

Ohjeita pysyvien tunnisteiden käyttöönottoon

TYÖDOKUMENTTI

Askelmerkkejä PID:ien käyttöönottoon

A. Vaatimusten kartoittaminen

- ulkoiset vaatimukset
 - käyttäjien tarpeet
 - kansalliset linjaukset
 - rahoituspäätökset? TTV?
 - Organisaatiot
 - esim EU-direktiivit
 - alakohtaiset tms standardit ja käytännöt
- mille tarvitaan tunnisteet ja mitä näistä ylläpidetään itse?
 - datasetti = tutkimusaineisto
 - DOI
 - tietokantakysely
 - DOI (dynaamisten tietokantojen staattisille osajoukoille)
 - tiedosto /bitstream
 - URN (PAS-palveluun siirrettävien julkaisujen tiedostot)
 - tutkimusinfrastruktuurit (data, equipment, service; <https://koodistot.suomi.fi/codescheme;registryCode=research;schemeCode=ServiceTypeCode>)
 - laite, sensori (ulkoinen?)
 - handle? <http://www.handle.net/>
 - TTV infra tunnisteet URN
 - tietokanta/tietojärjestelmä/ohjelmisto
 - TTV infra tunnisteet URN
 - palvelu?
 - projekti (ulkoinen)
 - RAID?
 - rahoituspäätös (ulkoinen)
 - TTV?
 - palvelu?
 - tutkijat
 - ORCID (ulkoinen)
 - ISNI (ulkoinen)
 - opiskelijat
 - Oppijnumero (Opetushallitus)
 - Julkaisut (ulkoinen)
 - ISBN, ISSN, ...
 - DOI
 - Organisaatiot (ulkoinen)
 - ISNI
 - RoR
 - GRID
 - (Y-tunnus)/Business-ID
 - (EU) PIC
 - Sanastot- ja ontologiat (ulkoinen?)
 - URN
 - finto URI
 - laji (tun) URI
 - Muuttujat (ulkoinen?)
- palvelutaso (SLA)
 - saavutettavuusvaatimus
 - back up/peilit
 - skaalautuvuus
- toiminnallisuus
 - resoluutivuus

B. Ratkaisujen vertaileminen

- tietomallit
- kustannukset
 - palveluun liittyminen/palvelun pystyttäminen
 - Ylläpito
 - Lyhyen ja pitemmän aikavälin kustannus
 - elinkaari/exit-mahdollisuus
- luotettavuus
 - teknologian ylläpidettävyys ja teknisen migraation mahdollisuus
 - mahdollisen ulkopuolisen palvelun toimittajan vakaus ja uskottavuus

C. Ylläpidon vastuut

- resolvointi
- tunnisteiden hallinnointi
- aineiston kuratointi

D. Eri tunnistejärjestelmät

Taulukko erillisellä sivulla

- (<https://koodistot.suomi.fi/codescheme;registryCode=research;schemeCode=PID>)

E. Käyttötapauksia

Hyödyt tunnisteiden käytöstä:

Olisiko ideaa niputtaa osa alla olevista hyödyistä FAIR-periaatteen alle?

