisessä.
- Käytä oman alasi terminologiaa ja asiasanoja.
- Muista myös vapaamuotoinen ja ainakin osittain yleistajuinen kuvaus: mistä on kysymys? (description).
- Suosi valmiita sanastoja, ontologioita ja yleisessä käytössä olevia termejä löytyvyyden takaamiseksi.
- Lisenssi/käyttöehdot ovat tärkeä metatieto, koneluettava formaatti kuten CC-lisanssi on suositeltava. Käytä mahdollisimman sallivaa lisenssiä. (emt.)



#DATAVIITTAUS

FAIR DATA

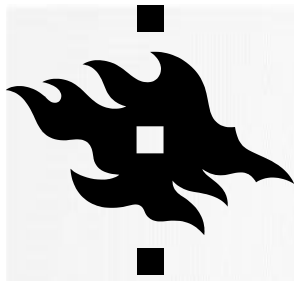
- **F**indable - löydettävissä
- **A**ccessible - saavutettavissa
- **I**nteroperable - yhteentoimivaa
- **R**eusable – uudelleen-/jatkokäytettävää

(Ks. Wilkinson, M. D. et al. 2016)

HYVÄ TIETEELLINEN KÄYTÄNTÖ, sen lähtökohtia ovat mm.

- muiden tutkijoiden työn ja saavutuksien asianmukainen huomioon ottaminen ja viittaaminen;
- eettisesti kestävät tiedonhankinta-, tutkimus- ja arviointimenetelmät sekä avoimuus ja vastuullisuus julkaisemisessa

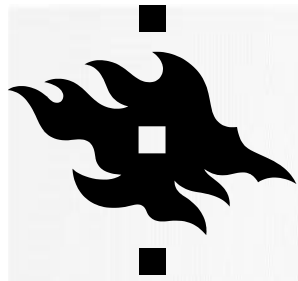
(Ks. TENK 2012 ja HY 2017)



MIKSI VIITATA DATAAN?

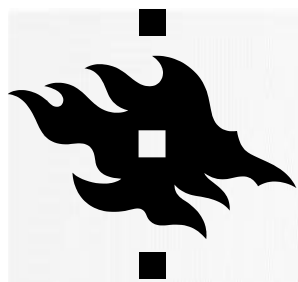
Siksi, että hyvään tieteelliseen käytäntöön kuuluu, että tutkijat ja tieteelliset asiantuntijat:

- 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 ja toteuttavat tieteellisen tiedon luonteeseen kuuluvaa avoimuutta tutkimuksensa tuloksia julkaistessaan
- ottavat muiden tutkijoiden työn ja saavutukset asianmukaisella tavalla huomioon niin, että he kunnioittavat näiden työtä ja antavat heidän saavutuksilleen niille kuuluvan arvon ja merkityksen omassa tutkimuksessaan ja sen tuloksia julkaistessaan. (Ks. TENK 2012)
- **Data tulee nähdä tutkimuksen konkreettisena ja merkityksellisenä tuotoksena siinä kuin perinteinen tutkimusjulkaisu (esim. vertaisarvioitu artikkeli tai monografia). Datalla on myös merkittävä potentiaali tulla uudelleen- tai jatkokäytetyksi. Avoin data ja avoimet metatiedot lisäävät tieteen läpinäkyvyyttä ja ovat omiaan parantamaan tieteen laatua.**



KOHTI MERIITTEJÄ?

- Julkaisut, joiden data on avointa, saavat merkittävästi enemmän (9-69 %) viittauksia kuin julkaisut, joiden dataa ei ole avattu. (Ks. esim. Piwowar & Vision 2013)
- Paitsi eri julkaisutyyppejä, niin myös viittauslähteiden kirjo tulee tunnistaa. Mm. altmetristen työkalujen kehittäminen tärkeää. Arvioinnissa käytettyjen lähteiden ei tulisi olla ns. liikesalaisuuksia, vaan mahdollisimman avoimia.
- Mutta: dataviittauksissa ei ole ensisijaisesti kyse mittaamisesta, vaan ennen kaikkea tieteen laadusta ja vaikuttavuudesta.



