

Pysyvien tunnisteiden kansallinen tiekartta

Jäljitettävä ja yksilöity tieto (objekteista) on
löydettävissä ja luotettavasti linkitettävissä
nyt ja tulevaisuudessa

<https://urn.fi/URN:NBN:fi-fe2023042138021>

Tiekartan on laatinut PID-verkosto. Lisää tiekartan kirjoitusprosessista ja siihen osallistuneista [verkoston kotisivuilla](#).

Tiivistelmä

Tässä dokumentissa on kuvattu pysyvien tunnisteiden kenttää ja toimijoita Suomessa. Tiekartan ensimmäinen versio on keskittynyt erityisesti tutkimuksen tunnisteisiin, mutta sen ehdottamat toimet ovat laajennettavissa myös laajempaan datanhallinnan kenttään.

Tiekartassa ehdotetuilla toimilla tavoitellaan seuraavaa [Suomen PID-verkoston](#) asettamaa tavoitetilaa: "Jäljitettävä ja yksilöity tieto (objekteista) löytyy ja on luotettavasti linkitettävissä nyt ja tulevaisuudessa".

PID-verkosto ehdottaa keskeisesti seuraavia toimia tavoitetilän saavuttamiseksi kansallisesti:

Tavoite	Kuvaus	Toimenpiteet
Luottamuksen ja yhteistyön rakentaminen	Opitaan tuntemaan toisia toimijoita, jaetaan tietoa ja verkostoidutaan	-rakennetaan yhteistyötä toimijoiden välillä PID-asioihin liittyen, mm: - tutkimusyhteisö - PID-palvelut - KAM -sektori - OKM, DVV, VM - linkitetyn datan yhteisö - tiedon laatuun liittyvä työ -tuotetaan PID-politiikka -selkeytetään tarpeita ja hyötyjä
Kansallisen infrastruktuurin järjestäminen	Organisoidutaan, suunnitellaan ja varmistetaan rahoitus	-kehitetään PID-työn alustaa -tunnistetaan kehittämistarpeet -sovitaan vastuut -rahoituksen selvittäminen
Tietoisuuden ja osaamisen lisääminen	Huolehditaan tiedon jakamisesta ja koulutuksista	-viedään PID-asioita mm. AVOTT-koulutus ryhmien agendalle -kehitetään kansainvälistä verkostoitumista ja toimintaa järjestelmällisesti

Tiekarttaa tullaan päivittämään yhdessä toimijoiden kanssa.

Sisällys

Tiivistelmä	2
Johdanto	5
Pysyviin tunnisteisiin liittyvät peruskäsitteet	6
Pysyvyys	6
Ainutkertaisuus	6
Käyttökonteksti	6
Resoluutio ja toiminnallisuus	7
Semanttisuus	8
Kate	9
Metatietomalli	10
Tarve kansalliselle pysyvien tunnisteiden politiikalle	11
Pysyvät tunnisteet osana julkista tiedonhallintaa	11
Tutkimuksen pysyvät tunnisteet	15
Tutkimuksen erityispiirteet	15
Toimijat ja roolit tutkimuksen käyttökontekstissa	16
PID-auktoriteetit	17
Konsortiot	18
Muut	18
Tutkimuksen pysyvien tunnisteiden järjestelmät ja katteet	18
Tutkijat	18
Väitöskirjat	18
Tieteelliset artikkelit	19
Tutkimusaineistot	19
Lähdekoodi, työvuot ja muut menetelmät	20
Sanastot, muuttajat ja koodistot	20
Organisaatiot	21
Rahoituspäätökset	21
Rahoituspäätökset Tutkimustietovarannossa	21
Euroopan unionin rahoituspäätökset	21
Kansalliset rahoituspäätökset	21
Tutkimushankkeet	22
Palvelut ja infrastruktuurit	22
Tavoitetila ja tarvittavat askeleet sen saavuttamiseksi	23
Luottamuksen ja yhteistyön rakentaminen	24
Luottamusta lisäävät asiat	24
Tarvittavat toimenpiteet	25
Kansallisen infrastruktuurin järjestäminen	25
Tarvittavat toimenpiteet	27
Tietoisuuden ja osaamisen lisääminen	27
Tarvittavat toimenpiteet	28

Johdanto

Tämän tiekartan on laatinut [Suomen PID-verkosto](#), joka on avoin asiantuntijaverkosto pysyviä tunnisteita hallinnoiville tahoille.

Hyvin hallintoitujen pysyvien tunnisteiden (PID) avulla objekti on aina löydettävissä, eriteltävissä, tunnistettavissa ja jäljitettävissä. Pysyvät tunnisteet mahdollistavat luotettavan tiedon hakemisen ja löytymisen, vaikka teknologinen ympäristö tiedon tarjoamiseksi muuttuu. Toimivasta pysyvien tunnisteiden hallinnoinnista ja tarkoituksenmukaisesta käytöstä on hyötyä tieteelle ja tutkimukselle tutkimuksen toistettavuutta ja tarkkaa dokumentaatiota tukevin toimintoina. Lisäksi ne tukevat tutkijoiden ja muiden toimijoiden meritoitumista, sillä tutkimustuotosten ja resurssien hyödyntämistä ja uudelleen käyttöä pystytään seuraamaan pysyvien tunnisteiden avulla. Oikein käytetyt tunnisteet tekevät viittaamisesta yksiselitteistä ja luotettavaa.

Pysyviä tunnisteita voi käyttää myös toistettavien, toiminnallisten prosessien kuten esimerkiksi konversion tai analyysin dokumentaatioon ja toistamiseen. Lisäksi tutkimustiedon hallinta tehostuu huomattavasti.¹ Saman voi olettaa pätevän kaikkeen hallinnollisen tiedon hallintaan. Tiedon elinkaareen seuraaminen ja dokumentaatio on edellytys tiedon pitkäaikaissaatavuuden varmistamiseksi. Tunnisteisiin liittyvät metatiedot mahdollistavat tiedon löydettävyyden, jakelun ja integroinnin sekä erilaisten palveluiden tarjoamisen.

Pysyvän tunnisteiden voi antaa eri tyyppisille objekteille, kuten:

- Teoksille tai tietosisällöille erilaisina kokonaisuuksina, esimerkiksi hallinnollisille päätöksille tai niiden osille.
- Digitaalisille objekteille, esimerkiksi tiedostoille, ohjelmistoille, tietokantakyselyille tai tietoaaineistolle; näiden tieto-objektien osajobjekteille sekä niitä koskeville metatiedoille erilaisina kokonaisuuksina
- Fyysisille objekteille, esimerkiksi henkilöille, organisaatioille, tutkimuslaitteille ja niiden osille kuten lämpötila-antureille.
- Käsitteille ja termeille, jotka voi tallentaa koneluettavaan muotoon siten, että myös käsitteiden väliset relaatiot ovat koneymmärrettäviä.

Tämä tiekartta on suomalaisten asiantuntijoiden yhdessä laatima dokumentti. Työ on tehty avoimen PID-verkoston puitteissa vuosina 2021 ja 2022. Katsomme, että pysyvät tunnisteet ovat kestävä ja tehokkaan tiedonhallinnan ytimessä ja ne vaativat huomiota osana kansallista tietoinfrastruktuuria - ei vähiten tutkitun ja luotettavan tiedon merkityksen kasvaessa.

Tämä dokumentti keskittyy erityisesti tutkimuksen tunnisteisiin. Tiekartalla esitettyä on tavoitteena laajentaa myös muuhun tiedonhallinnan kenttään sopivin osin.

Tunnisteisiin ja niiden hallintaan liittyy käsitteitä, joita käsittelemme seuraavassa luvussa. Tämän jälkeen kuvaamme tämän hetkistä tilannetta Suomessa - käytössä olevia

¹ Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022). Revised cost-benefit analysis for the UK PID Support Network (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7356219>

PID-tunnistejärjestelmiä ja kentällä toimivia toimijoita. Lopuksi ehdotamme toimenpiteitä dokumentin alkuun määrittelemämme tavoitetilan saavuttamiseksi.

Pysyviin tunnisteisiin liittyvät peruskäsitteet

Tässä luvussa kerrotaan pysyvistä tunnisteista niihin liittyvien peruskäsitteiden avulla, ja määritellään miten terminologiaa käytetään tässä dokumentissa. Käsitteiden järjestys on siksi rakentunut tässä siten, että luku muodostaa mahdollisimman loogisen kokonaisuuden, jossa myös esimerkein avataan termien merkitystä mahdollisimman selkeästi.

Pysyvyys

Pysyvyys tarkoittaa, että samalle objektille ei anneta uutta tunnusta samasta järjestelmästä. Pysyviä tunnisteita hallinnoidaan koko identifioidun objektin elinkaaren ajan ja vielä sen päätyttyäkin: identifioitu objekti voidaan hävittää tai poistaa käytöstä, mutta pysyvää tunnustetta ei tuhota eikä käytetä uudestaan.

Pysyvyys on ehdoton vaatimus ja se vaatii sekä tunnistejärjestelmän että tunnusten dokumentointia. Tunnisteen pysyvyys on käytännössä lupaus siitä, että identifioitu objekti on löydettävissä ja hyödynnettävissä koko sen elinkaaren ajan, ja jotain tietoa siitä myös sen jälkeen.

Julkisen hallinnon tiedonhallinnan näkökulmasta tunnisteiden pysyvyys on keskeinen hyvän hallinnon edellytys, koska viranomaisen tulee huolehtia sekä tarjoamansa tiedon oikeellisuudesta, saavutettavuudesta että jäljitettävyydestä.

Ainutkertaisuus

Ainutkertaisuus tarkoittaa, että kerran annettua tunnusta ei koskaan anneta uudestaan jollekin toiselle objektille. Pysyvän tunnisteen syntaksin tasolla tämä edellyttää sopimista ja hallintaa.

Käyttökonteksti

Tunnuksilla on usein tietty käyttökonteksti, eli niitä käytetään jossain ennalta sovitussa yhteydessä, kuten yksittäisellä hallinnonalalla tai sektorilla. Konteksti voi perustua lainsäädännössä annettuun veloitteeseen – esimerkiksi paikkatiedolle on INSPIRE-direktiivin vaatimuksesta käytettävä ns. Cool URI –tunnisteita, ja tämän lisäksi ja siitä osin poiketen on annettu laajemmat kontekstit mahdollistava kansallisen tason suositus paikkatietojen yksilöivästä tunnisteesta (JHS193²).

² Paikkatiedon yksilöivät tunnisteet. JHS 193 (2015) saatavilla suomeksi: <https://www.avoindata.fi/data/fi/dataset/jhs-193-paikkatiedon-yksiloivat-tunnukset> ja englanniksi <https://www.suomidigi.fi/en/jhs-193-unique-identifiers-geographic-information>

Kontekstin rajaaminen helpottaa hallinnointia ja mahdollistaa tarkemman ohjauksen esimerkiksi tietomallien suhteen. Kontekstin rajaaminen ei itsessään kuitenkaan takaa yhteentoimivuutta vaan on syytä huolehtia esimerkiksi tunnistejärjestelmän perusteella tallennettavien metatietojen yhdenmukaisuudesta.

