

Käytetyimpiä monitieteisiä viittaustiedonlähteitä

Viittaustietokannat sisältävät tiedon tietokannassa olevista julkaisuista sekä niiden välisistä [viitauksista](#). Viittaustiedot kertovat, kuinka usein muissa tietokantaan kuuluvissa julkaisuissa on viitattu tarkasteltavaan julkaisuun. Viittaustietokannoista voi tarkastella yksittäisen julkaisun viittaussuhteita lisäksi tyypillisesti mm. sitä, kuinka monta kertaa tiettyyn kirjoittajaan tai tietyn organisaation julkaisuihin on viitattu. Julkaisun saamien viittausten lukumäärä eri tietokannoissa vaihtelee. Viittaustietokannat keskittyvät yleensä tieteellisiin lehtiartikkeleihin, mutta ne voivat sisältää myös monografioita, konferenssijulkaisuja ja raportteja. Mikään viittaustietokanta ei kata kaikkia julkaisuja. On myös hyvä tiedostaa, että tietyt tieteenalat ovat edustettuna tietokannoissa paremmin (esim. lääke- ja luonnontieteet) kuin toiset (esim. yhteiskunta- ja humanistiset tieteet sekä taidealat). Viittausten lukumäärä tietokannoissa riippuu myös siitä, kuinka pitkältä ajalta tietokanta sisältää viittaustietoa ja kuinka usein viittaustiedot päivitetään tietokantaan. Kaikissa tietokannoissa on myös jonkin verran virheitä. Tietokantojen sisältö myös muuttuu jatkuvasti, koska ne lisäävät uusia aineistoja kokoelmaansa ja poistavat entisiä sekä päivittävät viittaustietoja vanhempien julkaisujen osalta.

Kaksi tunnetuinta monitieteistä viittaustietokantaa ovat maksulliset tilausperusteiset Clarivate Analyticsin [Web of Science](#) (WoS) ja Elsevierin [Scopus](#). Sekä WoS että Scopus sisältävät kuratoitua aineistoa ja niillä on tietyt laatuvaatimukset tietokantansa sisällölle: lehdistä pitää mm. olla vertaisarviointiprosessi ja toimituskunta, ISSN-tunnus ja artikkelin otsikko sekä abstrakti tulee olla englanninkielisiä. Viittaustietoja voi tarkastella myös avoimesti saatavilla olevista [Dimensions](#) ja [Google Scholar](#) (GS) -palveluista. Sekä Dimensions että GS indeksoivat automaattisin menetelmin aineistoja kokoelmiinsa ilman erillistä validointiprosessia. Dimensions tietokannan perussisältö ja toiminnot ovat tarjolla avoimesti ja maksuttomasti. Maksuville tilaajille on lisäksi tarjolla sisällöltään laajempi ja monipuolisempia analyysitoimintoja sisältävä versio. GS on maksuton hakukone, joka on erikoistunut tieteellisen tiedon etsimiseen, mutta sieltä löytyy myös ei-tieteellistä aineistoa kuten opinnäytteitä. Sen viittaustieto sisältää myös paljon viitauksia ei-tieteellisistä julkaisuista, joten viittaussuhteet eroavat suuresti WoS- ja Scopus-tietokantojen lukemista. Tässä oppaassa GS esitellään viittaustietokantojen yhteydessä, vaikka kyseessä ei ole varsinainen tietokanta. Kaikki kolme edellä mainittua tietokantaa sekä Google Scholar on esitely tarkemmin omissa luvuissaan.

Viittaustiedonlähteiden vertailua

Koska viittaustietokantojen aineistojen indeksointimenetelmät ja kokoelmapoliittiset linjaukset ovat keskenään hyvin erilaisia, eroavat myös tietokantojen sisällöt merkittävästi toisistaan. WoS ja Scopus noudattavat valikoivaa kokoelmapoliittikkaa ja käyttävät eri alojen asiantuntijoista koostuvaa toimituskuntaa sisältönsä valintaan ja indeksointiin. Dimensions nojaa laajasti koneoppimiseen ja automatisoiduilla algoritmeilla tapahtuvaan tiedonkäsittelyyn, asiantuntijoiden manuaalisesti suorittaman kuratoinnin sijaan. Dimensions indeksoi pääsääntöisesti kaikki tieteelliset julkaisut ja datasetit, joilla on DOI-tunniste. Google Scholarin automatisoidut menetelmät puolestaan kahlavat verkkoa ja indeksoivat kaikki julkaisut, joilla tunnistavat olevan akateeminen rakenne.