DATAN REILUUDEN JA VIITATTAVUUDEN PAHIMMAT ESTEET?

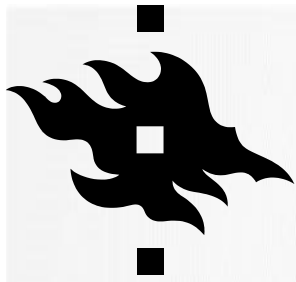
- huono datanhallinta – huonosti hallittua ei voi myöskään avata
- puutteelliset metatiedot, erit. pysyvät yksilöivät tunnisteet



- ei niinkään piittaamattomuutta hyvästä tieteellisestä käytännöstä, vaan pikemminkin tietojen ja taitojen sekä tuen puutetta

- suunta on **selkeästi parempaan**, lisääntyneen tietoisuuden myötä on kehittynyt vahva tahtotila ja myös konkreettisia palveluita, tallennuspalveluita, ohjeistuksia jne

- esimerkiksi HY:n tutkijoilleen tarjoamat DMPTuuli - datanhallinnan koulutukset vetävät osallistujia ei vain rahoitushakemusten vuoksi, vaan enenevästi koska datanhallinta koetaan muutoin tärkeäksi ja ajankohtaiseksi ottaa haltuun. Esim. Mildred-hankkeessa kehitetään konkreettisia tukipalveluita, ohjeita jne.

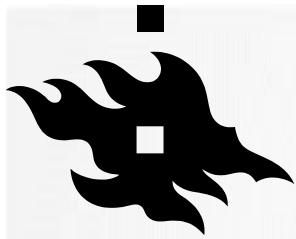


DATAN SAAVUTETTAVUUS?

Helsingin yliopistossa toteutettiin 2016 kysely tutkijoille (vastauksia 260):

- 56 % ei tallenna dataansa mihinkään digitaaliseen repositorioon
- 43 % tallensi yhteen tai useampaan, 21 % kahteen tai useampaan ja 10 % kolmeen tai useampaan

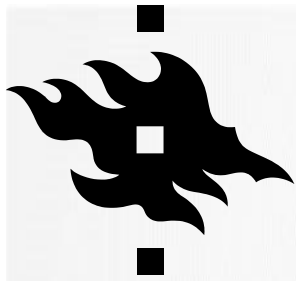
(Salmi & al. 2016, Kuusniemi & Salmi 2017)



- MIKSI EI?

- Tiedon puute yleisin syy olla tallentamatta ja jakamatta dataansa digitaalisten repositorioiden kautta
- Datan sensitiivisyys esteenä 11 %:lle vastaajista
- 11 % vastaajista eivät pitäneet kysymystä alansa kannalta relevanttina
- 8 % piti dataansa liian pienenä
- 8,5 % piti repositorioita turhina
- Vain 4,6 % vastaajista piti silloisia yliopiston tarjoamia ja muita tallennuspalveluita riittävinä datanhallintaan
- Usein esitetty toive oli, että tallennus- ja jakomahdollisuuksista olisi saatavilla enemmän tietoa ja opastusta sekä parhaimmat käytännöt –esimerkkejä

(Emt.)



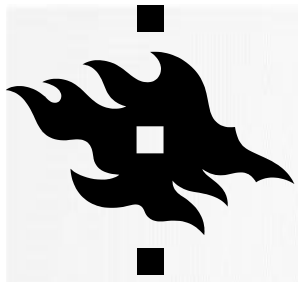
YLEISESTI KÄYTETTÄVIÄ PYSYVIÄ YKSILOIVIA TUNNISTEITA ESIM.

