



Tutkimustietovarannon ulkoinen rajapintakokonaisuus

IR-tekijäverkoston tapaaminen 11.11.2022
Joonas Nikkanen, CSC



Tutkimustietovaranto - Tiedejatutkimus.fi / Research.fi

- Kokonaisuus käynnistyi vuonna 2017 ja se rahoitetaan osana OKM:n koulutus- ja tiedepolitiikan osaston ja CSC:n välistä vuosisopimusta.
- Tutkimustietovarannon ohjaus perustuu vuodesta 2021 alkaen yhteisen ohjausryhmään, jossa ovat edustettuina tutkimusorganisaatioiden ja muiden keskeisten sidosryhmien edustajat.
- Vuonna 2020 avattu [Tiedejatutkimus.fi](https://tiedejatutkimus.fi)-palvelu on tutkimustietovarannon näkyvin osa, joka kokoaa yhteen tutkimusta kuvaavia tietoja sekä sisältää mm. tilastotietoa tutkimuksen henkilöstön, rahoituksen sekä julkaisujen tieteellisen vaikuttavuuden kehityksestä sekä tietoa suomalaisesta tutkimusjärjestelmästä.
- Lisätietoa hankkeesta [Tutkimustietovaranto-wikisivuilta](#)

Tutkimustietovarannon sisällön hyödyntäminen rajapinnan avulla

- Tutkimustietovarantoon kootut tiedot olleet selattavissa Tiedejatutkimus.fi sivustolla 06/2020 alkaen, laajentuen uusilla kokonaisuuksilla (esim. tutkimusaineistot, avoimet rahoitushaut) ja tietolähteillä (esim. Fairdata-palvelut, ESR-rahoitustiedot)
 - Uusien integrointien teko lähdejärjestelmiin on ollut ja on jatkuvaa toimintaa
- Keskeisenä tavoitteena on avata tiedot koneluettavassa muodossa lukurajapinnan (API) kautta laajemmin hyödynnettäväksi
- Tietojen hyödyntämisen rajapintojen kautta mahdollistaa 1.1.2022 voimaantullut [laki tutkimustietovarannosta](#).
- Tavoiteaikataulu on saattaa ensimmäinen tuotantoversio lukurajapinnasta valmiiksi vuoden 2023 alussa

[Laki tutkimustieto-
varannosta \(1238/2021\)](#)

Tutkimustietovarannon sisällön hyödyntäminen rajapinnan avulla

- Rajapintakokonaisuus rakentuu osissa ja ensimmäisessä vaiheessa mahdollistaa Tutkimustietovarannon tietojen lukemisen ja niiden jatkohyödyntämisen osana muita palveluita
 - Mahdollinen tietojen tuottaminen kirjautusrajapintojen kautta pilotointiin rahoitustietojen osalta vuoden 2023 aikana, muiden tietokokonaisuuksien osalta ei suunniteltu vielä
- Rajapintojen hyödyntäminen on mahdollista
 1. Nykyisiin Tutkimustietovarannon tietokokonaisuuksiin tutkimustoimijoille (=tutkimusta tekevät, rahoittavat ja tukevat organisaatiot)
 2. Tutkijan tietoihin niille tahoille, joille luovuttamiseen tutkija on antanut suostumuksen
- **Tietokokonaisuudet** = tiedot rahoituspäätöksistä, infrastruktuureista, julkaisuista, aineistoista, rahoitushauista ja organisaatioista
- Tutkijan profiilitiedot sekä tutkimusaktiviteetit ovat **lukurajoitettuja** ja niiden käyttö edellyttää tutkijan itsensä antamaa lupaa jakaa näitä tietoja eteenpäin tutkimustietovarannon rajapinnan kautta.
 - Luvittaminen tehdään toimijakohtaisesti ja se tapahtuu tutkijan toimesta Tutkijan tiedot -työkalussa.