Tieteenalojen julkaisukäytännöt kuten julkaisutyyppi ja julkaisukieli heijastuvat tietokantojen kattavuuteen. Perinteiset viittaustietokannat kuten WoS ja Scopus sisältävät pääsääntöisesti tieteellisiä lehtiä. Erityisesti tieteenalat, joille on tyypillistä julkaista kirjoissa, ovat huonosti edustettuina kyseisissä tietokannoissa. Myös konferenssijulkaisujen osalta on puutteita. Heikointa kattavuus on tyypillisesti yhteiskuntatieteiden ja humanististen tieteiden osalta. Myös tietokantojen keskittyminen pääasiassa englanninkieliseen kirjallisuuteen rajoittaa kattavuutta. Vaikka tietokantoihin hyväksytään myös muun kuin englanninkielisiä julkaisuja, yhtenä valintakriteerinä on kuitenkin julkaisukanavissa julkaistavien artikkelien englanninkieliset tiivistelmät ja otsikot. Automaattisiin indeksointimenetelmiin perustuvat Dimensions ja GS yltyvät merkittävästi suurempaan kattavuuteen useimmilla tieteenaloilla. Dimensions on huomattavasti kattavampi kuin WoS tai Scopus erityisesti toimitettujen kirjojen, niiden sisältämien kappaleiden, sekä monografioiden osalta.

Tietokantojen eroista ja siitä, miten hyvin ne kattavat eri tieteenalojen aineistoja, on tehty lukuisia tutkimuksia. Useat tutkimukset osoittavat WoS Core Collection -tietokannan kattavuuden monien tieteenalojen osalta heikommaksi kuin kilpailevissa tietokannoissa (mm. Mongeon, P. and Paul-Hus, A. 2016, Martín-Martín, A. et al. 2021, Singh, V.K. et al. 2021, Visser, M. et al. 2021). Tässä oppaassa esitellyistä viittaustiedonlähteistä kattavuudeltaan parhaimmaksi lähes kaikkien tieteenalojen osalta nousee useissa tutkimuksissa Google Scholar (mm. Harzing, A.-W. 2019, Martín-Martín, A. et al. 2021). Tuoreita tutkimuksia viitetietokantojen aineistojen kattavuudesta on listattu tämän luvun loppuun. Lisäksi liitetiedostona on Jyväskylän yliopiston julkaisutuotannon perusteella tehty tietokantojen kattavuusvertailu.

Tietokantaa valittaessa aineiston kattavuus ei ole kuitenkaan ainoa kriteeri. Myös tietokannan hakutoiminnoilla ja indeksoidun aineiston laadulla on merkitystä. Vaikka GS yltyä kilpailijoitaan suurempaan kattavuuteen, on sen hakutoiminnoissa paljon rajoitteita, kuten rajoitettu Boolean operaattoreiden käyttö sekä hyvin rajalliset suodatusominaisuudet hakutulokselle. Perinteiset, pitkään markkinoilla olleet WoS ja Scopus ovat kattavuudeltaan kilpailijoita heikompia, aineiston indeksointi tietokantaan on hitaampaa ja ne ovat maksullisia. Toisaalta molempien hakutoiminnot ovat monipuoliset ja niiden indeksoima aineisto on validoitua. Viittaustietokantojen vertailua helpottamaan on taulukkoon 1. kerätty tiedot Web of Science, Scopus ja Dimensions tietokantojen sekä Google Scholar -palvelun keskeisistä ominaisuuksista ja sisällöistä. Vastuullisen julkaisumetriikan nimissä analyysissä käytettävä viittaustietokanta tulisi valita käyttötarkoituksen lisäksi analysoinnin kohteena olevan tieteenalan tarpeet huomioiden. Mikäli mahdollista, on suositeltavaa käyttää julkaisuperusteisissa analyyseissä useampaa kuin yhtä viittaustietokantaa.