1. Datan laatu, saatavuus ja käytettävyys jatkossa
2. Organisaatioiden julkaisutiedonkeruu
 - a. ORCID: tutkijalle meriittiä (Tutkimustietovaranto)
 - b. ISBN/ISSN tieteellisen (A- ja C -tyypit) julkaisutyypin edellytys => meriittiä sekä tutkijalle että organisaatiolle
3. Organisaatioiden julkaisuarkistot
 - a. Pysyvä linkki julkaisuun (yleensä URN?), lisäksi julkaisutietueessa DOI (mikäli julkaisulla on)
4. Altmetriikka:
 - a. DOI tai muu pysyvä tunniste oleellista mainita aina, kun postaa someen uutisen tutkimustuotoksestaan
 - b. esim. DOI:n avulla voidaan saada tietoa julkaisujen some-huomiosta esim. Altmetric.com -palvelussa
 - c. esim. rahoitushakemuksissa tutkimusryhmän tuottamien julkaisujen saama some-huomio on osoitus tutkimuksen yhteiskunnallisesta vaikuttavuudesta
5. Tutkimuksen toistettavuus viittausten avulla
6. Dataan viittaaminen antaa meriittiä
 - a. tekstin sisäiset viittaukset
 - b. dataviittaus referenssiliistassa koneluettava
 - c. viittaus dataan lisää viittauksia julkaisuun (Piwowar, H.A., Day, R.S., Fridsma, D.B. [Sharing detailed research data is associated with increased citation rate \(2007\) PLoS ONE, 2 \(3\), art. no. e308](#))
 - d. viittauksista on hyötyä tutkijalle ja organisaatiolle
7. Tiedon jäljitettävyys ja alkuperä näkyväksi
8. Kansainväliset tietokannat
 - a. Scopus/Elsevier: ORCID helpottaa tutkijan tunnistamista (varsinkin, jos Scopus-profiili hajonnut)
 - b. Data Citation Index/Clarivate: tietokanta, josta voi hakea dataviittauksia
9. Hyöty tutkijalle
 - a. nopeuttaa rahoitushakemusten laatimista
 - b. nopeuttaa julkaisujen submittausta
 - c. on osoitus tutkimusprojektin hyvästä hallinnoinnista ja johtamisesta
10. Tukee avointa tiedettä ja avointa vertaisarviointia. Plan S:n mukainen
11. Maksuttomuus
12. Kuuluu hyvään tieteen tekemiseen

PID politiikat

Oulun Yliopisto: Oulun yliopiston suositus pysyvistä tunnisteista (PID) tutkimusaineistoille <https://www oulu.fi/yliopisto/node/44005>

CSC

Kansainväliset politiikat, organisaatioiden ym. politiikat

- The UK National PID Consortium: A Pathway to Increased Adoption <https://scholarlykitchen.sspnet.org/2020/06/29/the-uk-national-pid-consortium-a-pathway-to-increased-adoption/>
- EOSC PID Policy Europe <https://zenodo.org/record/3780423#.Xvq9XkxuK3B>, liitteen PID-sanasto

Mille tarvitaan tunnistetunnisteita

Julkaisut

- jo yleisesti vakiintunut käytäntö varsinkin kv-julkaisuilla (DOI)
- monografialle ISBN (linkki oppaaseen)
- sarjajulkaisut ja lehdet: ISSN
- kotimaisilla (tiede)julkaisuilla tunnistekäytäntö vielä osin vakiintumaton
- ORCIDin käyttökin yleistyy hyvin kustantajien (ja tutkimusrahoittajien) vaatimuksesta
- DOI:n käyttö julkaisujen näkyvyyden (altmetriikka) edistämiseksi oleellista
- yliopistojen CRIS-järjestelmiin kerättävät tunnisteet julkaisujen osalta

Infrat

- <https://dwidrihfe.csc.fi/> tutkimusinfrastruktuurien tietojen toimitus tutkimustietovarantoon
- UOulussa osalle infoja jo hankittu URN, osalle hankinnassa. Ohjausryhmä, jos vie asioita eteenpäin.
- tunniste mahdollistaa viittaamisen

Tutkimusaineistot (data)

- UOulussa tutkimusaineistokataloogi-hanke ideoinnissa.
- Mihin viitataan, jos datalla on useampi tunniste?
- Koneluettavat aineistohallintasuunnitelmat: esim. DMPOnlineassa testissä RoR-tunnisteiden integrointi systeemiin. Testissä myös DOI-integraatio. DOI DMP:lle? Julkaisujen DOI:t DMP:hen?
- Kansallisista FAIRdata-palveluista saatavat tunnisteet: QVaimesta URN. DOI vain IDA:SSA oleville aineistoille? Miksiköhän näin?? URN, jos julkaistaan vain metadataa.