Tutkimustietovarannon rajapinnan käyttöoikeusprosessin luonnos

- Tietojen hyödyntämisessä rajapinnan kautta on keskeistä se, onko käytöstä kiinnostunut taho tutkimustoimija vai ei (esim. yritykset)
 - Erillislain määritelmä tutkimustoimijasta laeva 2 § “Tässä laissa tarkoitetaan: *tutkimustoimijalla* opetus- ja kulttuuriministeriötä, tutkimusta tekeviä organisaatioita sekä tutkimusta rahoittavia ja tukevia organisaatioita”
- Tutkimustoimija saa käyttöoikeuden myötä
 - Pääsyn Tutkimustietovarannon **tietokokonaisuuksiin** ja
 - Tarvittaessa luvituksen jälkeen myös tutkijakohtaisiin **lukurajoitettuihin** kokonaisuuksia
- Jos kyseessä ei ole tutkimustoimija, on mahdollisuus hakea käyttöoikeutta vain luvitettaviin **lukurajoitettuihin** kokonaisuuksiin

Tutkimustietovarannon rajapinnan käyttöoikeusprosessin luonnos

1. Käyttöoikeutta pyydetessä tahon tulee toimittaa seuraavat tiedot tutkimustietovarantoon:
 - Käyttötarkoitus
 - Yhteyshenkilö
 - Tarvitseeko lukurajoitettuja (=luvitettavia) tietoja
2. Tehdään päätös opetus- ja kulttuuriministeriössä (tai epäselvissä tapauksissa viedään keskusteluun tutkimustietovarannon ohjausryhmään) onko taho a) tutkimustoimija b) muu taho, jolle käyttöoikeus voidaan myöntää (mahdollisen väärinkäytön esto)
3. Tutkimustietovarannon puolella tehdään tarvittavat avaukset, luodaan rajapintaan käytettävät pääsyavaimet ja toimitetaan ne takaisin taholle
4. Taho alkaa käyttää rajapintaa ja hyödyntää **tietokokonaisuuksia**
Lukurajoitetut kokonaisuudet vain tutkijan luvitettua ne profiilityökalussa

Tietojen luvitus

- **Lukurajoitettuja kokonaisuuksia** varten rajapintaan ja profiilityökaluun rakentuu suostumuksenhallinta, jonka avulla tutkija voi itse:
 - nähdä mitä tietoja hänestä on kertynyt tutkimustietovarantoon eri organisaatioista
 - luvittaa omia tietojaan 3. osapuolille
 - hallita tietojensa siirtymistä palveluiden välillä
-> Mydata / Omadata ajattelun toteutuminen tutkijoille & asiantuntijoille
- Tietojen luvitus pyritään rakentamaan osaksi Tutkijan tiedot –työkalua sekä rajapintakokonaisuutta vuoden 2023 aikana

Tiedot ja tietolähteet

Seppo Sammakko

0000-0002-0002-0963

Rajaa

Tila profiilissa

- Julkinen 19
- Ei julkinen 19

Tietokokonaisuus

Hae

- Muut nimet 3
- Sähköposti 1
- Linkit 2
- Avainsanat 1
- Kuvaus 1
- Affiliaatiot 3
- Koulutus 1

[Näytä enemmän](#)

[Näytä vähemmän](#)

Lähde

- ORCID (test) 17
- Tiedejatutk... (test) 21

0 valittu

<input type="checkbox"/>	Tieto	Sisältö	Tiedejatutkimus.fi	Lähde
<input type="checkbox"/>	Affiliaatiot	Tutkija, Tampereen yliopisto	Julkinen	ORCID (test)
<input type="checkbox"/>	Koulutus	Frogdesign	Julkinen	ORCID (test)
<input type="checkbox"/>	Julkaisut	Frog community compo...	Näytä	Ei julkinen ORCID (test)
<input type="checkbox"/>	Julkaisut	Detection of cultured bre...	Näytä	Julkinen ORCID (test)
<input type="checkbox"/>	Julkaisut	Congolius, a new genus ...	Näytä	Ei julkinen ORCID (test)
<input type="checkbox"/>	Julkaisut	Antioxidant potential an...	Näytä	Ei julkinen ORCID (test)
<input type="checkbox"/>	Julkaisut	Frogpaper	Julkinen	ORCID (test)
<input type="checkbox"/>	Julkaisut	Literacy instruction activ...	Näytä	Ei julkinen Tiedejatutkimus.fi (test)
<input type="checkbox"/>	Julkaisut	Fermion mass hierarchy ...	Näytä	Julkinen Tiedejatutkimus.fi (test)
<input type="checkbox"/>	Julkaisut	Natural neutrino sector i...	Näytä	Ei julkinen Tiedejatutkimus.fi (test)

< Edellinen

1

2

3

4

Seuraava >

Mitä dataa rajapintakokonaisuudesta on mahdollista saada?