Viittaustietokantojen tutkija- ja organisaatioprofiilit

- Tutkija- ja organisaatioprofiilit ovat tyypillisesti palveluntarjoajien omien algoritmien avulla muodostettuja profiileja julkaisuissa esiintyvälle organisaatioille ja kirjoittajille.
- Viittaustietokantojen tutkija- ja organisaatioprofiileissa on usein virheitä. Julkaisujen tekijöiden tunnistamista vaikeuttavat erityisesti yleiset sukunimet, nimenvaihdokset ja englannin kielessä esiintymättömät merkit. Vastaaavasti organisaatioiden nimet kokevat vaihtoksia ja niissä voi tapahtua rakenteellisia muutoksia. Viittaustietokantojen algoritmit eivät välttämättä tunnista tällaisia muutoksia.
- Julkaisu liitetään kuuluvaksi tietylle tutkijalle ja organisaatiolle julkaisussa ilmoitettujen tietojen perusteella. Tämän vuoksi on ensisijaisen tärkeää, että kirjoittajat merkitsevät organisaationsa oikein kirjoittamiinsa julkaisuihin. Organisaatioiden ja niiden yksiköiden nimet on hyvä kirjoittaa kokonaisuudessaan auki ja lyhenteitä on syytä välttää.

- Web of Science ja Scopus tietokantojen osalta organisaatioiden ja tutkijoiden on myös itse mahdollista huolehtia profiileidensa ajantasaisuudesta ja oikeellisuudesta. Tutkija- ja organisaatioprofiilien virheettömyydestä huolehtiminen on jatkuva prosessi. Tutkijaprofiilit saattavat rikkoontua korjauksista huolimatta, ja jo korjattuihin organisaatioprofiileihin saatetaan liittää siihen kuulumattomia julkaisuja palveluntarjoajan toimesta.
- Viittaustietokantojen organisaatioprofiilien ajan tasalla pitäminen on tärkeä tehtävä koska tietoja käytetään erilaisissa benchmarking- ja yhteistyöselvityksissä. Lisäksi monet ranking-organisaatiot hyödyntävät tietoja

	Web of Science	Scopus	Dimensions	Google Scholar
Saatavuus	Maksullinen	Maksullinen	Perussisältö ja -toiminnot avoimesti saatavilla, edistyneempi käyttö ja kattavammat sisällöt ovat maksulliset.	Avoimesti saatavilla
Lehtien lukumäärä (Dimensions: Julkaisukanavien lkm.)	~22 000	~24 000	~99 000 + ~1M kirjaa (~45 500 julkaisukanavalle omaa kanavatason metadataa)	Ei julkinen
Muu sisältö	Julkaisut: konferenssijulkaisuja, monografioita. Pienissä määrin myös muita julkaisutyppejä. Sisältää lisäksi tietoa rahoittajista	Julkaisut: konferenssijulkaisuja, monografioita, kirjan /kokoomateoksen osia, ammattilehtiä ja patenteja. Pienissä määrin myös muita julkaisutyppejä. Sisältää lisäksi tietoa rahoittajista	Julkaisut: kirjan/kokoomateoksen osia, konferenssijulkaisuja, preprinttejä, monografioita, patenteja ja yhteiskunnallisia julkaisuja Sisältää lisäksi tietoa dataseteistä, tutkimuksen rahoituksesta ja klinisistä kokeista. Tieteelliset julkaisut ja tiedot dataseteistä sisältyvät ilmaisversioon. Muu sisältö maksullista.	Konferenssijulkaisuja , monografioita, preprinttejä, opinnäytteitä, ppt-esityksiä, www-sivuja
Keskeisimmät alat	Luonnontieteet, lääketiede, tekniikka, yhteiskuntatieteet, taiteet ja humanistiset tieteet	Luonnontieteet, tekniikka, terveystieteet, yhteiskuntatieteet, taiteet ja humanistiset tieteet	Parhainta kattavuus on aloilla: luonnontieteet, lääketiede, tekniikka ja yhteiskuntatieteet	Ei julkinen
Ajallinen kattavuus	Vuodesta 1900 lähtien (Science), vuodesta 1956 (Social Sciences) ja 1975 (Arts and Humanities), aineiston saatavuus riippuu organisaation hankkimista lisensseistä	Kattavuus vaihtelee, osa lehdistä 1780-luvulle asti, viitteissä lähdetiedot 1970 lähtien	Ei erillistä linjausta indeksoitavan aineiston iästä, indeksoi pääsääntöisesti kaikki julkaisut, joilla on DOI-tunniste.	Ei julkinen

