



Havaintoja OpenAlex-tietokannan hyödyntämisestä

Julkaisumetriikkaverkoston tapaaminen 11.12.2024

Joonas Nikkanen & Yrjö Leino, CSC



Sisältö

- Lyhyesti OpenAlex
- Kansallinen bibliometriikkatyö ja huomiot OpenAlex aineistosta
- Kansainvälisiä verrokkeja OpenAlex tietojen käytössä
- Tutkimustietovarannon hyödyntäminen?



OpenAlex

- “OpenAlex is a free and open catalog of the world's scholarly research system” <https://openalex.org/>
 - Avoin tietokanta viittaustietoihin, OurResearch organisaation ylläpitämä – taustalla [Microsoft Academic Graph \(MAG\)](#) sulkeminen 2021 lopulla
 - 2022 ensimmäinen versio ns. data snapshotista sekä APIsta, 2023 lopulla hakukäyttöliittymä verkossa
 - Aggregoi tietoja eri lähteistä, keskeisinä MAG sekä Crossref, mutta myös liuta muita (ROR, Unpaywall, tieteenalakohtaiset repositoriot, Pubmed + crawling)
 - Esim. CWTS julkaisee nykyään Leiden Rankingista version OpenAlex tietoja käyttäen <https://open.leidenranking.com/>
- Päälle rakennettu erilaisia lisäpalveluita 3. osapuolien toimesta, esim. <https://semopenalex.org/> -> semanttinen tietämysgraafi RDF formaatissa

Kansallinen bibliometriikkatyö

- OKM:n, SA:n ja CSC:n yhteistyössä tuottamia bibliometriikka-analyyseja
- Aineistona WoS-tietokannan tiedot (Scopus-aineistosta luovuttu 2024 aikana)



[suomeksi](#) | [på svenska](#) | [in English](#)

[Tietoja palvelusta](#)
[Koulutuksen yhteiset](#)
[Varhaiskasvatus](#)
[Esi- ja perusopetus](#)
[Lukiokoulutus](#)
[Ammatillinen koulutus](#)
[Ammattikorkeakoulutus](#)

[Yliopistokoulutus](#)
[Korkeakoulutus ja t&k-toiminta](#)
[Vapaa sivistystyö](#)
[Väestö-, koulutus- ja ammattirakenne](#)
[Ennakointi](#)

[Korkeakoulutuksen yhteiset ja t&k-toiminta](#) > Bibliometriikka (Web of Science)

Bibliometriikka (Web of Science)

Bibliometrinen raporttien avulla voit vertaillumaiden julkaisumääriä, tutkia julkaisu-yhteistyötä tieteenaloittain perusteella. Suomen osalta tarkastella kansallisten tutkimusorganisaatioiden ja Suomen yhteistyömaiden mukaan.

Bibliometrinen analyysien tiedot jollain kaikki viittausindikaattorit myös aikaisemmin julkaistujen raporttien viimeksi päivitetty huhtikuussa tapahtuu arviolta toukokuussa 2024.

Bibliometriikkaan liittyviä määritelmiä osittaminen ja top 10 -indeksi löydät

Maiden julkaisuprofiilit

Maaryhmä: EU15-maat | Maa: Kaikki | Päätieteenala: Kaikki | Tieteenalaryhmä: Kaikki