Tässä dokumentissa keskitytään pääosin tieteen ja tutkimuksen käyttökontekstiin.

Resoluutio ja toiminnallisuus

Resoluutiolla tarkoitetaan tunnistettuun objektiin liittyvien palvelujen tarjoamista. Palvelu voi olla esimerkiksi Internet-resurssin, kuvaavien metatietojen URL-osoitteen tai metatietojen palauttaminen vastauksena rajapintakyselyyn. Palautettava URL-osoite ei välttämättä ole lopullisen kohdesivun osoite, vaan käyttäjä voidaan uudelleenohjauksella viedä oikeaan URL-osoitteeseen.

Resolvoitavissa olevaa tunnistetta kutsutaan **toiminnalliseksi**. Perinteiset tunnisteen kuten kirjan ISBN eivät sellaisenaan ole toiminnallisia, vaikka niiden avulla voi tehdä esimerkiksi Google-hakuja. Monet perinteiset tunnisteen kuten ISBN ja ISSN voidaan esittää myös toiminnallisena tunnisteenä, eli esimerkiksi URN- tai DOI-tunnisteenä. DOI- ja Handle-tunnukset ovat aina toiminnallisia, URN-tunnukset ovat joko toiminnallisia tai ei-toiminnallisia riippuen nimialueesta. Esimerkiksi URN:ISBN-tunnukset ovat toiminnallisia, mutta URN:DEV-tunnukset (Internet of Things -tunnukset) eivät ole.

Resoluutiopalvelu (resolveri) on toiminnallisten tunnisteen resoluoinnista eli uudelleenohjauksesta vastaava sovellus, kuten Handle systemiin liittyvä [Handle.Net](#), jota käytetään sekä Handle- että DOI-järjestelmissä (Digital Object Identifier). URN- tunnuksesta (Uniform Resource Name) ei ole yhtä yhteistä sovelluspakettia, vaan periaatteessa jokaisella URN-tunnuksesta jakavalla taholla on oma resolverisovellus. Resoluutiopalvelu voi sisältää monenlaista älykkyyttä ja lisäpalveluita, joten on syytä valita tarpeita parhaiten palveleva luotettava palvelu.

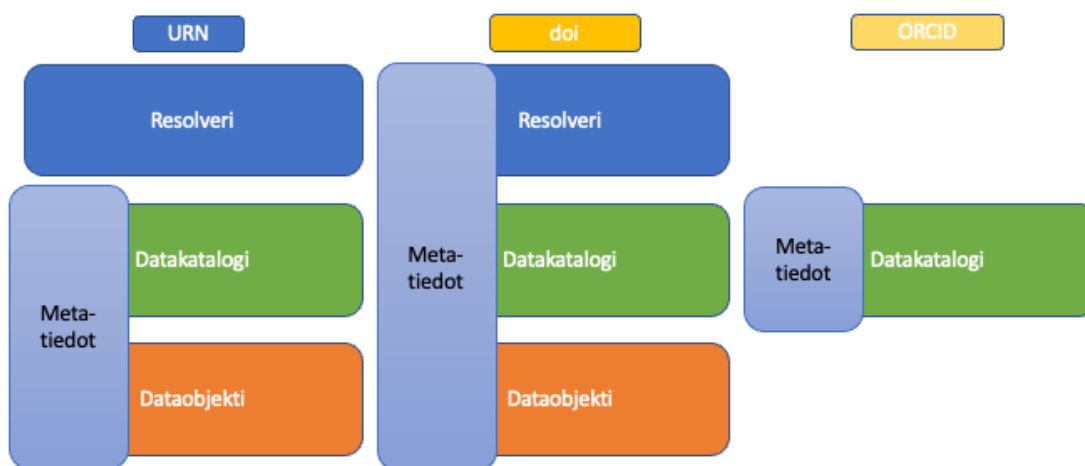
Tunnisteen resoluointia voidaan tehdä usealla eri tavalla. Tavallisin tapa on normaali verkkonimipalvelu, jota käytetään Cool URI-tyyppisissä tunnisteeissa. Näiden resoluoituvuus on siis verkko-osoitteen nimiavaruuden varassa, eikä siten sisällä mitään muuta pysyvyyttä antavaa mekanismia, mikäli esimerkiksi palvelua tarjoavan organisaation verkko-osoite muuttuu.

Käytännössä resolverien palvelutarjonta perustuu resolverille tallennettaviin tietoihin³. Erilliselle resolverille kopioitavien tai luotavien tietojen (kernel metadata, ydinmetatiedot) määrää suositellaan yleensä pidettävän minimissä, koska mastermetatietoja on syytä säilyttää ja ylläpitää samassa paikassa kuin digitaalista objektia tai rekisteriä. Useiden kopioiden luominen on paitsi turhaa myös riskialtista mikäli tietojen kuratoinnissa tai päivityksessä tapahtuu virheitä. Kopiointia vältetään käyttämällä tietoa alkuperäisestä lähteestään yksilöivän tunnisteen avulla. Metatietoja voidaan säilyttää joko erillisessä

³ DOI- ja Handle-resolvereille tallennettavista tiedoista on mahdollista tehdä aineistotyyppikohtainen kansallinen ohje palvelutarjonnan yhdenmukaistamiseksi. Kansalliskirjaston URN-resolverille on toistaiseksi mahdollista tallentaa vain URN-tunnus ja yksi tai useampia URL-osoitteita, joihin se linkittyy. Sovellusta voidaan kehittää muun muassa rikastamalla tallennettavia tietoja.

katalogissa tai osana itse dataobjektia. Metatietojen tulisi olla rakenteisia, jotta ne olisivat myös ohjelmallisesti tulkittavissa. Tällöin metatiedot voidaan myös helposti indeksoida haettaviksi.

Metatiedot mahdollistavat PID:eihin liittyvien palveluiden tarjoamisen, esimerkiksi parantaen löydettävyyttä tai tiedon koneellista hyödyntämistä. Metatiedot ja valitut protokollat ja standardit vaikuttavat myös suuresti niiden tekniseen yhteentoimivuuteen ja niistä saataviin hyötyihin. Koska katteet ja käyttökontekstit vaihtelevat, yhteentoimivuuden tasoja voi ja pitää todennäköisesti olla useita.



Eri järjestelmät tarjoavat erilaisia metatietojen tallennusmahdollisuuksia, jotka on syytä huomioida tiedonhallinnan suunnittelussa.

Semanttisuus

Tunnuksen semanttisuus tarkoittaa, että sillä on tulkittavissa oleva rakenne. Esim. kirjan ISBN-tunnuksesta (International Standard Bibliographic Number) voi päätellä sen julkaisumaan ja kustantajan.

Tunnisteita kehittävät ja luovat organisaatiot eivät ole yksimielisiä siitä, pitäisikö tunnisteiden olla semanttisia vai ei. Esimerkiksi kustantajat väittävät kirjan ensipainoksen ISBN-tunnukseen sisältyvän kustantajatiedon vanhenevan, kun kirjan myöhempien painosten kustantaja muuttuu. Mutta kirjastojen näkökulmasta ISBN-tunnukseen sisältyvä kustantajatieto on pysyvästi validi ja mahdollistaa kirjan alkuperäisen kustantajan tarkistamisen kansallisen ISBN-keskuksen kustantajarekisteristä. Semanttisuutensa ansiosta ISBN-tunnus kertoo, mistä resoluutiopalvelu löytyy. Ei-semanttisten tunnusten kuten ISSN:n resoluointi on hoidettava keskitetysti, koska tunnus ei anna mitään vinkkiä, jonka nojalla oikea resoluutiopalvelu voitaisiin valita. Resolverin löytäminen on tietenkin haasteellista vain silloin, jos resolverin osoite ei sisälly tunnukseen.

Kate

Tunnistejärjestelmän katteella tarkoitetaan sen avulla identifioitavien objektien joukkoa, joka on yleensä rajattu. ISBN-tunnuksilla voi identifioida vain kirjoja, ja ORCID-tunnuksilla (Open Researcher and Contributor Identifier) identifioidaan tutkijoita. Standardoiduissa tunnistejärjestelmissä kate määritellään yleensä jo itse standardissa. Katteen rajaaminen helpottaa tunnistejärjestelmän hallinnointia.

Toimijoiden tunnisteet ovat heikommalla tolalla: tutkijoiden tunniste ORCID ei kata kaikkia tutkijoita, kuten ei myöskään toimijoiden tunniste ISNI (International Standard Name Identifier), organisaatioiden tunnus ROR on vielä lähtökuopissa ja hankkeiden RAiD-tunniste on vasta tulossa käyttöön. Vielä heikompiakin osa-alueita on: kokoelmille on olemassa ISO:n standarditunnus ISCI (International Standard Collection Identifier), mutta sen käyttö on vähäistä ja soveltaminen PID-tunnuksena olematonta. Organisaatioiden omat tunnistejärjestelmät eivät yleensä perustu virallisiin standardeihin lukuun ottamatta ISNI-tunnusta.

Kehitettäessä ja käyttöön otettaessa paikallisia tunnistejärjestelmiä tulisi aina selvittää mahdollisuus käyttää standarditunnisteita. Merkittäviä tunnistejärjestelmiä joiden kate on periaatteessa rajaamaton, ovat esimerkiksi ISO:n OID, Handle, DOI, UUID (Universally Unique Identifier), URN sekä niin sanottu Cool URI. Käytännössä tunnisteiden soveltaminen ei kuitenkaan ole vapaata. DOI-tunnuksien jakelua rajataan Registration Agency -organisaatioiden (rekisteriviranomaisten) kuten Crossrefin ja DataCiten kanssa tehtävillä sopimuksilla. Jos esimerkiksi data-aineistosta on kaksi samansisältöistä versiota eri tiedostomuodoissa, niille on annettava yhteinen DOI-tunnus. Jos pitkäaikaissäilytyksen mahdollistamiseksi halutaan myös versioille omat tunnukset, on käytettävä jotakin muuta tunnistetta.

Opiskelijoiden julkisena tunnisteena voitaisiin periaatteessa käyttää mitä tahansa edellä mainituista tunnistejärjestelmistä, mutta Suomessa Opetushallitus on ottanut käyttöön OID-pohjaisen opiskelijanumeron opiskelijoiden tunnisteeksi, ja opiskelijanumerosta voidaan tehdä URN (URN:OID) tai muu toiminnallinen tunnus.