Kielellinen kattavuus ja indeksoitavan aineiston kielivaatimukset	Sisältää pääsääntöisesti englanninkielistä aineistoa. Pienessä määrin myös muunkielistä aineistoa. Indeksoitavaksi hyväksyttävien julkaisukanavien artikkeleiden otsikko ja tiivistelmä tulee olla englanninkielinen.	Sisältää pääsääntöisesti englanninkielistä aineistoa. Pienessä määrin myös muunkielistä aineistoa. Indeksoitavaksi hyväksyttävien julkaisukanavien artikkeleiden otsikko ja tiivistelmä tulee olla englanninkielinen.	Sisältää pääsääntöisesti englanninkielistä aineistoa. Myös muunkielistä aineistoa jos niillä on DOI -tunniste. Julkaisu-, patentti ja apurahatietoja myös muilla kielillä. Ei erillisiä kielellisiä vaatimuksia indeksoitavalle aineistolle. Apuraha- ja patenttitiedoissa käytetään tiivistelmien ja otsikoiden automaattista konekäännöstä.	Ei julkinen mutta sisältää englanninkielisen aineiston lisäksi myös muunkielistä aineistoa.
Kokoelmapolitiikka	Julkinen	Julkinen	Ei erillistä kokoelmapolitiikkaa. Indeksoidaan pääsääntöisesti kaikki julkaisut, joilla on DOI-tunniste.	Ei julkinen, sopimus useimpien merkittävien kustantajien kanssa
Viittausanalyysi	Citation Report -työkalu	Citation Overview -työkalu	Analytical Views hakutuloksen yleisnäkymä sekä erilliset näkymät tieteenaloille, tutkijoille ja julkaisukanaville	Hakutulosten yhteydessä Cited by -linkki, josta saa näkyviin julkaisuun viittaavat julkaisut sekä näkee viittausten määrän per julkaisu.
Viittaustietojen ajallinen kattavuus	Vuodesta 1900 lähtien (Science), vuodesta 1956 (Social Sciences) ja 1975 (Arts and Humanities)	Vuodesta 1970	Vaihtelee. Tietokannassa on paikoin indeksoituna yli 100 vuotta vanhoja viittauksia, toisaalta joistakin aivan uusista indeksoituista julkaisuista saattaa puuttua osa viitatuista lähteistä.	Ei julkinen
Indikaattoreita	Viittausten määrä, keskimääräinen viittaustilavuus per julkaisu, keskimääräinen viittaustilavuus per julkaisu, keskimääräinen viittaustilavuus vuosittain, h-indeksi, tietueen käyttötilasto (kuinka monta kertaa yksittäinen tietue on joko ladattu viiteidenhallintajärjestelmään tai avattu julkaisun kokoteksti)	Viittausten määrä, viittaustilavuus vuosittain, h-indeksi, tietueen katsontakerrat, PlumX käyttötilastoja kuvaavia indikaattoreita, tietuealanormalisointi viittaustilavuusindikaattori artikkeleille ja lehdille.	Viittausten määrä, viittaustilavuus viimeiseltä kahdelta vuodelta, keskimääräinen viittaustilavuus per julkaisu, keskimääräinen viittaustilavuus per julkaisu, Altmetric.com palvelun löytämä online-huomio yksittäiselle tutkimustuotokselle, tietuealanpainotettu ja suhteellinen viittaustilavuus	Viittausten määrä per julkaisu Google Scholar profiilissa lisäksi tutkijakohtainen viittaustilavuus, h-indeksi ja i10 -indeksi kaikilta vuosilta sekä vuodesta 2017 alkaen.
Dataa hyödyntäviä yliopisto-rankingeja	Academic Ranking of World Universities (ARWU) eli Shanghai Ranking University Ranking by Academic Performance (URAP) U.S. News & World Report's Best Global Universities Rankings U-Multirank National Taiwan University (NTU) Ranking CWTS Leiden Ranking CWUR (Center for World University Ranking, United Arab Emirates) Round University Ranking (RUR) Suomen Akatemian Tieteen tila -raportit	Times Higher Education World University Ranking Times Higher Education Impact Ranking QS Ranking Suomen Akatemian Tieteen tila -raportit	Ei tiedossa rankingeja, jotka hyödyntävät Dimensions dataa	Ei tiedossa rankingeja, jotka hyödyntävät Google Scholar dataa
Tutkijaprofiili	ResearcherID/Publons	Scopus Author Identifier	Dimensions researcher profile	Google Scholar Profile
Dataa hyödyntäviä työkaluja	InCites, Journal Citation Reports, Eigenfactor, ScienceWatch, Publish or Perish	SciVal, SCImago Journal and Country Rank, CWTS Journal Indicators, Publish or Perish	Dimensions Analytics, JYUcite	Publish or Perish

Taulukko 1. Yhteenveto neljän keskeisimmän monitieteellisen viittaustiedonlähteen ominaisuuksista.