Valitse: Multiple sel... | All | 1993-1996

Maa	Matematiikka, tilastotiede	Fysiikka, geotieteet, avaruustieteet	Kemian, tekniillinen kemia	ICT- ja sähkötekniikka	Materiaalitiede ja tekniikka	Tekniikan muut alat	Taloustieteet	Ekologia, ympäristötiede, kannubiologia	Maatalous- ja metsätieteet	Biolääketieteet, biotieteet	Kliiniset lääketieteet	Terveystieteet	Käyttötutkimus tieteet	Yhteiskuntatieteiden muut alat	Humantieteet
Alankomaat	0.51	0.75	0.50	0.73	0.69	0.60	1.26	0.79	0.68	1.05	1.69	1.50	1.63	1.74	1.74
Belgia	0.82	0.86	0.75	0.85	0.65	0.71	1.05	0.99	1.05	1.08	1.20	1.00	1.19	1.53	1.53
Espanja	0.97	0.75	0.81	0.70	0.54	0.74	1.11	1.10	1.22	0.84	0.95	1.30	1.88	1.78	1.78
Itävalta	0.56	0.59	0.58	0.98	0.61	0.58	1.24	0.72	0.66	0.86	1.21	1.80	1.96	1.79	1.79
Iso-Britannia	0.71	0.82	0.52	0.75	0.54	0.69	1.49	0.76	0.58	0.96	1.16	1.42	1.61	2.19	2.19
Italia	1.07	0.98	0.67	0.87	0.60	0.91	1.09	0.85	1.11	1.15	1.44	0.85	0.66	0.79	0.79
Itävalta	1.51	0.94	0.75	1.14	0.72	0.81	1.13	0.95	0.81	1.00	1.15	0.70	0.91	1.11	1.11
Ranska	0.91	0.89	0.65	1.31	0.55	1.03	1.62	1.01	1.14	0.99	1.38	0.87	0.95	0.81	0.81
Luxemburg	1.51	0.89	0.54	2.36	0.74	0.77	2.14	0.79	0.27	0.76	0.48	0.90	1.37	1.84	1.84
Portugali	0.82	0.72	0.82	1.07	0.88	1.18	1.54	1.23	1.24	0.87	0.86	1.15	1.44	1.40	1.40
Saksa	1.32	1.30	0.86	1.06	0.81	0.43	1.23	0.80	0.69	1.08	1.14	0.64	0.53	0.71	0.71
Ruotsi	0.63	0.81	0.63	0.88	0.66	0.87	1.29	0.87	0.62	0.97	1.15	1.32	1.26	1.71	1.71
Sveitsi	1.04	1.19	0.99	1.03	0.86	0.80	1.07	0.84	0.66	1.02	1.09	0.72	1.14	0.93	0.93
Suomi	0.75	0.84	0.63	2.21	0.65	0.65	2.01	1.08	0.97	0.74	0.65	1.14	1.68	1.86	1.86
Tanska	0.44	0.70	0.56	0.81	0.44	0.66	1.25	0.98	1.01	1.14	1.42	1.62	0.83	1.48	1.48

[Suomen Akatemia](#) / [Suomen Akatemian toiminta](#) / [Tietoaaineistot](#) / [Tieteen tila](#)

Tieteen tila

Suomen Akatemian tieteen tila -työn tavoitteena on tuottaa materiaalia, joka tukee korkeakoulujen ja valtion tutkimuslaitosten omaa kehitystyötä ja vahvistaa tiedepoliittain toimijoiden käytössä olevaa tietopohjaa. Tieteen tilassa julkaistaan säännöllisesti tilastollisia ja bibliometrisiä tarkasteluja tutkimuksen rahoituksesta ja henkilöstöstä sekä julkaisu-toiminnasta. Lisäksi julkaistaan teemaattisia analyysejä.

Uusimmat Tieteen tila -tilastot


 Etusivu Haku **Tiede- ja innovaatiopolitiikka** ▾ Tiede- ja tutkimusuutiset Suomeksi ▾

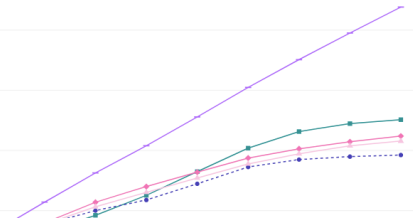
<< Takaisin

[Lukuja tieteestä ja tutkimuksesta](#)
[Suomen ja vertailumaaryhmien ja -maiden julkaisumäärät](#)

Suomen ja vertailumaaryhmien ja -maiden julkaisumäärät

Ohje näppäimistökäyttäjälle

Maaryhmä: Multiple sel... | Päätieteenala: All | Tieteenala: All | Valitse alousoineivotokausi: 1993-1996