Esimerkiksi data-aineistoille voidaan antaa DataCiten DOI, mutta niitä kuvaaville metadatatietueille voidaan antaa URN:NBN. Data-aineistolle (tai dataobjektille) DataCiten DOI toimii hyvin teostasolla. Jos siis dataobjektista on CSV- ja Excel-versiot, niillä on silloin yhteinen DOI, joka linkittyy versioiden yhteiseen laskeutumissivuun. Dataobjektin loppukäyttäjille tämä saattaa riittää. Mutta jos dataobjekti pitkäaikaissäilytetään esimerkiksi CSC:n PAS-palvelussa, tarvitaan PID-tunnukset myös dataobjektin versioille ja niitä kuvaaville metatiedoille. Lisäksi metatiedoissa yksittäisillä elementeillä kuten henkilöillä, organisaatioilla ja objektin sekä niiden sisältöä kuvaavilla termeillä voi olla omia PID-tunnuksia. Se miten ja millaisia pysyviä tunnisteita annetaan esimerkiksi aineistojen eri osille (metatiedot, metatietoelementit, tiedostot tai jopa yksittäisille tieto-objekteille tai arvoille) on kontekstiriippuvaista. Tärkeintä on luoda yhtenäiset ja dokumentoidut periaatteet, jotka palvelevat käyttäjäyhteisön tarpeita sekä tukevat tehokasta ja laadukasta tiedonhallintaa. Periaatteet ja käytännöt tule julkaista esimerkiksi osana datapolitiikka, palvelun dokumentaatiota, aineistohallintasuunnitelmaa tai erillisenä PID-politiikkana. On tärkeää muistaa, että pysyvät tunnisteet eivät ole laadun tae, vaan on olemassa paljon tunnisteita, joiden myöntämiseen ei ole minkäänlaista tarkistusprosessia.

Julkaisualalla standarditunnisteiden yhteinen kate on varsin hyvä mutta ei täydellinen, eikä kaikkia standardoitujakaan tunnistejärjestelmiä ole otettu laajasti käyttöön. Esimerkiksi tieteelliselle artikkelille voidaan antaa DOI-tunnus, mutta jos artikkelista on versioita eri tiedostomuodoissa, niille on annettava eri tunnus, eli esimerkiksi URN:NBN.

PID-tunnisteiden tehokas soveltaminen sekä päällekkäisyyksien ja aukkojokohtien välttäminen edellyttää tarkkaa suunnittelua ja eri tunnistejärjestelmien rinnakkaiskäyttöä.

Metatietomalli

Tunnisteisiin voi liittyä metatietomalli. Tunnusta annettaessa identifioitu objekti pitää kuvailla joko standardissa tai muulla tavoin määritellyllä tavalla. Esimerkiksi ISSN-standardi edellyttää, että jokainen tunnuksen saava kausijulkaisu kuvaillaan, ja tiedot lähetetään kansainvälisen ISSN-keskuksen ylläpitämään ISSN-tietokantaan. ISSN –verkosto on laatinut metatietojen tallennuksesta tarkat ohjeet. Metatietomallin kattavuus vaihtelee tunnistejärjestelmästä riippuen. ORCID-tunnuksen saadakseen tutkijan on kerrottava vain etunimensä ja sähköpostiosoitteensa, mutta ISNI-toimijatunnisteen (International Standard Name Identifier) saaminen edellyttää, että on riittävästi metatietoa, jotta voidaan yksiselitteisesti erottaa toimijat toisistaan (esim. yksiselitteinen nimi, syntymävuosi ja julkaisun nimi). Jos tietoja ei ole tarpeeksi (esim. henkilön syntymävuosi puuttuu), silloin odotetaan, että toisesta organisaatiosta tulee vahvistusta. Mitä kattavampi metatietomalli, sen tehokkaampaa on tiedonhallinta. ISNI-tietokannasta löytyvä Albert Einsteinin voidaan tunnistaa luotettavasti tietokannan sisältämien julkaisutietojen avulla, mutta ORCID-tietokannassa olevien Albert Einsteinien identiteetti on epävarmempi. Toisaalta vaatimus kattavien metatietojen tallentamisesta ja monimutkaiset tallennusohjeet tekevät tunnisteiden luomisesta työlästä. Lisäksi esim. toimijoiden tunnisteisiin voi liittyä henkilötietoja kuten syntymäaika, joita ei voida jakaa vapaasti, koska kyseessä on henkilörekisteri.⁴

⁴ Määritelmät perustuvat dokumenttiin: Pysyvät tunnisteet julkisessa hallinnossa. YTI-selvitys 2018. Versio 1.01. s. 2–5. <https://vm.fi/hanke?tunnus=VM032:00/2017>

Tarve kansalliselle pysyvien tunnisteiden politiikalle

Tässä luvussa kuvataan pysyvien tunnisteiden merkitystä hallinnolle ja tiedonhallinnalle yleensä.

Pysyvät tunnisteet osana julkista tiedonhallintaa

Viranomaisia veloitetaan tarjoamaan tietoja koneluettavassa muodossa rajapintojen kautta.⁵ Myös Kysy vain kerran -periaate ja eurooppalaiset yhteentoimivuuden vaatimukset edellyttävät tiedon laatu-, provenanssi- ja eheystietojen hallintaa.⁶ Keskeinen merkitys näiden toteutumiselle on tiedon yksiselitteiden tunnistettavuus ja luotettava viittaamismahdollisuus (linkittäminen). Käytännössä tämä vaatii pysyvien tunnisteiden hallintaa. Tällä hetkellä pysyvien tunnisteiden kenttä ja hallinta on kuitenkin sirpaloitunutta ja yhteisiä käytäntöjä kaivataan. Tunnisteet tulisi tunnistaa osaksi tiedonhallintakarttaa, sillä yhteisiä käytäntöjä tulisi ylläpitää ja päivittää yhteistyössä eri alojen toimijoiden kanssa. Käytäntöjen pohjana tulee olla lakisääteiset syyt, jotta pysyvien tunnisteiden käytössä ei ole tulkinnanvaraisia tekijöitä. Lähtökohtana tunnistejärjestelmien hallinnalle on oltava, että kaikilla tunnisteilla on vastuutaho, jonka toiminta on tarpeeksi organisoitua ja rahoitettua. On tärkeää määritellä mitkä digitaaliset objektit tarvitsevat pysyviä tunnisteita ja miksi nämä objektit ovat tärkeitä ja mikä taho tunnisteista vastaa. Pysyviin tunnisteisiin liittyy useita eri rooleja, joista kerrotaan tarkemmin *Toimijat ja roolit tutkimuksen käyttökontekstissa* - alaluvussa. Tutkimustiedon osalta on osoitettu, että pysyvien tunnisteiden käyttö tiedonhallinnassa myös säästää resursseja.⁷

Valtiovarainministeriö on nostanut esiini eri tasoisten linjausten tarpeen eurooppalaisen yhteentoimivuustyön pohjalta.⁸ Jotta yhteentoimivuutta voidaan toteuttaa, tarvitaan operatiivisen, taktisen sekä strategisen tason yhteentoimivuutta. Valtionhallinnon tiedonhallintakartan pohjalta pitäisi siis tehdä tarkempi arkkitehtuuri myös tunnisteille niiden keskeisen roolin vuoksi.

⁵ <https://www.finlex.fi/fi/laki/ajantasa/2019/20190906> (esim. §22, §24 §24a ja §24b)

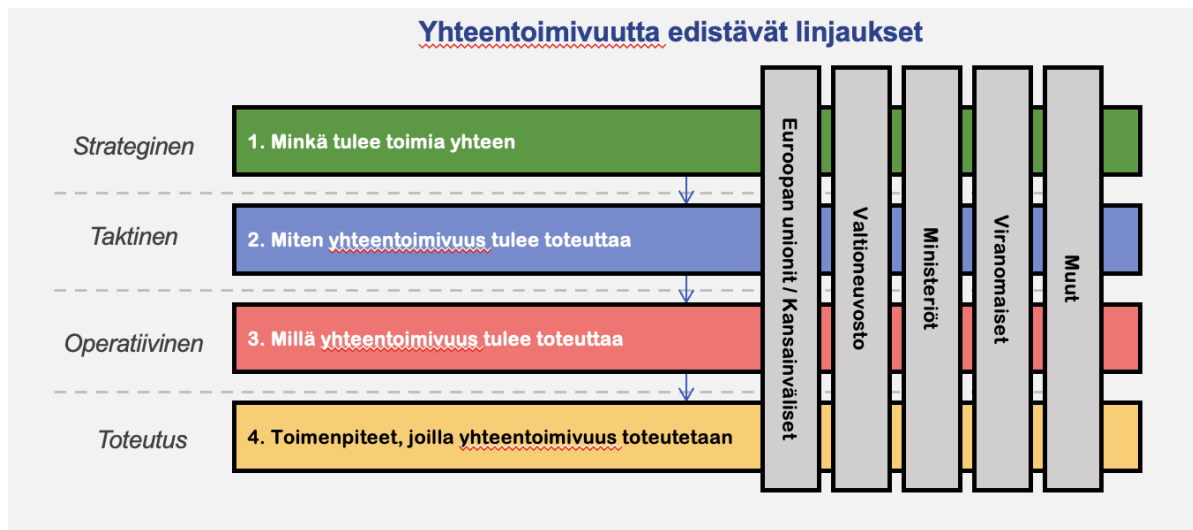
⁶ EU:n yhteentoimivuudesta lisää <https://joinup.ec.europa.eu/>. "Once only" on Euroopan unionissa määritelty tavoitteeksi (<https://ec.europa.eu/digital-building-blocks/wikis/display/DIGITAL/Once+Only+Technical+System>). Sama periaate liittyy myös OmaData-periaatteeseen ja se on mainittu kaikkien eduskuntapuolueiden tietopolitiikan toimijoiden yhteistyöryhmän julkaisemassa toimenpide-ehdotuspaperissa 12/22. <https://tietopolitiikka.fi/>. OmaDatan määritelmä SITRAn mukaan: <https://www.sitra.fi/tulevaisuussanasto/omadata/>

⁷ Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022). Revised cost-benefit analysis for the UK PID Support Network (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7356219>

⁸ Julkisen hallinnon yhteentoimivuuden linjaukset <https://vm.fi/julkisen-hallinnon-yhteentoimivuuden-linjaukset>
European Interoperability Framework
<https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail>

European Interoperability reference architecture

<https://joinup.ec.europa.eu/collection/european-interoperability-reference-architecture-eira/solution/egovera>



Kansallista PID-politiikkaa tarvitaan, sillä pysyvien tunnisteiden määrittämiseen tarvitaan selkeitä, esimerkiksi tiedonhallintalakiin ja valtionhallinnon tiedonhallintamalliin perustuvia ohjeita ja käytäntöjä. JulkICT:llä ja tiedonhallintalautakunnalla on keskeinen rooli. Jos tarkastellaan julkista hallintoa laajasti tiedon tuottajana (valtio ja myös kuntia) keskeinen kysymys on, minkä muotoisen (teknisen) tunnuksen he antavat asialle, asiakirjalle yms. objektille, joihin tulee voida prosessin myöhemmissä vaiheissa ja muissa prosesseissa teknisesti viitata. Esimerkiksi kaupungeilla voi olla satoja tietojärjestelmiä joissa on erilaiset tavat yksilöidä tietoaineistoja. Tässä kohdin myös resolvointi on haastava kysymys, koska tiedon käytettävyys ei ole usein kyseisessä ympäristössä vielä ole täysin avointa, mutta kenties joskus tulevaisuudessa näin halutaan toimia. Olennaista olisi se, että annetut tunnukset kulkeutuisivat aikanaan tiedon mukana arkistoihin asti ja sitä kautta niitä sitten jatkossakin voisi hyödyntää myös tutkimuksessa.