Lähteet

Clarivate Analytics (ei päivystä) *Web of Science LibGuides*. Saatavilla: <https://clarivate.libguides.com/home> (Viitattu 4.2.2022)

Digital Science (ei päivystä) *Dimensions*. Saatavilla: <https://www.dimensions.ai/products/free/> (Viitattu 12.4.2022)

Google (ei päivystä) *About Google Scholar*. Saatavilla: <https://scholar.google.com/intl/fi/scholar/about.html> (Viitattu 18.2.2022)

Harzing, A.-W. (2019) Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, 120(1), pp. 341–349. Saatavilla: <https://doi.org/10.1007/s11192-019-03114-y>

Martin-Martin, A., Thelwall, M., Orduna-Malea, E. and López-Cózar, E.D. (2021) Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 126: pp. 871–906. Saatavilla: <https://doi.org/10.1007/s11192-020-03690-4>

Mongeon, P. and Paul-Hus, A. (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106, pp. 213–228. Saatavilla: <https://doi.org/10.1007/s11192-015-1765-5>

Elsevier (ei päiväystä) *Scopus*. Saatavilla: <https://www.elsevier.com/solutions/scopus> (Viitattu 12.4.2022)

Singh, V.K., Singh, P., Karmakar, M. et al. (2021) The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126, pp. 5113–5142. Saatavilla: <https://doi.org/10.1007/s11192-021-03948-5>

Visser, M., van Eck, N.J. and Waltman, L. (2021) Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), pp. 20–41. Saatavilla: https://doi.org/10.1162/qss_a_00112

Viimeaikaisia viittaustiedonlähteiden kattavuutta ja ominaisuuksia käsitteleviä tutkimuksia

Baas, J., Schotten, M., Plume, A., Côté, G. and Karimi, R. (2020) Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1), pp. 377–386. Saatavilla: https://doi.org/10.1162/qss_a_00019

Birkle, C., Pendlebury, D. A., Schnell, J. and Adams, J. (2020) Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), pp. 363–376. Saatavilla: https://doi.org/10.1162/qss_a_00018

Gusenbauer, M. (2022) Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. *Scientometrics*, 127, pp. 2683–2745. Saatavilla: <https://doi.org/10.1007/s11192-022-04289-7>

Harzing, A.-W. (2019) Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, 120(1), pp. 341–349. Saatavilla: <https://doi.org/10.1007/s11192-019-03114-y>

Herzog, C., Hook, D. and Konkiel, S. (2020) Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, 1(1), pp. 387–395. Saatavilla: https://doi.org/10.1162/qss_a_00020

Huang, C.-K., Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K. and Ozaygen, A. (2020) Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies*, 1(2), pp. 445–478. Saatavilla: https://doi.org/10.1162/qss_a_00031

Martín-Martín, A., Orduna-Malea, E. and López-Cózar, E.D. (2018) Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: A multidisciplinary comparison. *Scientometrics*, 116(3), pp. 2175–2188. Saatavilla: <https://doi.org/10.1007/s11192-018-2820-9>

Martín-Martín, A., Orduna-Malea, E., Thelwall, M. and López-Cózar, E.D. (2018) Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), pp. 1160–1177. Saatavilla: <https://doi.org/10.1016/j.joi.2018.09.002>

Martín-Martín, A., Thelwall, M., Orduna-Malea, E. and López-Cózar, E.D. (2021) Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, 126, pp. 871–906 (2021). Saatavilla: <https://doi.org/10.1007/s11192-020-03690-4>

Mongeon, P. and Paul-Hus, A. (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106, pp. 213–228. Saatavilla: <https://doi.org/10.1007/s11192-015-1765-5>

Singh, V.K., Singh, P., Karmakar, M. et al. (2021) The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126, pp. 5113–5142. Saatavilla: <https://doi.org/10.1007/s11192-021-03948-5>

Visser, M., van Eck, N.J. and Waltman, L. (2021) Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1): pp. 20–41. Saatavilla: https://doi.org/10.1162/qss_a_00112

Waltman, L. and Larivière, V. (2020) Special issue on bibliographic data sources. *Quantitative Science Studies*, 1(1), pp. 360–362. Saatavilla: https://doi.org/10.1162/qss_e_00026

Jyväskylän yliopiston julkaisutuotannon perusteella tehty tietokantojen kattavuusvertailu

Seppänen, J.-T. (ei päiväystä) *Comparing citation database coverage: University of Jyväskylä research publications in Dimensions, Scopus, Web of Science and PubMed*. Kyseessä on vielä julkaisematon tuotos. Saatavilla: [Comparing citation database coverage _draft.pdf](#)