INDIKAATTORIT
 HAKU JA VALINTA
 OPISKELIJAT JA TUTKINNOT
 TUTKINNON SUORITTANEIDEN SJOITTUMINEN
 OPETTAJAKELPOISUUDET
 HENKILÖSTÖ
 TIETEEN JA TEKNOLOGIAN HENKILÖVOIMAVARAT
 TUTKIMUS- JA KEHITTÄMISTOIMINTA
 BIBLIOMETRIKKA (WEB OF SCIENCE)
 KANSAINVÄLISYYS
 JULKAISUT
 KORKEAKOULUTUS JA

Kansallinen bibliometriikkatyö

- Vuoden 2024 aikana otettu työstöön OpenAlex-tietokannan tiedot (snapshot)
 - Tähän mennessä rakennettu tietojen sisäänluku + käsittely
 - Alustavia vertailutietoja OpenAlex sisältämien tietojen:
 - Kattavuudesta (suhteessa WoS, myöhemmin myös tutkimustietovarantoon)
 - Laadusta (erit. affiliaatitietojen ja tekijätietojen osalta)
 - Korjata affiliaatitietojen osalta ne tiedot, jotka korjattavissa (WoS aineiston käsittelyn pohjalta tehdyllä mallilla)
 - Pl. täysin virheelliset affiliaatitiedot, näiden autom. korjailu lähes mahdotonta
- Tarkoitus tuottaa vuoden 2025 aikana vastaavat vaikuttavuusindikaattorit suomalaisten organisaatioiden osalta kuin WoS aineistolla
 - Näiden mahdollinen julkaiseminen Vipunen-palveluun WoS analyysien rinnalle

Kansallinen bibliometriikkatyö – huomiot OpenAlex aineistosta tähän mennessä



- Datan laadusta:

- Affiliaatio-ongelmat
 - Osa korjattavissa, osa täysin virheellisiä
- Nimiasut → erikoismerkkejä yms. ylimääräistä nimissä

- Datan kattavuudesta:

- Vertailu DOI-tunnisteellisten osalta WoS ↔ OpenAlex
- Vaikuttavuusarviot → miten tulokset eroaa WoS ↔ OpenAlex

Julkaisuvuosi 2010 - 2019	OpenAlex WoS	
Suomeen affilioituneiden tutkijoiden tekemiksi tunnistettuja julkaisuja	218407	198342
- englanninkielisiä	206451	197203
- englanninkielisten osuus	94,5 %	99,4 %
- suomenkielisiä	7189	211
- suomenkielisten osuus	3,3 %	0,1 %
- julkaisut, joilta puuttuu DOI	24925	36332
- näiden osuus	11,4 %	18,3 %
- julkaisut, jotka molemmissa aineistoissa (DOI)	135622	135622
- julkaisut, jotka vain kyseisessä aineistossa (DOI)	57860	26388

Kansainvälisiä verkkoja OpenAlex tietojen käytössä

- Ranska: ScanR sekä works-magnet ja OpenAlex affiliaatitietojen korjaukset
- Norja: Cristin-data ja OpenAlex vertailu
- Tanska: NORA-portaali ja OpenAlex integraatio

OpenAlex - Ranska

- Ranskassa ministeriön tasolla sitouduttu hyödyntämään OpenAlex tietoja – osa avoimen tieteen kansallista ohjelmaa
 - Keskitytty erityisesti affiliaatitiedon laatuun ja sen korjaamiseen
 - Tarjoavat organisaatioille työkalua, jolla OpenAlexista löytyviä affiliaatitietojen (ROR-tunnisteellisia) korjauksia voi lähettää OpenAlexille – tehtyjä korjauksia n. 7000 tämän vuoden puolella <https://works-magnet.esr.gouv.fr/>
 - Taustalla ajatus, että tietojen korjaamiseen liittyvä manuaalinen työ/tarkistukset eivät valu hukkaan kaupallisten toimijoiden tuotteisiin, vaan ovat läpinäkyvästi saatavilla
- Hyödyntävät myös kansallisessa ScanR—portaalissaan OpenAlexin tarjoamia aiheita, viittausmääriä ja yhteisjulkaisujen tekijätietoja <https://scanr.enseignementsup-recherche.gouv.fr/>
- Tiedostavat että OpenAlexin sisältämien tietojen kattavuus ei ole vielä samalla tasolla kuin WoS tai Scopus tietojen, mutta vuoden 2024 aikana on menty jo huomattavasti eteenpäin