Kansallisen PID-politiikan tavoitteena olisi luoda käytäntöjä, joiden avulla pysyvien tunnusten luonti tapahtuu osana tiedon tuotantoprosessia eikä jälkikäteen ja pysyviä tunnisteita hyödynnetään kattavasti yli organisaatorajojen. Opetus- ja kulttuuriministeriön luoma sääteley opinnäytetöiden hallinnasta on hyvä esimerkki onnistuneesta käytännöstä. Taustana pitää olla selvitys jossa on kartoitettu käyttötarpeita ja tapauksia.

Perusrekisteriviranomaisten ja esimerkiksi tutkimuslaitosten pitäisi tiedostaa, milloin niiden käyttämiensä tunnisteiden pitäisi olla pysyviä tunnisteta ja niiden itsensä tulee toimia referenssietopalveluina. Tämä tulee tärkeäksi viimeistään kun aineistoja avataan rajapintojen kautta verkkoon käytettäväksi, mutta yllä mainituista syistä (esim arkistointi ja pitkäaikaissäilytys) tunnisteiden hallinnasta organisaatiotasolla on hyötyä muutenkin esimerkiksi hallintorajojen yhteentoimivuuden edistämiseksi. Tunnisteiden hallintaa pitäisi koordinoita osana julkisen hallinnon tiedonhallintakarttaa mikä tehostaisi tiedonhallintaa kokonaisuutena. Avoimen datan direktiivissä tullaan määrittelemään arvokkaille tietoaineistoille teemaluokkia, ja näille pitäisi nimetä omistajat ja sen myötä ydintietojen pysyvien tunnisteiden hallinnasta vastaavat tahot. Tarvittaessa viranomainen voi toimia kaikissa yllä luetelluissa rooleissa. On kuitenkin syytä pohtia kestävä ja tehokasta toteutusmallia ja varmistaa sekä yhteentoimivuus että järjestelmien ja tunnisteiden koko elinkaari.

Toimijoita pitää velvoittaa käyttämään olemassa olevia tunnisteita, soveltamaan pysyviä tunnisteita myös omissa palveluissaan. Tämä edellyttää tietoa tarjolla olevista PID-tunnistejärjestelmistä ja niiden käyttömahdollisuuksista.

Myös elinkeinoelämää tulisi aktivoida toisaalta pysyvien tunnisteiden hyödyntämiseen tiedonhallintansa tehostamiseksi, toisaalta omien arvokkaiden tietoresurssien avaamiseen pysyvin tunnistein identifioituna. Eurooppalaiset data-aineistojen hyödyntämiseen luotavaat data-avaruudet⁹ lisäävät vielä jäljitettävyyden, yhteentoimivuuden ja luotettavan viittaamisen tarvetta myös kaupallisella puolella.

Pysyvien tunnisteiden tarjoamat palvelut ja niiden edellytykset esim. koskien metatietoja on tärkeää kuvata ja tuoda koko kansallisen toimijakentän saataville. Erilaiset jo käytössä olevat tunnisteet (DOI, Handle ja eri URN-nimialueet; toimijoiden tunnisteet) sekä niiden hallintamallit ja prosessit on oleellista kuvata kattavasti kansallisella tasolla. Pysyvien tunnisteiden soveltamista ohjaavat kansainväliset käytännöt kuten Crossrefin ja DataCiten DOI-jakelua koskevat vaatimukset, sekä tunnistejärjestelmien taustalla olevat standardit. Nämä rajoitukset on otettava huomioon, kun mietitään eri järjestelmien soveltuvuutta meidän oloihimme. Suomessa ei kannata ottaa käyttöön tunnisteita, jotka soveltuvat aiottuun käyttötarkoitukseen huonosti. On myös vältettävä PURL-tunnisteen kaltaisia hiipuvia järjestelmiä, joiden kehittäminen ei ole enää aktiivista ja jotka eivät perustu mihinkään standardiin.

Paikkatiedon tunnisteet ja tutkimuksen toimijoiden tunnisteet (ISNI, ORCID ja ROR) ovat esimerkiksi käytännössä URL-osoitteita. Toimijoiden tunnisteiden pysyvyys ja toiminnallisuus perustuvat globaaleihin tietokantoihin, joihin toimijoita koskevat tiedot tallennetaan keskitetysti. Kansallisen tason mahdollisuudet vaikuttaa siihen, miten nämä tunnukset resolvoituvat, ovat rajalliset. ISNI-tunnuksien osalta Kansalliskirjasto voi Registration Agency -organisaationa luoda uusia tunnuksia, korjata virheellisiä tunnuksia ja muokata tunnuksiin liittyviä metatietoja. ORCID-tunnistejärjestelmässä metatietojen ylläpitovastuu on tunnuksen luoneella tutkijalla. Hallinnossa on monia tunnus-järjestelmiä, kuten Y-tunnukset, joiden hallinta pysyvien tunnisteiden tavoin olisi luontevaa ja kaikille sidosryhmille hyödyksi (esim. toiminnallisuus ja pysyvyys).

Kansallinen yhteistyö on tärkeää, jotta rinnakkaisia PID-järjestelmiä voidaan käyttää järjestelmällisesti yhdessä. Usean rinnakkaisen pysyvän tunnisteen käyttö mahdollistaa valmistautumisen myös tuleviin Euroopan Unionin vaatimuksiin. Suomessa tulisi sopia yhtenäisestä tavasta ilmaista eri tunnistetyyppejä ja niiden relaatiota. Suomen ei kuitenkaan pidä jäädä odottamaan vaan valmistautua laajentuvaan pysyvien tunnisteiden hyödyntämiseen.

Riippuvuudet kansainvälisiin toimijoihin ja järjestelmiin ovat myös potentiaalinen riski, jota on syytä minimoida esimerkiksi osallistumalla kansainväliseen yhteistyöhön pysyvien tunnisteiden osalta ja esimerkiksi peilata kriittisiä pysyviä tunnisteita Suomessa.

⁹ Euroopan parlamentin ja neuvoston asetus (EU) 2021/694, annettu 29 päivänä huhtikuuta 2021, Digitaalinen Eurooppa -ohjelman perustamisesta ja päätöksen (EU) 2015/2240 kumoamisesta. <https://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=OJ:L:2021:166:FULL&from=EN>

Tunnisteiden huolellista ja johdonmukaista käyttöä on edistettävä paitsi Suomessa myös kansainvälisesti. Resolvointi-infrastruktuuriin liittyvistä linjauksista pitää myös ohjeistaa. Kenen hallussa resolvointi-infrastruktuurin tulisi olla? On olemassa perusteita miksi näiden pitäisi olla kansallisessa kontrollissa. Myös tunnisteiden omistajuus on huomioitava PID-palveluita valittaessa.

Tutkimuksen pysyvät tunnisteet

Tässä luvussa kuvataan tieteen ja tutkimuksen käyttökontekstin nykytilaa ja käytössä olevia tunnisteita sekä toimijoita

Tutkimuksen erityispiirteet

Tieteen ja tutkimuksen alalla FAIR-periaatteet muodostavat kansainvälisen, keskeisen ohjenuoran palveluiden ja datainfrastruktuurin kehittämisessä. FAIR-periaatteet tarkoittavat, että erilaisten tutkimusaineistojen, tutkimusmenetelmien ja tutkimusjulkaisujen tulee olla löydettävissä, saatavilla, siirrettävissä tai yhdistettävissä, uudelleen käytettävissä ja toistettavissa. Tämä edellyttää pysyvien tunnisteiden hyödyntämistä ja hallintaa. Lisäksi aineiston pitää olla tarvittaessa myös pitkäaikaissäilytettävissä. Sekä Suomen Akatemia että EU:n tutkimusrahoittajat edellyttävät FAIR-periaatteiden huomioimista.¹⁰ FAIR edellyttää pysyviä tunnisteita tutkimuksen toistettavuuden varmistamiseksi, ja niiden soveltamisen pitäisi tapahtua niin automatisoidusti kuin mahdollista.

Vaikka tutkijalta edellytetään tuotosten julkaisemista FAIR-periaatteiden mukaisesti, hän ei pysty käytännössä tekemään sitä yksin, vaan tarvitsee kehysorganisaationsa tukea ja palveluita. Korkeakoulut, tutkimuslaitokset ja muut PID-tunnuksia käyttävät julkishallinnon organisaatiot puolestaan saavat tukea kansallisilta PID-palveluiden tarjoajilta. Näiden organisaatioiden pitää ohjata tutkijoitaan luotettavien PID-palveluiden luokse. Tarjolla on eritasoisia kansallisia ja kansainvälisiä palveluita kuten Fairdata-, Eudat ja Zenodo-palvelut tutkimusaineistoille, tieteellisten kustantajien DOI-palveluita esimerkiksi tieteellisille artikkeleille sekä Software Heritage tutkimusohjelmistoille. Nämä palvelut mahdollistavat pysyvän tunnisteiden hankkimisen tutkimuksen tuotoksille.

Pysyvä tunniste annetaan tutkimuksessa useimmiten julkaistuille tai julkaistaville tuotoksille. Tutkimusprosessin aikana saatetaan tarvita jo pysyviä tunnisteita organisaation ja useamman organisaation sisällä. Tähän käytetään joko sisäisiä tunnisteita tai esimerkiksi data-aineistoja voidaan julkaista karttuvina tai dynaamisina aineistoina. PIDien käyttöön tulisi laatia yhteiset linjaukset esimerkiksi sellaiseksi, että objektille annetaan vain yksi PID, mutta saman sisällön erilaisille manifestaatioille (eri tiedostomuodoissa julkaistut sisällöltään identtiset versiot) voidaan antaa omat PIDit.