- Tutkimustoimijoille käyttöoikeusprosessin myötä avoimet Tutkimustietovarannon **tietokokonaisuudet**
- Tutkijan tiedot ovat **lukurajoitettuja**, vain tutkijan toimijalle luvittamana profiilyökalun kautta

Tutkimustietovarannon tietokokonaisuudet

Rahoitus-
päätökset

Aineistot

Rahoitus-
haut

Infrastruk-
tuurit

Julkaisut

Organisaa-
tiot

= se mikä nähtävillä Tiedejatutkimus.fi

Tutkijan tiedot - kokonaisuus

Profiili-
tiedot

Affiliaatio

Yhteys-
tiedot

Koulutus

Aktiviteetit

Meriitit

= se mitä tulee profiileihin Tiedejatutkimus.fi
(pl. käyttäjän piilottama sisältö)

Tietokokonaisuus	Kuka toimittaa tietoja?	Miten tietoja toimitetaan?	Milloin tietojen toimittaminen on mahdollista?
Julkaisut	Tutkimusorganisaatiot	VIRTA-julkaisutietopalveluun XML-tiedostona tai JUSTUS-palvelun kautta	Käytössä
Rahoituspäätökset	Rahoittajat	XML-siirtotiedostona tutkimustietovarantoon	Käytössä
Rahoitushaut	Rahoittajat	Suorat tiedonsiirrot järjestelmätoimittajilta, sähköisellä lomakkeella	Suoria tiedonsiirtoja viimeistellään, sähköinen lomake tuotannossa
Tutkimusaineistot	Tutkimusorganisaatiot, aineistopalvelut	Fairdata Metax -rajapinnan kautta, haravointi Metaxiin tai tietojen syöttäminen Qvain-palvelussa	Käytössä
Tutkimusinfrastruktuurit	Tutkimusorganisaatiot, infrastruktuurien vastuuorganisaatiot	Infrastruktuurien syöttölomakkeella tai siirtotiedostona järjestelmistä	Käytössä
Tiedeutiset	Tutkimusorganisaatiot, rahoittajat ym. tiedettä tukevat organisaatio	Haravoidaan RSS-syötteestä	Käytössä
Hankkeet (esim. tutkimusprojektit)	Tutkimusorganisaatiot	Siirtotiedostoina tutkimustietovarantoon	Pilotointi vuoden 2023 aikana
Tutkijan tiedot -kokonaisuus	Tutkimusorganisaatiot (CRIS-järjestelmät, Justus), tutkijat (ORCID-palvelu)	XML-siirtotiedostona tutkimustietovarantoon tai JUSTUS-palvelun kautta (tulossa), Yksittäiset tutkijat voivat tuoda tietoja ORCID-profiileistaan	Pilotointi käytössä, betaversio vuoden 2022 marraskuussa. Tuotanto keväällä 2023

Käyttötapaukset tietojen hyödyntämiselle

- Käyttötapauksia on kerätty tutkimustietovarantohankkeen edetessä
- Tutkimustietovarannon tietokokonaisuuksista kiinnostusta ovat herättäneet eniten rahoituspäätökset ja rahoitushaut sekä tutkijan tiedot
- Seuraavilla kalvoilla muutamia esimerkkejä kertyneistä käyttötapauksista

Tietokokonaisuus	Käyttötapaus	Tarkempi kuvaus
Rahoituspäätökset	Tekstipohjainen vapaahaku tutkimushankkeista ja aineistoista	Meillä yksi mahdollinen käyttötapaus liittyisi tekstipohjaiseen hakuun tutkimushankkeista Rajapinnan kautta tulisimme todennäköisesti myös hakemaan tätä laajempaa dataa omalle tietoaalustallemme (Data Platform)
Rahoituspäätökset	Päätösten vienti sisäiseen tietovarantoon	Järjestelmä hakee tiedot, viehdään heidän sisäiseen tietovarantoon ja sieltä edelleen tutkimustietojärjestelmään rikastamaan nykyisiä tietoja
Rahoitushaut	Hakujen ylläpito	Ylläpidetään sivustoa, johon haetaan tällä hetkellä tietoja RSS-syötteen avulla Aurora-tietokannasta.
Tutkijan tiedot	Liittyvän tutkimuksen koonti	Kiinnostuneita hyödyntämään TTV:oon koottuja tietoja aivoihin liittyvistä tutkimuksesta ja tutkijoista omassa palveluissaan. Tietoa käytetään ainakin Neurokeskuksen verkkosivuilla, jonne on koottu tietoja alan tutkijoista
Julkaisut	Liittyvän tutkimuksen koonti	Lukuisia verkkosivustoja, joilla olisi tarve esittää erilaisiin aihepiireihin liittyviä tutkimuksia