Instrumentit

- <https://www.rd-alliance.org/groups/persistent-identification-instruments-wg>
- <https://github.com/rdawg-pidinst/schema>
- Stocker, M., Darroch, L., Krah, R., Habermann, T., Devaraju, A., Schwardmann, U., D'Onofrio, C. and Häggström, I., 2020. Persistent Identification of Instruments. *Data Science Journal*, 19(1), p.18. DOI: <http://doi.org/10.5334/dsj-2020-018>
 - "...Currently, there is no globally implementable way to persistently identify measuring instruments. Addressing this challenge, the present article describes the results of the work conducted by PIDINST, an 18 month RDA Working Group that aimed at establishing a cost-effective, operational solution based on existing PID infrastructures, combined with a robust metadata schema for accurate identification, retrieval and automation into workflows. The solution was demonstrated at two institutional instrument providers...."

Sanastojen, tesaurosten, ontologioiden tunnisteet

- termikohtaiset? käsitekohtaiset? käytetäänkö missään?? mitä hyötyä voisi olla?? käykö ilmi käytetty sanasto pelkän tunnisteiden perusteella?? Fintossa on YSON tunnisteet.

Käyttötarkoitus, mihin tietoa kerätään, viittaamien?

- julkaisutietojen, datojen, infojen, tutkijoiden, organisaatioiden yksiselitteinen yhdistäminen toisiinsa
- infraan/tutkimusaineistoon liittyvien julkaisujen ja julkaisujen saamien viittausten mappaus tietokannoista
- datametriikka kuten views, downloads, citations

Rajoitukset

- monenmoiset järjestelmät?
- tiedon/osaamisen/motivaation puute?
- osa tunnisteista - kuten ORCID - vaatii tutkijalta omaa aktiivisuutta tunnisteiden luomisessa ja ylläpidossa
- useita tunnisteita samasta asiasta/samalle asialle: esim. tutkijalla voi olla useita ORCID-tunnisteita
- lopullisen tunnisteiden saamisessa viive esim. julkaisulla (voi olla viittauksia + muuta huomiota jo ennen pysyvää tunnistetta) > tunnisteiden varaamismahdollisuus tärkeä

Kustannukset

Saatavuusvaatimukset

Muut vaatimukset, esim metatiedot, näkyvyys, jäljitettävyys?

- erilaisten tunnisteiden haettavuus/hyödynnettävyys tietokannoissa ei läheskään kattavaa

Esimerkkejä:

Tutkimusorganisaation tutkimusaineistojen säilytyspalvelu ja sen metatietovaranto

- DataverseNL: Online storage, sharing and registration of research data, during the research period and after its completion <https://dataverse.nl/>

SYKEN metatietojärjestelmä ja data-arkisto CKAN (Hanna)

- Nytkyinen toteutus/ arkkitehtuuri
- DOI:n käyttöönotto
- SYKEN tutkimuksen tuottama data
- Käytön myötä havaittuja kehittämiskohteita / puutteita / tarpeita

Toiveita

Ohjeita metatietojen tekemisestä, esimerkkejä!!

CrediT taxonomy

<https://casrai.org/credit/>:

Conceptualization
Data curation
Formal Analysis
Funding acquisition
Investigation
Methodology
Project administration

Resources
Software
Supervision
Validation
Visualization
Writing – original draft
Writing – review & editing

DataCite:

If Contributor is used, then contributorType is mandatory.

Controlled List Values:

ContactPerson
DataCollector
DataCurator
DataManager
Distributor
Editor
HostingInstitution
Producer
ProjectLeader
ProjectManager
ProjectMember
RegistrationAgency
RegistrationAuthority
RelatedPerson
Researcher
ResearchGroup
RightsHolder
Sponsor
Supervisor
WorkPackageLeader
Other

Tällä hetkellä valmiit ratkaisut eivät tarjoa kovin hyviä mahdollisuuksia PIDien käyttöön.

M.m. sensitiiviset datat aiheuttavat tarvetta omiin ratkaisuihin.

Linkitykset voi tehdä myös CRIS-järjestelmässä.