Kansainväliset standardit ja parhaat käytännöt kehittyvä nopeasti. Tähän kehitystyöhön on syytä osallistua resurssien sallimissa rajoissa. Suomi on osallistunut esimerkiksi DOI- ja ISNI-standardien kehittämiseen. Myös Research Data Alliancessa (RDA) tehdään paljon data-aineistoihin ja tunnisteisiin liittyvää työtä, mutta mandaatti on epäselvä. DataCite on vahvistanut asemaansa viime aikoina laajentamalla tietomalliaan ja sen katetta.

Tutkimustiedoissa käytettyjen tunnisteiden avulla esimerkiksi organisaatiot ja rahoittajat pystyvät seuraamaan tutkimuksen kehitystä. Tutkimuksesta kerätään eri lähteistä tietoja

¹⁰ EOSC:in strategia https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf, European Research Councilin ohjeet <https://erc.europa.eu/manage-your-project/open-science> sekä Suomen Akatemian ohjeita <https://www.aka.fi/en/research-funding/apply-for-funding/how-to-apply-for-funding/az-index-of-application-guidelines2/data-management-plan/data-management-plan/>

tietovarantoihin, kuten Tiedejatutkimus.fi¹¹ ja OpenAIRE, joiden avulla tutkimuksen löydettävyyttä ja vaikuttavuutta voidaan seurata. Nykyisin näistä tiedoista tuotetaan linkitetyn datan kokonaisuuksia joissa tieto muuttuu entistä paremmin rakenteistetuksi ja paremmin löydettäväksi. Tietograafit asettavat tiedot asiayhteyteen linkittämisen ja semanttisen metatiedon avulla sekä mahdollistavat näin tietojen integroinnin, yhdistämisen, analysoimisen, visualisoinnin ja jakamisen automatisoidusti.

Tutkimuksessa myös tutkimusprosessin dokumentointi, tiedostoformaattit ja toistettavuus sekä oikeuksien ja pääsyn hallinta voivat olla tärkeitä ja samalla erityisen haasteellisia.

Toimijat ja roolit tutkimuksen käyttökontekstissa

Tutkimuksen puolella on EOSC:in PID-politiikassa¹² on määritelty seuraavat yleiset toimijaroolit:

- PID-auktoriteetti (vastaa standardin ylläpidosta)
- PID-palvelun tarjoaja (vastaa PID-resolverista ja nimiavaruuksien hallinnasta)
- PID-manageri (vastaa tunnisteiden ja niiden ydinmetatietojen hallinnasta)
- PID-omistaja (luo ja allokoi tunnisteiden, vastaa (meta)tietojen sisällön ylläpidosta)
- PID-käyttäjä (loppukäyttäjä)

Pysyvien tunnisteiden alueella tapahtuu tällä hetkellä nopeaa kehitystä myös Euroopan tasolla. European Open Science Cloud -hankkeisiin liittyen tehdään mm. resolvointipalvelua, kehitetään PID-politiikkoja ja muodostetaan PID-palveluntarjoajien yhteistyöfoorumia.¹³ Lisäksi yhteentoimivuutta esim. URN:ien ja DOI:n välillä kehitetään ja uutta RAIID-tunnistetta ollaan standardoimassa.

Suomessa yhteistyön koordinointi on pääosin PID-verkoston varassa ja epävirallista toimintaa lähinnä Kansalliskirjaston ja CSC:n toimesta. Mikäli yhteisen kansallisen infrastruktuurin luomisessa rahoitus tulisi useammilta sektoreilta voitaisiin luoda varsinainen PID-keskus sekä verkosto missä rahoitettaisiin joukko PID-resolvereita, joilla hajautettu ylläpito. Tässä PID-tiekartassa on tunnistettu seuraavat kansalliset palveluiden tuottajat tutkimuksen alueella:

¹¹ Tutkimustietovarannon tunnistamat pysyvät tunnisteet (luonnos) löytyvät täältä:

<https://koodistot.suomi.fi/codescheme:registryCode=research:schemeCode=PID>. (20.2.2023)

¹² European Commission, Directorate-General for Research and Innovation, Schwarzmann, U., Fenner, M., Hellström, M., et al., PID architecture for the EOSC : report from the EOSC Executive Board Working Group (WG) Architecture PID Task Force (TF), Publications Office, 2021, <https://data.europa.eu/doi/10.2777/525581>

¹³ Ks. tarkemmin <https://fair-impact.eu/> ja <https://faircore4eosc.eu/> sekä <https://www.eosc.eu/advisory-groups/pid-policy-implementation> (20.3.2023)

Tyyppi	Määritelmä	Palveluntuottaja	Esimerkit katteesta tai käytöstapauksesta
PID-palveluiden tarjoaja	Organisaatio, joka tarjoaa PID-palveluita tietyn hallinnoidun PID-skeeman mukaisesti. Palvelun tarjoaja vastaa luotettavan palvelun tuottamisesta, jotta sen myöntämät tunnisteet pysyvät eheinä ja resoluoituvat. PID-palvelun tarjoaja vastaa siitä, että palvelu skaalautuu, on yhteentoimivia ja uusia tunnisteita voi luoda.	Kansalliskirjasto	julkaisut, URN toimijat, ISNI
		DataCite (CSC) CrossRef (Tieteellisten Seurain Valtuuskunta)	DataCite DOI CrossRef DOI tutkimustuotokset, kuten artikkelit tai datasetit
		ORCID	tutkijat, ORCID
Aineistopalveluiden tarjoajat	Aineistopalveluiden tuottajat vastaavat pysyvien tunnisteiden hallinnasta ja niiden objektien eheydestä, metatietojen oikeellisuudesta sekä resolverissa että muualla, sekä oikeasta käyttökattesta ja PID-skeeman mukaisesta käytöstavasta. Tyypillisesti aineistopalvelun tarjoaja hankkii tunnistepalveluita sopivalta PID-palveluiden tarjoajalta. Tunnisteiden kustannuksia vältetään siirtämistä loppukäyttäjille.	Kansalliskirjasto Tieteellisten Seurain Valtuuskunta	Julkaisut
		Korkeakoulut Tutkimuslaitokset	Organisaation omissa palveluissa olevat tutkimustuotokset
		Kielipankki, Fairdata-palvelut, Tietoarkisto, Laji.fi	Tieteenalakohtaisesti ja yhteisissä palveluissa olevat tutkimusaineistot
		Arkistot ja museot	Kulttuuriperintöaineistot
		Tutkimustietovaranto	tutkimusinfrastruktuurit, rahoituspäätökset

PID-auktoriteetit

PID-auktoriteetilla tarkoitetaan tahoa joka omistaa PID-skeeman ja ohjaa sen hallintaa ja kehitystä. Tällaisia tahoja ovat esimerkiksi, DONA (DOI) ja IANA (URN), jotka m.m. hallinnoivat nimiavaruuksia, jotta ne ovat uniikkeja ja voivat määrittellä niiden katteen.

Joissain tapauksissa PID-palveluiden tarjoajat luottavat DNS-järjestelmään ja luovat sitten oman skeemansa, kuten esim ORCID on tehnyt.

Konsortiot

Suomessa on muutamia PID-konsortioita, joiden avulla pyritään helpottamaan ja tukemaan organisaatioiden PID-palveluiden hyödyntämistä käytännössä.

DataCiten tutkimusaineistoille sopivaa PID-palvelua (DOI) organisaatiot voivat hyödyntää liittymällä suoraan DataCiten jäseniksi tai CSC:n hallinnoiman DataCite Finland -konsortion kautta. Jäsenyydellä pääsee hyödyntämään DataCiten Fabrica-järjestelmää sekä DataCite REST APIa. Vuonna 2022 tällä konsortiolla on 6 jäsentä. CSC on myös ePIC-konsortion jäsen, ja handle tunnisteita voi saada esimerkiksi tämän kansainvälisen konsortion kautta.

Samalla tavalla kansallinen ORCID-konsortiokaan ei suoraan hallinnoi tutkijoille tarkoitettuja ORCID-tunnisteita tai niiden käyttöä. Se huolehtii suomalaisten tutkimusorganisaatioiden ORCID-jäsenyyksistä, jonka avulla organisaatiot saavat oikeuden käyttää ORCIDin kattavinta kaksisuuntaista rajapintaa. Tällä hetkellä tämä CSC:n hallinnoima Suomen konsortio on pääasiassa hallinnollinen toimija ja sillä on 32 jäsentä.

Muut

PID-verkosto on epämuodollisena verkostona toiminut jo useita vuosia. Verkosto on koonnut tietoa yhteisille wikisivuille, ja osaamista on jaettu säännöllisissä tapaamisissa yleensä pari kertaa vuodessa. Verkoston asemaa asiantuntijaverkostona olisi hyvä vahvistaa, mutta yhteentoimivuuden ja koordinaation edistäminen on tällä hetkellä edelleen epävirallista ja vapaaehtoista toimintaa oman toimen ohessa. Jotta päästäisiin tarpeeksi vahvaan koordinaatioon, varsinkin tutkimuksen tiedonhallinnan kehityksen rajoilla kansallisen ohjauksen tarve on selkeä. Ohjauksen tulisi olla linjaustasoista ja käyttötapauksia ja -tarpeita huomioivaa, jotta implementaatiot onnistuvat. Tämä vaatii panostusta esimerkiksi edellä mainittujen kartoittamiseen.

Tutkimuksen pysyvien tunnisteiden järjestelmät ja katteet

Tutkijat

ORCID-käyttöä voidaan suositella ja ISNI-tunniste paikkaa ORCIDin aukkokohdat, eli edesmenneet ja vastahakoiset tutkijat jotka eivät voi/halua hakea itselleen tunnusta.

Väitöskirjat

Opinnäytetöiden säilyttäminen on säädetty lakiperustaisesti, mutta Opetus- ja kulttuuriministeriön ohje koskee vain pysyvästi säilytettävän aineiston tiedostomuotoa (PDF/A), ei PID-tunnusta tai muita metatietoja. Väitöskirjan verkossa julkaistava (pdf-)versio saa ISBN-tunnuksen ja siihen pohjautuvan URN:in, ks. esim. <https://urn.fi/URN:ISBN:978-951-29-9043-6>. Suomessa ei ole tapana antaa väitöskirjoille DOI-tunnistetta. Jos väitöskirjalle annetaan Crossrefin DOI, se on teoksen tunniste joka kattaa väitöskirjan kaikki nykyiset ja myös tulevat manifestaatiot (eri tiedostomuodoissa julkaistut sisällöltään identtiset versiot). Crossrefin DOI on silloin teostason tunnus, URN:ISBN julkaisun PDF- ja EPUB-versioiden tunnus ja esimerkiksi voidaan lisäksi antaa URN:NBN julkaisun metatietojen tunnuksiksi. Yhden julkaisun eri osien tai

manifestaatioiden identifiointiin voidaan siis tarvita useita PID-järjestelmiä.¹⁴ Väitöskirjalle annettu URN:ISBN on ISBN-järjestelmän sääntöjen mukaisesti manifestaatiokohtainen, mutta jos väitöskirjan yksittäinen manifestaatio koostuu useista tiedostoista, niillä on kaikilla sama URN:ISBN mutta jokaisella tiedostolla pitää lisäksi olla oma URN:NBN eli kansallisbibliografioiden ID-tunnus tiedostojepitkäaikaissäilytyksen mahdollistamiseksi.