Works-magnet to improve affiliations metadata

You can search


- by string
- by ROR id

You can add ROR children

You can exclude ROR id

Use cases

- I want to make sure all the publications from ESSEC have the ROR <https://ror.org/02dga6j42>
- I want to make sure all the publication with the ROR <https://ror.org/02dga6j42> are from the ESSEC institution

Affiliation name, ROR of your institution  *

Press ENTER to search for several terms / expressions. If several, an OR operator is used.

02dga6j42 X

École Supérieure des Sciences Économiques et Commerciales X

ESSEC X

ESSEC Business School X

Institut Economique X

 Get children from ROR

ROR to exclude: exclude affiliation strings already mapped to a specific ROR in OpenAlex

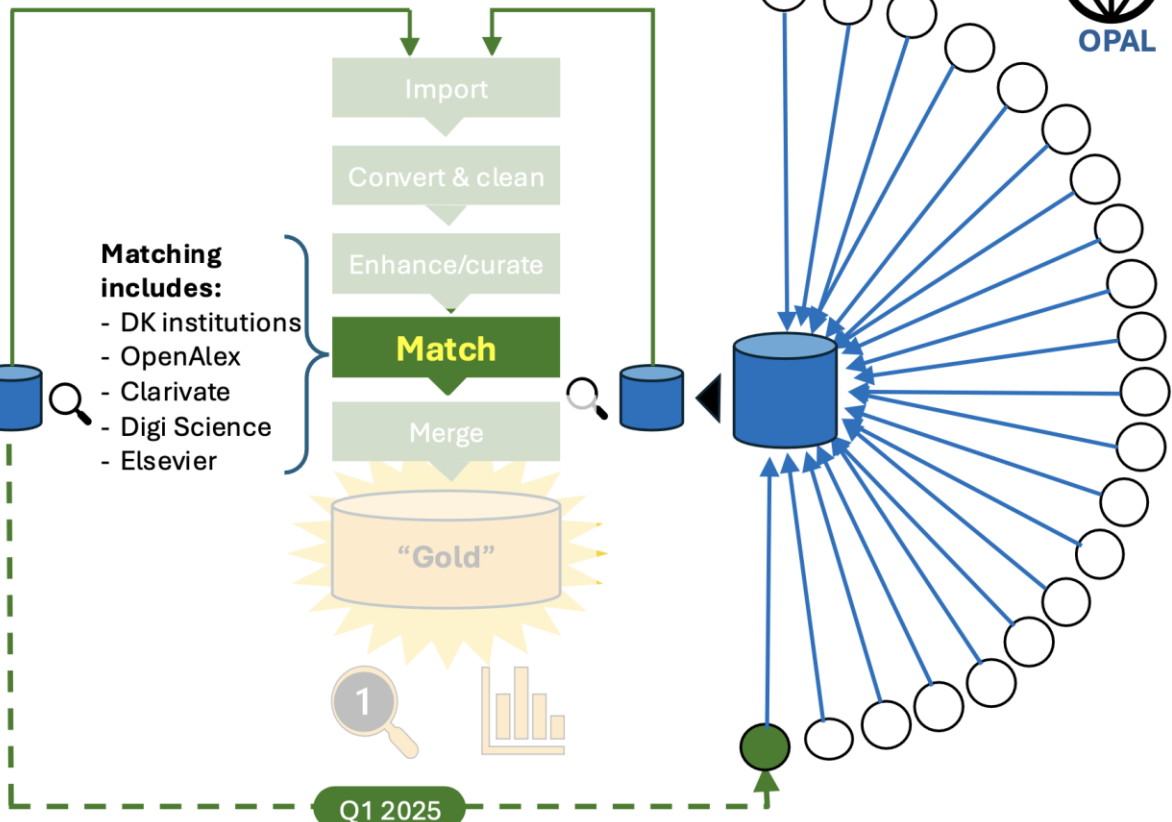
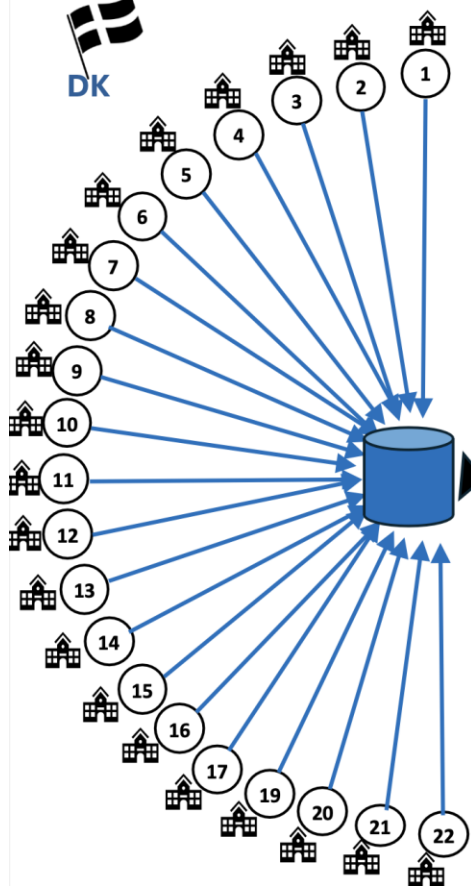
You can focus on recall issues in OpenAlex (missing ROR). This way, only affiliation strings that are NOT matched in OpenAlex to this specific ROR will be retrieved. If several ROR to exclude, separate them by space.