- **julkaisulle ja datalle** DOI (*Digital Object Identifier*), URN (*Uniform Resource Name*), Handle
- **henkilölle/tutkijalle** ORCID (ks. <https://orcid.org/> ja <https://tutkijatunniste.fi/>)
- **henkilölle ja yhteisölle** ISNI (*International Standard Name Identifier*)
- **paikkatiedolle** YSO-paikat (ks. <https://finto.fi/yso-paikat/fi/>)
- **käsitteelle** URI (*Uniform Resource Identifier*, ks. <https://finto.fi/yso/fi/>)

Tunnisteiden käyttö parantaa identifioimista, löydettävyyttä, viitattavuutta, tiedonkeruuta ja linkittämistä

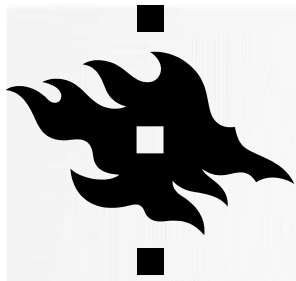


yhteentoimivien metatietojen ansiosta tietojärjestelmissä tekijät linkittyvät julkaisuihin, tutkimusdata julkaisuihin, tekijöihin ja uudelleenkäyttöön jne.



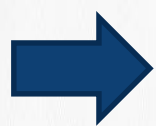
LINKKIMÄTÄ - URL EI RIITÄ

- V. 2014 Klein & al. tutkivat v. 2012 julkaistuja tieteellisiä julkaisuja. 13 % linkeistä arXiv-dokumenteissa ja 22 % linkeistä Elsevierin lehdissä ei enää toiminut. (Useimmiten julkaisu kuitenkin muutoin löydettävissä, ei siis täysin hävinnyt.)
- Ilmiö valitettavan tuttu myös esim. kotimaisille informaatioalan lehdille. (Hakala 2017)



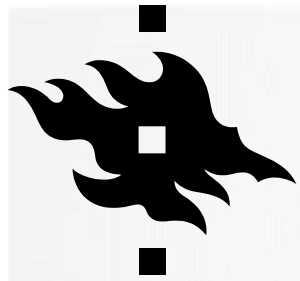
“THE IMPACT FACTOR OF AN OPEN ACCESS JOURNAL DOES NOT CONTRIBUTE TO AN ARTICLE’S CITATIONS”

- “It is better to publish in an OA journal for more citations. It may not be worth paying high publishing fees for higher IF journals, because there is minimal gain in terms of increased number of citations. On the other hand, if one wishes to publish in a non-OA journal, it is better to choose one with a high IF.”
- “In the past, most libraries could possess only a limited number of journals, and librarians used the IF to decide which journals to buy. Consequently, high IF journals were more likely to be purchased, read, and cited. With low IF journals, availability was a constraint. Scientists, wanting a greater audience for their research, preferred to publish in high IF journals. There was plenty of evidence that publishing in a higher IF journal resulted in more citations.”
- “In contrast, at present, open access (OA) journals are universally available. Libraries have no need to subscribe, and researchers can access OA articles freely. Expectedly, OA publication is associated with increased citations, so researchers are likely to prefer this path.” (Chua & al. 2017)

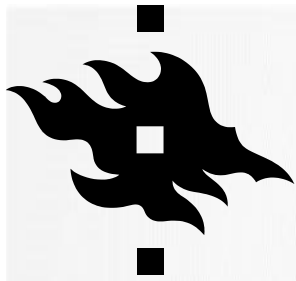


Vrt. datarepositorioiden laatu ja maine? Avointen toimintatapojen (ml.rajapintojen), (loppukäyttäjälleen) maksuttomuuden, datan ja metatietojen kuratoinnin ja neuvonnan merkitys!

Ei kiitos hybrideille ja maksumuureille!



DATAMETRIKKA?