Tieteelliset artikkelit

Muiden muassa TSV puolestaan tarjoaa Crossrefin DOI-tunnuksia sellaisten tieteellisten lehtien artikkeleille, jotka julkaistaan heidän palveluissaan (artikkeleiden kohdalla Journal.fi ja monografioiden kohdalla Edition.fi). Muutamilla muilla toimijoilla (ainakin Helsingin yliopisto, Jyväskylän yliopisto, SKS) on itsenäisesti hankittuja DOI-tunnuksia.

Kansalliskirjaston URN ISBN tunnuksia annetaan kirjoille ja URN:NBN-tunnuksia annetaan kaikenlaisille objekteille. Ulkopuolisista toimijoista näitä voivat käyttää ne, joilla on sopimus kansalliskirjaston kanssa. Järjestelmässä on edelleen joitakin aukkoja:

- Vanhan lehtiartikkelin tarvitsema DOI-tunnus, tällöin hallinnoijan on tehtävä sopimus suoraan CrossRefin kanssa.
- Uuden kotimaisen lehtiartikkelin tarvitsema tunnus, mikäli sitä ei julkaista Journal.fi -palvelussa.
- Mikäli artikkelista on kaksi versiota eri tiedostomuodoissa, DOI on niiden yhteinen eikä artikkeleita ei voi siirtää PAS-palveluun.

Tutkimusaineistot

Tutkimusaineisto on dataa, jota hyödynnetään tutkimuksessa tai syntyy sen tuloksena. Mikä tahansa objekti - kartta, video, tietokanta - voi olla tutkimusaineistoa. Tutkimuksissa tuotettuja data-aineistoja julkaistaan monissa eri palveluissa, joista Fairdata-palvelut ja useat muut käyttävät aineistojen identifiointiin DataCiten DOI-tunnuksia. Niiden rinnalla käytetään myös muun muassa Handle- ja URN:NBN-tunnuksia. Dynaamisille tutkimusaineistoille ei ole olemassa erityisesti niille kehitettyä tunnistejärjestelmää.

Hajautettuun ja/tai karttuvaan dataan on mahdollista viitata antamalla tunniste sekä tietolähteelle, että sinne kohdistuvalle haulle. Esimerkiksi Lajitietokeskus tarjoaa DOI:ta haku-tuloksille, mikä mahdollistaa viittaamisen saatuun hakutulokseen, vaikka sama haku myöhemmällä ajanhetkellä tuottaisi eri tuloksen. Hakutulokselle annettava DOI mahdollistaa myös useampaan datalähteeseen viittaamisen yhdellä tunnisteella, ja samalla viittaamisen alkuperäisiin lähteisiin. Fairdata-palvelussa on mahdollista luoda karttuva datasetti, johon on mahdollista lisätä dataa ilman, että muuttaa tai poistaa olemassa olevia tunnisteita, ja rikkomatta aineistoa. Pitkäaikaissäilytys Fairdata-PAS hyödyntää DOI-tunnisteita.

DOI-tunnusten osalta olemme riippuvaisia tunnisteiden käytöstä sovitusta kansainvälisistä sopimuksista. Muiden (Handle ja URN:NBN) osalta käyttö on vapaampaa. Suomessa CSC tarjoaa DataCiten DOI-tunnuksia pienille toimijoille. Nämä tunnisteet on tarkoitettu pääasiassa tutkimusaineistoille, mutta niitä voi soveltaa myös julkaisuarkistoihin tallennettuihin aineistoihin.

¹⁴ 'Elektonisten julkaisujen identifiointi'-opas kertoo miten PID:ien avulla voidaan tunnistaa sähköisiä julkaisuja. <https://urn.fi/URN:ISBN:978-951-51-7661-5>

DOI-tunnuksen voi saada kaikelle aineistolle, jonka julkaisee Zenodo-palvelussa. Näiden metatietojen luotettavuus riippuu siitä, pitääkö tallentaja itse niitä yllä. Siksi Zenodoa ei voi pitää vartenotettavana vaihtoehtona esimerkiksi TSV:n ja CSC:n DOI-tunnuksille.

On todennäköisesti olemassa tarvetta organisaatioiden omille data-arkistoille, joissa kuitenkin datan eheyden hallinta ja aineistojen kuratointi ovat vaativaa toimintaa. Tunnisteiden myöntäminen vaatii huomattavia panostuksia tiedonhallinnan organisointiin. DataCiteen liittyminen on mahdollista Suomen konsortion kautta tai suoraan, mutta edellyttää riittäviä omia palveluita. Aukkoja tällä hetkellä ovat muun muassa seuraavat:

- Samasta data-aineistosta on kaksi versiota eri tiedostomuodoilla, näille ei saa omia tunnisteita. Jos eri tiedostomuodot laitetaan pitkäaikaissäilytykseen, niillä olisi oltava omat tunnisteet.
- Tietokantojen ja muiden kompleksisten tietojen tunniste-käytännöt ja -palvelut vaativat kehitystä

Lähdekoodi, työvuot ja muut menetelmät

Tutkimuksessa syntyy ja hyödynnetään myös sovelluksia ja erilaisia malleja ja muuta koodia. Näiden dokumentointi ja viitattavuus on usein edellytys tutkimuksen toistettavuuden varmistamiseksi. Näille tuotoksille on tekeillä avoimen tieteen osalinjaus, joka julkaistaan kesäkuussa 2023¹⁵.

Koodia saatetaan nykyisin julkaista Zenodossa palvelussa olevan GitHub-integraation avulla (DataCite DOI) tai Software Heritagen arkistossa (<https://archive.softwareheritage.org/>), jossa käytössä on oma sisäinen tunniste, jota tosin markkinoidaan vahvasti esim EOSCissa. Uudelleenkäytettävyyden varmistaminen vaatii tutkijalta erityistä huolellisuutta dokumentaation ja riippuvuuksien suhteen.

Työvuot ovat vielä suhteellisen kartoittamaton kate, mutta tiedon laadun ja tutkimuksen toistettavuuden puolesta näiden kehittäminen olisi tärkeää. Tutkimusmenetelmiin viittaaminen edellyttää, että niille on oma repositorio, jossa niitä hallitaan, kuvaillaan ja niiden eheydestä huolehditaan.

Muita tutkimusmenetelmiä voi julkaista erilaisissa palveluissa, mutta palvelutarjonta ja käytännöt ovat usein vielä puutteellisia.

Sanastot, muuttujat ja koodistot

Varsinkin tutkimustiedon osalta rinnakkaisten, kopioitujen koodistojen kohdalla pitäisi viitata koneluettavasti lähdekoodiston oikeaan versioon. Tavoitteena olisi, että koodistoista olisi yksi ylläpidetty masterversio, johon voi luoda paikallisia lisäobjekteja tai omia näkymiä. Tutkimusaineistojen ontologioille on tehty oma suositus, jota on suositeltava noudattaa myös asiasanastojen osalta.¹⁶ Muuttujien standardinomainen kuvaaminen ja kuvausten ylläpito olisi ensisijaisen tärkeää, mutta tätä tulisi ainakin pääosin tehdä kansainvälisellä tasolla.

¹⁵ Linjaukset julkaistaan Avointiede.fi-sivustolla Avoimen tieteen kesäpäivien yhteydessä:

<https://avointiede.fi/fi/linjaukset-ja-aineistot/kotimaiset-linjaukset>

¹⁶ D2.8 : 3th set of Recommendations for FAIR Semantics. FAIRsFAIR. <https://doi.org/10.5281/zenodo.6276576>

Organisaatiot

Myös organisaatioilla on pysyviä tunnisteita. Tällaisia ovat käyttökontekstista riippuen esimerkiksi RoR, PIC, ISNI, Y-tunnus ja CrossRef_Funders. Kontekstilla tarkoitetaan, että rahoituksen saajana organisaatiolla saattaa olla eri tunniste (PIC), kuin sopimuksen osapuolena (Y-tunnus), kuin rahoituksen myöntäjänä (CrossRef_Funders). Linkitys löytyy esim Wikidatasta (QID). Organisaatioiden erityispiirre on niiden muutokset, jotka edellyttävät relaatioiden hallintaa mieluiten koneluettavan ontologian avulla. Suomessa kansallisille toimijoille tällainen ontologia on KANTO - Kansalliset toimijatiedot (<https://finto.fi/finaf/en/>). OPH ylläpitää koulutuksen ja tutkimuksen suomalaisten julkisten toimijoiden organisaatiotietoja alayksikköineen Organisaatiopalvelussa¹⁷.

Rahoituspäätökset

Rahoituspäätökset Tutkimustietovarannossa

Tutkimustietovarannossa on vuoden 2022 lopussa tietoja sekä kansallisista että EU:n rahoituspäätöksistä.

Euroopan unionin rahoituspäätökset

EU:n päätöksistä tietoja on H2020-puiteohjelman, aluekehitysrahaston (fi:EAKR/en:RDF) ja sosiaalirahaston (en:ESF) rahoituspäätöksistä. Aiempien puiteohjelmien päätöksiä ei ole toistaiseksi tuotu ja 2021 alkaneen Horizon Europe puiteohjelman päätöksiä lataaminen on työn alla. EUn rahoituspäätösten primääri tunnuksina on pitkään aikaan ollut hankkeiden rahoituspäätösnumero, joka on EUn paikallinen tunniste. (aiemmin RCN, nyk. Grant ID) on eri asia. Tämä paikallinen tunniste on siinä määrin vakioitu, että kaikkien projektien www-sivulle Cordiksessa voi laskeutua vakioitu URLilla, joka on kiinteä alkuosa+grant id (esim: <https://cordis.europa.eu/project/id/101057264>). Kuitenkin, vuoden 2022 aikana rahoituspäätöksille on Cordiksessa alkanut ilmestyä DOI tunnisteita, (esim <https://doi.org/10.3030/101057264>) jotka ilmeisesti käyttävät vakioitua muotoa, jossa DOI:n loppuosa on EU:n Grant ID. Tutkimustietovaranto valmistautuu lataamaan nämä uudet tunnisteet.

Kansalliset rahoituspäätökset

Tiedossamme ei ole, että mikään kansallinen tutkimusrahoittaja hallinnoisi pysyviä tunnisteita omille päätöksilleen. Kaikki käyttävät omia paikallisia tunnisteitaan, ja tyytyvät näiden käyttöön. Tämä on jo aiheuttanut konflikteja kun esim Suomen Akatemian paikallinen tunnus on ollut identtinen toisen rahoittajan tunnuksen kanssa.