OpenAlex - Norja

- Toteutettu vertailu kansallisen Cristin-dataan (tutkimustietovarantoa vastaava kansallinen toteutus)
- Kaikkiaan vertailtiin n. 250 000 julkaisua
 - Artikkelien osalta erittäin korkea kattavuus -> 99% julkaisuista, joilla DOI, löytyi OpenAlexista
 - Kirjoista (vertailu vain nimitiedoilla) noin 48%
 - Alakohtaista vaihtelua paljon
 - Kielien suhteen norjankielisistä julkaisuista n. 2/3 löytyi myös OpenAlexin tiedoista
 - Eroja erityisesti julkaisujen luokittelussa -> artikkeliksi merkittyjä huomattavan paljon enemmän kuin Cristinissä
 - Duplikaatteja erit. monografioissa, 1/3 norjalaisista monografioista useampaan kertaan OpenAlexin tiedoissa

OpenAlex - Tanska

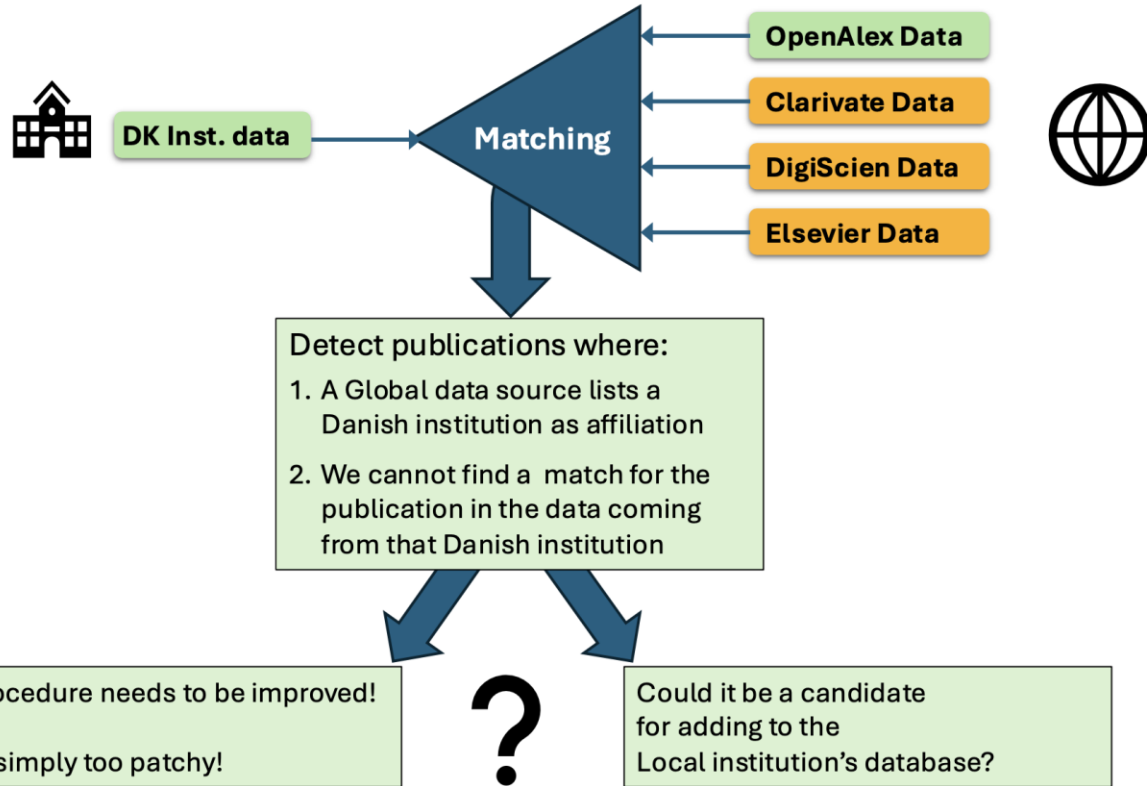
- Tanskassa kansallinen CRIS-toteutus (<https://forskningsportal.dk/>) ryhtyy hyödyntämään OpenAlexin tietoja (CWTSn avustamana)
 - Tarjoaa myös organisaatioille mahdollisuuden hyödyntää OpenAlexista löytyviä organisaatiolle merkittäviä tuotoksia omissa CRIS-järjestelmissään
 - Lisäksi lisäävät kansalliseen avoimen tieteen monitorointiin CWTSn tarjoamia avoimuuden indikaattoreita
- Tarkoitus tuottaa myös väylä päivittää tietoja OpenAlexin suuntaan organisaatioiden omista CRIS-järjestelmistä -> tietojen korjailu/parantelu