DATAMETRIIKKAA ON JO – MUTTA VÄHÄNLAISESTI

- Ei tule nähdä vain arviointivälineenä, vaan laajemmin kuten bibliometriikka
- Monet repositoriot tarjoavat jo numeerista tietoa datan käyttöön liittyen (katselu- ja latausmäärät, viittaukset)
- Tulevaisuudessa ehkä myös osana altmetriikkaa, mutta välineet ja muut edellytykset eivät vielä riittävän kehittyneitä
- Metodologisia haasteita – kuinka pitkälle meneviä johtopäätöksiä voimme tehdä dataviittaus-lukujen perusteella? Entä miten tarkastella laajemmasta lähtökohdasta kuin yhdestä datajulkaisusta lähtöisin? (Vrt. yliopistoilla varsin kattava käsitys mitä tutkijansa julkaisevat, jos tarkasteltavana vertaisarvioidut artikkelit, monografiat jne, mutta varsin niukasti tietoa yliopistonsa tutkijoiden tuottamasta datasta.)
- Viittausanalyysi evaluointimenetelmänä: riski yliyksinkertaistaa tieteellinen työ ja menestys. Taustalla ajatus, että julkaisut=tuottavuus ja viittaukset=vaikuttavuus.
- Datan sisällyttäminen tarkasteltaviin tutkimustuotoksiin kuitenkin omiaan parantamaan mm. viittausanalyysin ”oikeellisuuden” edellytyksiä. Huomioitava kuitenkin käytäntöjen ja välineiden jne ns. kypsyystaso.

ESIMERKKI: **HY & PLUMX**, YLEISNÄKYMÄ (11.10.2017)
(URL: <https://plu.mx/helsinki/g/>)

[Home](#) / [University of Helsinki](#)



University of Helsinki

Links: [🔗 Website](#)

 Show Profile Data Embed Widget

Artifact Summary

26064



Article

1621



Review

724



Conferen

678



Chapters

388

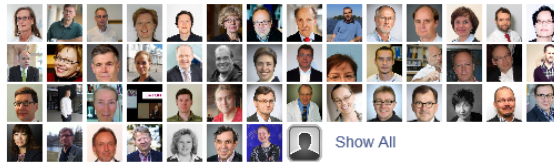


Book

Researchers



Showing 50 of 273

[Show All](#)Narrow by 

Showing 5 of 5

- City Centre Campus
- HiLIFE Fellows (Helsinki Institute of Life Science)
- Kumpula Campus
- Meilahti Campus (Academic Medical Center Helsinki)
- Viikki Campus

All (30218) Article (26064) Review (1621) Conference Paper (724) Book Chapter (678) Book (388) Research Artifact (243) Correction (90) Other (90)

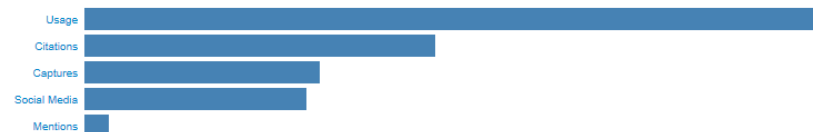
Guideline (69) Abstract (54) Patent (43) Report (37) Letter (36) Clinical Trial (22) Video (17) **Data (12)** Reference Entry (7) Lecture / Presentation (5)

Retraction (5) Thesis / Dissertation (4) Interview (2) Journal (2) Poster (1) Speech (1) Project (1) Bibliography (1) Web Page (1)

Filter...

Analytics

Impact by Type: All





EI VARS. DATASETTI, VAAN "SUPPLEMENTARY TEXT" – PUUTTUU YHTEYS TUTKIMUSTULOKSIA RAPORTOIVAAN ARTIKKELIIN



Supplementary text

Citation data: The Annals of Applied Statistics, ISSN: 1932-6157, Vol: 7, Issue: 1, Page: 369-390

Publication Year: 2013

Researchers: [Matti Pirinen](#)

USAGE ^

4

Abstract Views

3

EBSCO

3

Link-outs

1

EBSCO

1

CAPTURES ^

31

Readers

31

Mendeley

31

CITATIONS ^

27

Citation Indexes

27

Scopus

27

CrossRef

26

DATA SUMMARY

GROUPS:

Meilahti Campus (Academic Medical Center Helsinki)

• FIMM - Institute for Molecular Medicine Finland

HiLIFE Fellows (Helsinki Institute of Life Science)

RESEARCHERS:

[Matti Pirinen](#)