Tietokokonaisuudet	Käyttötapa	Tarkempi kuvaus
Organisaatiot	Ajantasaisen tiedon ylläpito	Olisi tarve esim. tuottaa ajantasainen koodisto esim. organisaatioista, jotka TTV:llä käytössä. Järjestelmä hakee tiedot, päivittävät näillä järjestelmän koodistoja ja/tai käyttävät ensisijaisina koodistoina, vievät koodistot käytettäväksi käyttäjien hyödynnettäväksi
Useita tietokokonaisuuksia	Omien tietojen rikastaminen ja ylläpito	Rajapintapalvelu kiinnostaisi meitä erityisesti, sillä meillä on tarvetta hyödyntää rajapintoja ylläpidon lisäksi tietojen noutamisessa omalle sivustollemme. Erityisesti meitä kiinnostaisi, josko rajapinnan kautta saisi jatkossa rakenteisessa muodossa tutkimusjulkaisut, tutkimushankkeiden ja mahdollisesti tämän vuoden loppupuolella myös tutkijaprofiilien tiedot. Näistä tiedoista voisi siis muodostaa dynaamisesti verkkosivuja.
Useita tietokokonaisuuksia	Oman tietovarannon rikastaminen	Käyttävät esim. VIRTJA-jtp ja Metax rajapintoja, ovat olleet aktiivisia seuraamaan mitä / milloin TTV-rajapintaa voisi käyttää. Oma järjestelmä hakee tiedot rajapinnasta ja ne vievät heidän sisäiseen tietovarantoon ja sieltä edelleen tutkimustietojärjestelmään

Tutkimustietovarannon sisällön hyödyntäminen rajapinnan avulla – tekninen toteutus

- Teknisesti kokonaisuus rakentuu REST-rajapinnaksi, johon rakentuvilla GET-kutsuilla voidaan hakea tutkimustietovarannon sisältämiä tietoja käyttäen rajaavia parametreja
 - Tieto tuotetaan tutkimustietovarannon tietomallin mukaisesti ja JSON-formaatissa
 - Suunniteltu myös GraphQL pohjaista toteutusta uloisen rajapinnan osalta tukemaan erityisesti linkitetyn datan jatkohyödyntämistä
- Rajapinnan käyttöoikeuden myötä toimija luovutetaan pääsyavain (*token*), jolla tietoja rajapinnasta voidaan hakea määriteltä käyttötarkoitusta varten
- Rajapinnan käyttöä tulee tukemaan seuraavat kokonaisuudet:
 - Swagger-dokumentaatio: haku- ja rajauseroimien kuvailu, tiedon muodon tutkiminen ja rajapinnan testaaminen - avataan pilotoinnin yhteydessä
 - Tietomallit.suomi.fi –palvelussa sijaitsevat tietomalli: tiedon muodon ja määrittelyiden tutkiminen - <https://tietomallit.suomi.fi/model/ttv>
 - Wiki-sivu: käyttöoikeuden haku, tekniset edellytykset ja yleiskuvaus - <https://wiki.eduuni.fi/x/PA6XE>

Tutkimustietovarannon lukurajapinnan aikataulusta

- Lukurajapintaa pilotoitu rahoitukseen liittyvillä tiedoilla (avoimet rahoitushaut, rahoituspäätökset) kesän 2022 kuluessa
 - Tavoitteena oli saada palautetta teknisestä toteutuksesta, autentikoinnista, tiedon muodosta yms.
 - Organisaatioita mukana vain vähän - data ei vielä jatkuvasti päivittyvää / ei kytkettävissä järjestelmiin
- Seuraava etappi on julkaista ensimmäinen tuotantoversio lukurajapinnasta vuoden 2023 alussa
 - Sisältää rahoitukseen liittyvät tiedot (sekä päätökset että avoimet haut), aineistotiedot, julkaisutiedot
 - Ns. 1.0 versio, johon päivityksiä ajetaan ja tietokokonaisuuksia lisätään mukaan (esim. infrastruktuurit, organisaatiot)
 - **Lukurajoitettu kokonaisuus** ja tiedon luvittaminen pyritään lisäämään vuoden 2023 aikana (liittyy profiilityökalun kehityksen etenemiseen)
 - Huomioidaan taaksepäin yhteensopivuus, eli rajapinnan jo olemassaolevia endpointteja ylläpidetään päivityksistä huolimatta