Q1 2025

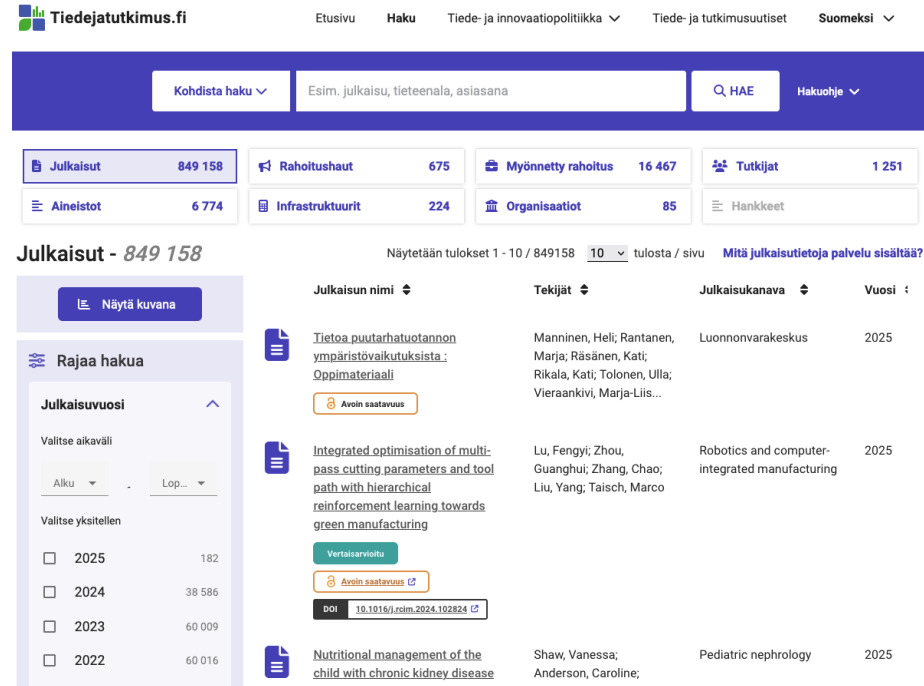


New collaboration opportunities around matching



Tutkimustietovarannon hyödyntäminen?

- Tutkimustietovarantoon kertyy kansallisen julkaisutiedonkeruun myötä n. 60 000 julkaisua vuodessa, selailtavissa Tiedejatutkimus.fi – portaalissa – REST rajapinnan kautta saatavilla
 - Tiedon lähteinä organisaatiot, joissa eri CRIS-järjestelmiä (Pure, Converis, SoleCRIS) tai vastaavia (Justus) käytössä -> julkaisutietojen keruu ja kuratointi silti työntensivistä
 - Tiedon yhdenmukaistamista tehdään tutkimustietovarannossa esim. yhteisjulkaisujen osalta – tarkistuksia myös kuvailutietoihin (koodien oikeellisuus yms.)
 - Tiedoissa ei kuitenkaan viittaustietoja (GraspOS projektissa tutkitaan OpenCitations-palvelun sisältämien tietojen hyödyntämistä)



The screenshot shows the Tiedejatutkimus.fi portal interface. At the top, there is a navigation bar with 'Etusivu', 'Haku', and dropdown menus for 'Tiede- ja innovaatiopolitiikka' and 'Tiede- ja tutkimusuutiset'. A search bar contains the text 'Esim. julkaisu, tieteenalä, asiasana' and a 'HAE' button. Below the search bar, there are statistics for various categories: Julkaisut (849 158), Aineistot (6 774), Rahoitushaut (675), Infrastruktuurit (224), Myönnetty rahoitus (16 467), Organisaatiot (85), Tutkijat (1 251), and Hankkeet.

The main content area is titled 'Julkaisut - 849 158' and shows a list of search results. The first result is 'Tietoa puutarhatuotannon ympäristövaikutuksista: Oppimateriaali' by Manninen, Heli; Rantanen, Marja; Räsänen, Kati; Rikala, Kati; Tolonen, Ulla; and Vieraankivi, Marja-Liis... It is published by Luonnonvarakeskus in 2025. The second result is 'Integrated optimisation of multi-pass cutting parameters and tool path with hierarchical reinforcement learning towards green manufacturing' by Lu, Fengyi; Zhou, Guanghui; Zhang, Chao; Liu, Yang; and Taisch, Marco. It is published by Robotics and computer-integrated manufacturing in 2025. The third result is 'Nutritional management of the child with chronic kidney disease' by Shaw, Vanessa; and Anderson, Caroline, published by Pediatric nephrology in 2025.

Tutkimustietovarannon hyödyntäminen?

- Voisiko näitä tarkistettuja julkaisutietoja viedä edelleen OpenAlexin suuntaan?
- Tiedonsiirto tutkimustietovarannosta on olemassa esim. OpenAIRE-palveluun
 - OpenAIRE ei kuitenkaan sisällä laajasti viittaustietoja
- Organisaatioiden kuratointityö hyödylliseksi myös OpenAlexin käyttäjille
 - Erityisesti affiliaatitiedon ja tekijätietojen osalta tarkat tiedot tutkimustietovarannossa
 - Ei edellytä organisaatioilta lisätyötä
- Mahdollista katsoa OpenAlexin dataa esim. kansallisen julkaisutyypiluokituksen suhteen