THL aineistokatalogi

Ohjelmistoista:

<https://ngr.coar-repositories.org/>

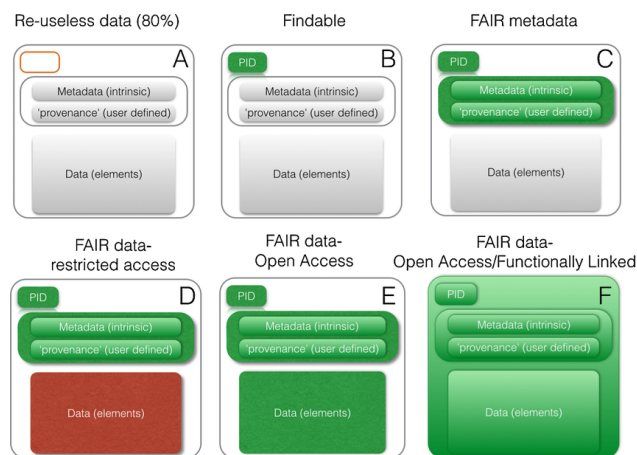
<https://dataverse.org/>

<https://documents.egi.eu/public/RetrieveFile?docid=3419&filename=EOSC-hub%20D2.8%20v1%20Approved%20by%20EC%20Public.pdf&version=5>

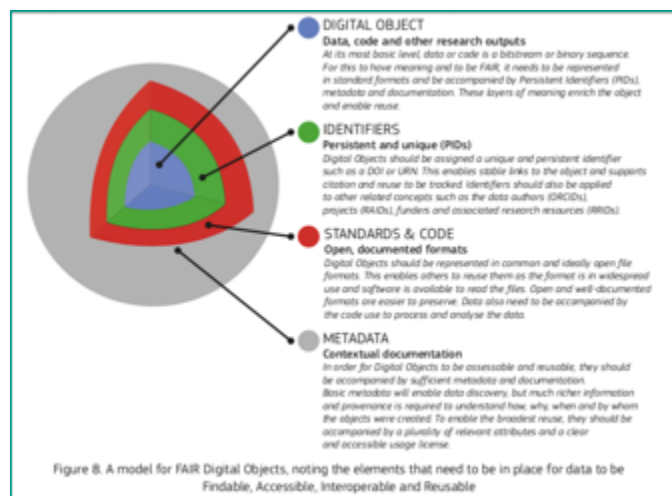
Tutkimusprojektin oma relaatiotietokanta

FAIR Data Objects

Data as increasingly FAIR Digital Objects



Mons, Barend & Neylon, Cameron & Velterop, Jan & Dumontier, Michel & Bonino da Silva Santos, Luiz Olavo & Wilkinson, Mark. (2017). Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use. 37. 1-8. <https://doi.org/doi:10.3233/ISU-170824>



TFIR <https://doi.org/doi:10.2777/1524>

DataCite relation types

https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf

IsCitedBy

Cites

IsSupplementTo

IsSupplementedBy

IsContinuedBy

Continues

IsDescribedBy

Describes

HasMetadata

IsMetadataFor

HasVersion (DC)

IsVersionOf (DC)

IsNewVersionOf

IsPreviousVersionOf

IsPartOf (DC)

HasPart (DC)

IsReferencedBy (DC)

References (DC)

IsDocumentedBy

Documents

IsCompiledBy

Compiles

IsVariantFormOf

IsOriginalFormOf

IsIdenticalTo

IsReviewedBy

Reviews

IsDerivedFrom (DC: Source)

IsSourceOf

IsRequiredBy (DC)

Requires (DC)

IsObsoletedBy (DC: IsReplacedBy)

Obsoletes (DC: Replaces)

**

Asiat, joille on/saattaa olla tunnisteet valmiina (sisällytetään, varattava paikka tietojärjestelmässä)

- aineiston tuottaneet henkilöt
- liittyvät julkaisut
- liittyvät muut aineistot
- liittyvät infrastruktuurit ja laitteet

- instrumentit
- liittyvät hankkeet/projektit
- liittyvät rahoittajat
- liittyvät rahoituspäätökset
- liittyvä omistaja/hallinnoija
- liittyvät tiedostot /jakelut

Asiat, joille järjestelmä tuottaa mastertiedot sekä pysyvän tunnisteiden:

- aineistot (metatiedot, sis. tiedostojen tunnisteet/linkit)
- aineistokokeilut?
- aineistoversiot