DOI:

10.1214/12-aos586supp, 10.1214/12-aos586

AUTHOR(S):

Matti Pirinen, Peter Donnelly, Chris C. A. Spencer

PUBLISHER(S):

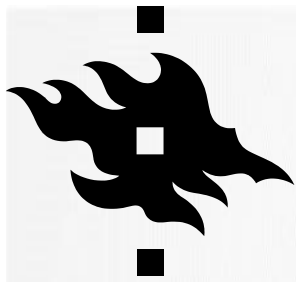
Institute of Mathematical Statistics

TAGS:

Mathematics, Decision Sciences

DATA DESCRIPTION

Motivated by genome-wide association studies, we consider a standard linear model with one additional random effect in situations where many predictors have been collected on the same subjects and each predictor is analyzed separately. Three novel contributions are (1) a transformation between the linear and log-odds scales which is accurate for the important genetic case of small effect sizes; (2) a likelihood-maximization algorithm that is an order of magnitude faster than the previously published approaches; and (3) efficient methods for computing marginal likelihoods which allow Bayesian model comparison. The methodology has been successfully applied to a large-scale association study of multiple sclerosis including over 20,000 individuals and 500,000 genetic variants. © 2013 Institute of Mathematical Statistics.



Huom. Laajat
metatiedot,
viitattavuus,
pysyvä tunniste,
käyttöehtojen
koneluettavuus
+ numeerista
käyttötietoa

DATASET

DATASET

TEXT

2016-07-25_mildr... .csv (45.24 kB)

2016-08-01_mildr... .csv (55.01 kB)

readme.txt (2.49 kB)

Download all (102.74 kB)

Share

Cite

Embed

+ Collect (you need to log in first)

3 files

Salmi, Anna; Ojanen, Mikko; Kuusniemi, Mari Elisa (2016): Project MILDRED Research Data Repository Survey, University of Helsinki. figshare.

<https://doi.org/10.6084/m9.figshare.3806394.v4>

Retrieved: 12:07, Oct 17, 2017 (GMT)

Project MILDRED Research Data Repository Survey,
University of Helsinki

Version 4 ▾ 05.09.2016, 11:43 by Salmi, Anna, Ojanen, Mikko, Kuusniemi, Mari Elisa

This dataset is part of Project MILDRED, Development Project of Research Data Infrastructure at University of Helsinki. The project started on April 29, 2016. Project aim is to provide University of Helsinki with state-of-the-art research data management service infrastructure. To gain knowledge about researchers' data storage and preservation practices in 2016, an e-survey was sent to the UH research staff about 1) what data repositories they use for depositing their research data; 2) what reasons they had for not depositing data and 3) what alternative storage devices and repository services they used for their data.

The dataset consists of e-survey report master file and analysis of the original master file. The files have been anonymized. A readme.rtf file is included to provide full project and data level documentation.

REFERENCES

- <http://blogs.helsinki.fi/mildred/>

FUNDING

University of Helsinki

1163
views

85
downloads

0
citations

CATEGORIES

- Database Management
- Information Systems Organisation
- Data Format
- Library and Information Studies
- Information Retrieval and Web Search
- Data Structures
- Records and Information Management (excl. Business Records and Information Management)

KEYWORD(S)

Research Data Management

Data Repository

Research Data Infrastructure Planning

Research Data Management Practices

LICENCE

CC BY 4.0

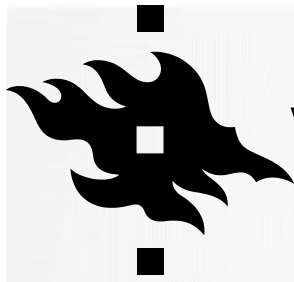
Log in to write your comment here...