Knowledge Exchangessa vastikään julkaistussa raportissa tutkimusrahoittajien roolista PID-kentällä¹⁸ nostetaan esille ilmeisimmät hyödyt yhteisestä maailmanlaajuisesta, uudesta ID:stä. Tässäkin todetaan että rahoittajan hallinnoimissa omia tunnistejärjestelmiä, riskinä on aina että tunnisteita kaksinkertaistetaan eri rahoittajien välillä. Myös inhimillisiä virheitä sattuu helposti kun kirjoittajat viittaavat teoksiinsa, eli rahoituksen tunnisterakenteen tulee olla mahdollisimman yksinkertainen ja myös tukea koneluettavuutta. Lisäksi, rahoittajien

¹⁷ Opetushallituksen palvelukokonaisuus.. Palvelukortti: Organisaatiopalvelut. (20.2.2023) <https://wiki.eduuni.fi/display/OPHPALV/Organisaatiopalvelu>

¹⁸ de Castro, Pablo, Herb, Ulrich, Rothfritz, Laura, & Schöpfel, Joachim. (2022). The role of research funders in the consolidation of the PID landscape. Zenodo. <https://doi.org/10.5281/zenodo.7258210>

yhtenäinen tunniste ID mahdollistaisi PID Graphin hyödyntämistä, mikä mahdollistaisi myös kaikkien julkaisujen ja tietoaaineistojen linkittämistä koneluettavasti suoraan tutkimusprojekteihin. Tämä yhtenäisyys vaikuttaa myös rahoittajiin päin suotuisalla tavalla kun pystyvät helposti seuraamaan projektien laskeutumissivuilta mitä rahoittamisissaan projekteissa on saatu aikaan.

Tutkimushankkeet

RAiD¹⁹ (Research Activity identifier) on mitä todennäköisimmin tulevaisuuden tunniste tutkimushankkeille. Teknisesti RAiDit ovat DOI-, Handle- tai Cool URI-tunnisteita eli suositeltavaa on DOI-muotoisen RAiDin käyttö, joka on tulossa EOSCin palveluna lähivuosina.

Palvelut ja infrastruktuurit

Tutkimustietovarannossa luodaan URN-tunnisteita infrastruktuureille. Palveluille ei ainakaan vielä luoda tunnisteita, mikä kuitenkin helpottaisi niiden metatietojen hyödyntämistä esimerkiksi palvelukatalogeissa tai aineistonhallinnan suunnittelussa. Lisäksi käyttön on kansainvälisesti tulossa tunnisteet laitteille.²⁰

¹⁹ RAiD Research Activity Identifier Service. (20.2.2023)

<https://ardc.edu.au/services/ardc-identifier-services/raid-research-activity-identifier-service>

²⁰ PIDINST White paper. Persistent Identification of Instruments (PIDINST)

working group output report <https://docs.pidinst.org/en/latest/white-paper/index.htm> |

Tavoitetila ja tarvittavat askeleet sen saavuttamiseksi

Tässä luvussa kuvataan tarvittavia toimenpiteitä pysyvien tunnisteiden käytön ja hallinnoinnin kehittämiseksi tieteen ja tutkimuksen käyttökontekstissa.

Suomen PID-verkosto aloitti tämän tiekarttatyön 2021 tuottamalla yhteisen tavoitetilan kuvauksen, joka on seuraava:

Jäljitettävä ja yksilöity tieto (objekteista) löytyy ja on luotettavasti linkitettävissä nyt ja tulevaisuudessa

Seuraavaksi kuvataan konkreettisesti niitä askelia, joita tämä asiantuntijaryhmä katsoo tarpeelliseksi tavoitetilan saavuttamiseksi.

Tavoite	Kuvaus	Toimenpiteet
Luottamuksen ja yhteistyön rakentaminen	Opitaan tuntemaan toisia toimijoita, jaetaan tietoa ja verkostoidutaan	-rakennetaan yhteistyötä toimijoiden välillä PID-asioihin liittyen, mm: - tutkimusyhteisö - PID-palvelut - KAM -sektori - OKM, DVV, VM - linkitetyn datan yhteisö -tiedon laatuun liittyvä työ - tuotetaan PID-politiikka -selkeytetään tarpeita ja hyötyjä
Kansallisen infrastruktuurin järjestäminen	Organisoidutaan, suunnitellaan ja varmistetaan rahoitus	-kehitetään PID-työn alustaa -tunnistetaan kehittämistarpeet -sovitaan vastuut rahoituksen selvittäminen
Tietoisuuden ja osaamisen lisääminen	Huolehditaan tiedon jakamisesta ja koulutuksista	-viedään PID-asioita mm. AVOTT-koulutusryhmien agendalle -kehitetään kansainvälistä verkostoitumista ja toimintaa järjestelmällisesti

Luottamuksen ja yhteistyön rakentaminen

Kansallisen tason datanhallinnan ja yhteentoimivuuden tilan selvittäminen ja PID-politiikan laatiminen yhdessä olisi hyvä pohja luoda yhteistyötä ja yhteisiä linjauksia.. Tutkimuspuoli ei voi tehdä koko pysyvien tunnisten kenttää yksin, vaan tarvitaan laajempaa viitekehystä esimerkiksi valtionhallinnon kanssa, jotta yhteentoimivuus, käytettävyys ja tehokkuus voidaan taata osana datanhallintaa.

Kun rakennetaan luottamusta tunnisteiden tarjoajien kesken sekä heidän ja tunnisteita hyödyntävien tahojen välille, on yhteisistä pelisäännöistä sopiminen keskiössä. Niillä taataan, että kaikilla osapuolilla on sama ymmärrys kunkin tunnistepalvelun käyttötarkoituksesta ja sen ylläpidosta. Koska toimitaan kansainvälisellä areenalla, myös kansainvälinen näkökulma kuten EOSC PID Policy on otettava huomioon pelisääntöjä laadittaessa.

Tavoitteena tulisi olla rakennettaa toimintatapa, jossa pystytään toimimaan itseorganisoituvasti yhteisiä suuntaviivoja seuraten. On yhdessä kuvailtava millaisia kaikkien käytössä olevia tietoteknisiä palveluita (=järjestelmiä) tarvitaan. VM:n ja tiednhallintalautakunnan osalta toivotaan myös ohjausta ja pysyvien tunnisteiden tunnistamista osaksi tiedonhallintakarttaa ja kokonaisarkkitehtuuryötä. Formaalien sopimusten kautta voidaan myös organisoida toiminnan ennustettavuutta ja hallintaa. Esimerkiksi Kansalliskirjasto valmistele parhaillaan URN-sopimusta, joka formalisoi URN-tunnisteiden käyttöä ja hallintaa.

Luottamusta lisäävät asiat

Luottamusta rakennetaan eri toimijoiden välille: tutkijan ja palveluntarjoajan välille, palveluntarjoajan ja PID-palvelun tuottajan välille. Kansalliset PID-linjaukset tukevat palveluntarjoajien ja PID-palvelun tuottajien välistä luottamusta, mikä vaikuttaa myös tutkijan ja palveluntarjoajan suhteeseen.

Luottamus loppukäyttäjän ja palveluntarjoajan välillä perustuu seuraavanlaisiin tekijöihin:

- Palvelun ja sen tarjoajan tunnettuus
- Palvelun dokumentaatio ja avoimuus
- Palvelun tarjoaman sisällön laatu (onko tilastopalvelua, metatietomalli, näkyvyys)

Luottamus palveluntarjoajan ja PID-palvelun tuottajan välillä

- Dokumentaatio
- Palvelun taso (esim. tuki, onko aina saatavilla)
- Palvelun luotettavuus (vastaako lupaus toteutumaa)
- Sopimukset
- Toimijoiden luotettavuus
- Auditoinnit ja sertifikaatit

Luottamus toimivaan PID-infrastruktuuriin kansallisesti

- Selkeät toimintaperiaatteet ja -ohjeet
- Selkeät vastuut
- Pysyvä rahoitus
- Pysyvien tunnisteiden hyötyjen osoittaminen

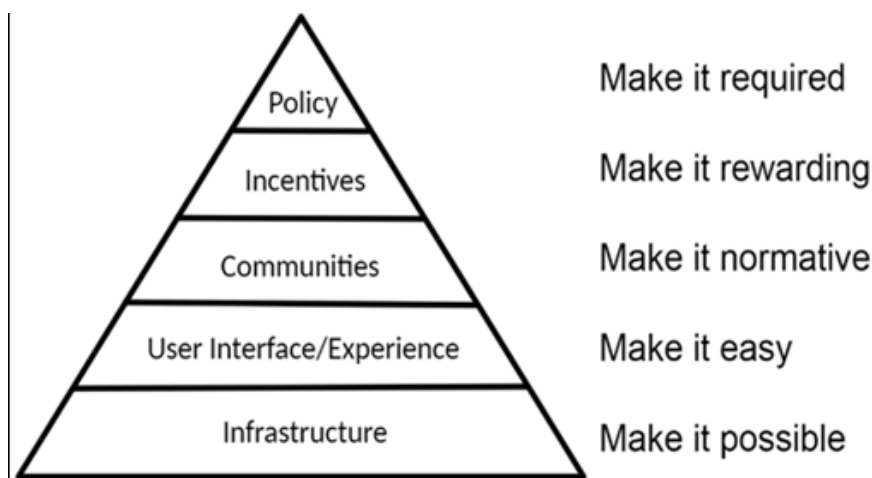
Tarvittavat toimenpiteet

1. Vakiinnutetaan yhteistyö ja asemoidaan se suhteessa muihin toimijoihin (etenkin OKM, Kansalliskirjasto, Kansallisarkisto, VM, JulkICT, DVV ja muut ministeriöt ja toimijat jotka tuottavat tutkimusaineistoja esim STM, Findata, Tulanet sekä Linked data -yhteisö)
2. Tuotetaan uusi tarkempi PID-politiikka
 - a. PID-verkoston tulee laatia ja ylläpitää suosituksia siitä, mitä tunnisteita käytetään eri objekteille ja käyttökonteksteihin liittyen.
 - b. Poliitikassa huomioidaan eri toimijat ja heidän roolinsa
 - c. Sovitaan kansallisesti millä perusteilla valitaan pysyvien tunnisteiden skeemoja ja palveluita.
 - d. Sovitaan kansallisesti miten paikataan tunnistetut puutteet PID-tarjonnassa.
2. Tunnistetaan ja tarvittaessa luodaan tarvittavat muut dokumentit: muut tarvittavat PID-politiikat, sopimukset, sertifikaatit
3. Varmistetaan toiminnan avoimuus – Knowledge exchangen selvityksestä luottamusta edistävät asiat
4. Osoitetaan pysyvien tunnisteiden käytön hyödyt käyttötapauksesimerkein, vaikuttavuusarviointia (vrt JISC, ARDC, RDA selvitys). Huomioidaan myös varoittavat esimerkit.