Dataviittausten merkitys avoimen tieteen edistämässä

/ Susanna Nykyri

19/10/2017

17



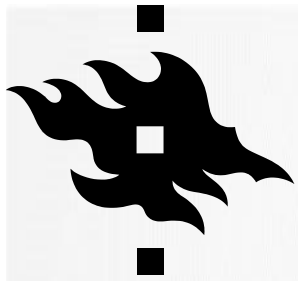
VIITATTAVAN DATAN TULEE OLLA

- Avointa vähintään metatietojensa osalta
 - Pysyvällä tunnisteella varustettu
- Hyvien datanhallintakäytäntöjen mukaista
- Kun huolehdimme datan viitattavuudesta, huolehdimme myös avoimen tieteen perusperiaatteiden toteutumisesta



Esityksen otsikko olisi yhtä lailla voinut olla:

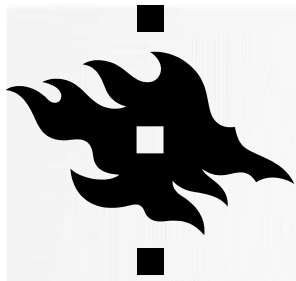
Avoimen tieteen merkitys dataviittausten edistämisessä



Kiitos!

susanna.nykyri (at) helsinki.fi

ORCID ID: orcid.org/0000-0002-5018-5176



LÄHTEET

Avoimien tieteiden ja tutkimuksen käsikirja (ATT, 2015). Avoimen tieteen käsikirja - Tutkijan muistilista metatiedoista. URL: <https://avointiede.fi/muistilista-mt> (Viimeksi luettu 16.10.2017)

Chua, S., Qureshi, A. M., Krishnan, V., Pai, D. R., Kamal, L. B., Gunasegaran, S., ... Sood, S. (2017). The impact factor of an open access journal does not contribute to an article's citations. *F1000Research*, 6, 208. <http://doi.org/10.12688/f1000research.10892.1> (Viimeksi luettu 11.10.2017)

Hakala, J. (2017). E-viittaamisen ihanuus ja kurjuus. *Informaatiotutkimus* 2(36). <https://doi.org/10.23978/inf.65190>

Helsingin yliopisto (HY, 2017). *Tutkimusetiikka*. URL: <https://www.helsinki.fi/fi/tutkimus/tutkimusymparisto/tutkimusetiikka> (Viimeksi luettu 11.10.2017)

Klein, M., Van de Sompel, H., Sanderson, R., Shankar H., Balakireva L., Zhou K., & Tobin, R. (2014) Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. *PLoS ONE* 9(12): e115253. <https://doi.org/10.1371/journal.pone.0115253> (Viimeksi luettu 17.10.2017)

Kuusniemi, E. & Salmi, A. (2017). Project MILDRED: Charting ground for research data management services at the University of Helsinki. *JEAHIL Vol. 13 No. 2 June 2017* — Open Science 2: Research Data. URL: <http://eahil.eu/wp-content/uploads/2017/06/journal-2-2017-web-1.pdf> (Viimeksi luettu 17.10.2017)

Moedas, C. (2016). Introduction to Open Innovation, Open Science, Open to the World – a Vision for Europe. Brussels: European Commission. <https://doi.org/10.2777/061652> (Viimeksi luettu 16.10.2017)

Piowar, H. & Vivion, T. (2013). Data reuse and the open data citation advantage. *PeerJ* 1:e175; DOI 10.7717/peerj.175

Salmi, A., Ojanen, M. & Kuusniemi, M. (2016). *Project MILDRED Research Data Repository Survey*, University of Helsinki. figshare. <https://dx.doi.org/10.6084/m9.figshare.3806394.v4> (Viimeksi luettu 17.10.2017)

Suomen Akatemia (SA, 2017): Avoimien tieteiden avoin julkaiseminen ja aineistojen avaaminen. URL: http://www.aka.fi/avoin_tiede (Viimeksi luettu 19.10.2017)

Tutkimuseettinen neuvottelukunta (TENK, 2012): *Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitleminen Suomessa*. 44 s. URL: http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf (Viimeksi luettu 11.10.2017)

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (1): 160018. doi:10.1038/sdata.2016.18. <http://dx.doi.org/10.1038/sdata.2016.18>. (Viimeksi luettu 11.10.2017)