Kansallisen infrastruktuurin järjestäminen

Suomalainen tunnisteinfrastruktuuri on muodostunut organisesti tarpeisiin ja luonteviin vastuisiin perustuen eikä keinotekoinen keskittäminen ole tarpeen. Pääpainon tulee olla siinä, että edellä listatut tavoitteet toteutuvat ja kaikki tunnistetarpeet katetaan riittävän resursoinnin avulla. Infrastruktuurin tulee olla kestävä. Tärkeää on yhteentoimivuus ja luotettavuus sekä toiminnan yhteistyö. Hallinnoinnin tulisi olla toimintamallikeskeistä eikä toimijakeskeistä.

Riittävä rahoitus mahdollistaa investoinnin palveluihin jotka puolestaan mahdollistavat toimintakulttuurin muutoksen ja sen myötä säästöjen syntymisen tiedonhallinnan parantumisen ja tehostumisen kautta.



Theory of change model (Nosek B., 2019)

Riittävän rahoituksen turvin voidaan:

- turvata palveluiden luotettava toiminta ja järjestelmien resilienssi riittävällä avoimuudella ja monimuotoisuudella
- varmistaa kestävä kehitys (vähennetään päällekkäisyyksiä, hyvä datanhallinta, sähkö on kallista)
- tukea toimijoita tunnistamaan ydintietonsa ja tarjoamaan niitä rajapintojen kautta pysyvin tunnistein muiden käyttöön
- ottamaan pysyvät tunnisteet osaksi yleistä tiedonhallintaa (tiedonhallintamallit ja yhteentoimivuus)
- Pitää yllä kerran myönnettyjen tunnisteiden linkitystä uudempiin järjestelmiin (jos tulevaisuudessa syntyy rinnakkaisia järjestelmiä)
- varmistaa kansainvälinen yhteentoimivuus ja kontribuutio kansainvälisiin järjestelmiin, standardeihin ja palveluihin

Pysyviä tunnisteita tarjoaa usein toinen taho kuin se, joka vastaa metatiedoista tai sisältöjen laadusta ja eheydestä. Tavoitetilassa PID-järjestelmät ovat kiinteä osa muuta datan hallintaa. Loppukäyttäjän näkökulmasta objektitsaisivat PIDit 'automaattisesti' eri palveluissa kuten tutkimusinfrastruktuureissa tai data-arkistoissa. Tarvittavat PIDit tulevat määrittelyksi loppukäyttäjän toimien kautta, loppukäyttäjät haluavat helppoja, automatisoituja ratkaisuja. Tämä tarkoittaa, että aineistojen omistajat tai aineistopalveluiden tarjoajat ovat vastuussa pysyvien tunnisteiden valinnasta ja implementoinnista.

Saatavilla on useita PID-järjestelmiä, jotka voivat olla yksityisten toimijoiden tuottamia tai kansallisten organisaatioiden tuottamia. Suomessa ei liene tarvetta ryhtyä tuottamaan uusia uutta PID-järjestelmiä tutkimuksen tarpeisiin vaan voimme valita saatavilla olevista palveluista. Eri PID-järjestelmiä pyritään käyttämään erilaisille objekteille, joten järjestelmiä tulee olemaan useita.

Kansallisten suositusten luominen järjestelmien valintaa varten voidaan tehdä osana yhteisiä linjauksia. Tällaisen kriteeristön pohjaksi syntynee samalla perusteltuja linjauksia. Esimerkki linjauksesta voisi olla, että PID-järjestelmän valinnassa täytyy huolehtia PIDien pysyvyyden toteutumisesta eli PID-järjestelmän ja siihen liittyvien PID-palveluntuottajiin tulisi olla sitoutuneita ja varautuneita muutoksiin ja yksi kriteeri voisi olla esim. ISO 27001 sertifikaatti palvelutuotannon järjestämisen varmistamiseksi. Tällaisia kriteereitä ollaan luomassa EOSC-hankkeissa.

Suositukset auttavat toimijoita valitsemaan sopivia ratkaisuja. Tunnistejärjestelmät ovat myös osa kansainvälistä infrastruktuuria ja tarvitsevat sekä hallintoa että palveluita. Osa PID-järjestelmistä on vahvasti ulkomaisten organisaatioiden hallinnassa (esim. ORCID). Tällöin standardointi on erityisen tärkeää, sillä se tuo läpinäkyvyyttä ja ennakoitavuutta toimintaan. Osa tunnisteista (esim. CrossRef DOI) on puolestaan vahvasti kaupallisten toimijoiden hallussa. Julkishallinnon puolella tietenkin Euroopan tiedonhallinnan ohjaus

Eurooppalaisella tasolla PID-järjestelmien hallinnointiin ja yhteisten linjausten tekemiseen on tahtotilaa. Linjaukset tulevat vaikuttamaan myös kansallisella tasolla infrastruktuurin järjestämiseen. eArchiving Common Services Platform (EARK-CSP) tuotetaan ns. synergiakokonaisuuksia Digital Europe hankkeen suuntaan. Yhtenä työkohteena on tukea

pysyvien tunnisteiden hyödyntämistä ja tuottaa yleiset suositukset sovellettavista standardeista ja käytännöistä.

DOIn osalta hallinnointi on kolmiportainen: International DOI Foundation, Registration Agency (Crossref / DataCite) ja kansallinen sopimusosapuoli (CSC/TSV). Löysimmin hallinnoituja ovat ARK, Handle ja muutamat URN-nimialueet kuten URN:UUID joilla ei ole keskitettyä valvontaorganisaatiota, ja tunnuksen käyttäjiä sitoo vain standardi ja ARKIn ja Handlen tapauksessa resolverisovellukset.

Kansainvälinen kehitys tulee todennäköisesti selkeyttämään tilannetta lähivuosina ja helpottaa siten PID-ratkaisujen valintaa ja integrointia tai soveltamista eri järjestelmiin.

Tarvittavat toimenpiteet

1. Kehitetään yhteistä alustaa, jolla ylläpidetään PID-politiikkaa ja muuta PID:eihin liittyvää tietoa.
2. Tunnistetaan kansainvälinen konteksti, seurataan kehitystä ja toimitaan ennakoivasti ja koordinoitusti myös kansainvälisesti
3. Selvitetään miten aukkoja ja puutteita voidaan ratkoa.
4. Sovitaan kansallisesti miten palvelut rahoitetaan ja organisoidaan tutkimuksen käyttökontekstissa eri katteille.

Tietoisuuden ja osaamisen lisääminen

Yhteinen käsitys PID-toimintakentän nykytilasta sekä siihen liittyvistä tavoitteista on edellytys tehokkaalle viestinnälle ja tietoisuuden lisäämiselle. Yhteistyön rakentamista on käsitelty yllä, mutta yhteistyössä saavutettujen tulosten ja hyötyjen viestiminen keskeistä. Tietoisuutta pysyvistä tunnisteista tulee lisätä kullekin kohderyhmälle merkityksellä tavalla. Onnistunut viestintä konkretisoi linjaukset ja osoittaa hyötyjä kullekin ryhmälle.

Jokaisen tutkijan on hyvä olla perillä siitä, mitä PID-tunnistetta voi ja tulisi kulloinkin käyttää ja mistä sen voi hakea mutta parasta olisi, jos kaikki tapahtuisi automaattisesti osana tiedon tuottamisprosessia ja palveluiden toimesta. Tarjonnan pitäisi olla kattavaa niin, että mikään PIDiä tarvitseva objekti ei jää ilman tunnusta. PID-tunnisteiden hankintaprosessit tulisi mahdollisimman pitkälle automatisoida sekä integroida käyttöliittymiin ja järjestelmiin siten, ettei niiden käyttö vaikeuta palveluiden hyödyntämistä ja että tiedot linkittyvät optimaalisella tavalla. Datanhallinnan palveluissa valittujen PID-järjestelmien käyttö tulee ohjeistaa hyvin käyttäjälähtöisesti. Milloin niistä saa hyötyä ja milloin ne ovat välttämättömiä?

Julkishallinnon toimijoille tulee avata PID:ien merkitystä pysyvälle säilytykselle ja kansallisen perinnön jättämiselle, jotta valtion toimijat ymmärtävät miksi pysyviä tunnisteita tulee rahoittaa ja hallinnoida. Sama koskee myös tutkijan oman organisaation edustajia.

Tällä hetkellä kaikkea PID:ien soveltamiseen tarvittavaa tietoa ei ole kerätty yhteen paikkaan.. Tämän vuoksi kokonaiskuvan muodostaminen on vaikeaa. Perustietoa PID:eistä saa Kansalliskirjaston julkaisemasta Elektronisten julkaisujen identifiointioppaasta, mutta tunnistekohtaisia käyttöoppaita ei ole vielä laadittu.

Tietoisuuden lisäämisen lisäksi tarvitaan koulutusta, jonka se tulisi pohjautua tunnistettuihin osaamistarpeisiin. Osaamistarpeiden tunnistus seuraavilla sektoreilla tulisi olla ensimmäisiä

askelia osaamisen edistämiseksi: toimijat ja konsortiot, organisaatiot ja tukipalvelut sekä tutkijat. Yleensä kansallisella tasolla yhteentoimivuutta kehitetään Digi- ja väestövirastossa.

Osaamisen lisääminen tapahtuu parhaiten esimerkkien kautta opettamisella. Esimerkkitapausten esille tuominen koulutusmateriaalissa on näin ollen erityisen tärkeää. Osaaminen ja tietoisuus PID:eistä karttuu myös kun joutuu etsimään ja hyödyntämään avointa dataa eri lähteistä, jolloin samalla havainnoi helposti PID:ien käyttötarpeita. Kansainvälisen kentän kehitystä tulee myös tiiviisti seurata, erityisesti standardien ja teknisten ratkaisujen osalta ja sopia yhteisistä käytänteistä yhdessä yli maan rajojen.

Tutkimuksen tuki on avainasemassa ja heidät tulisi kouluttaa, jotta he osaavat suositella oikeita ratkaisuja tutkijoille. Harvoin tutkija itse valitsee tunnisteiden sillä hän valitsee ensisijaisesti palvelun. Hänelle tulisi silti avata eri tunnisteiden taustat, sekä niihin liittyvät hyödyt ja mahdolliset haasteet.

Tieteenalojen sisällä tarvitaan laajaa datanhallinnan osaamista, jonka osa osaaminen pysyvistä tunnisteista on. Koulutus pysyvistä tunnisteista on siis myös kotiorganisaatioiden harteilla, mutta tieto käytössä olevista PID-järjestelmistä tulee saattaa helposti saatavilla myös kouluttajille.

Tarvittavat toimenpiteet

1. Luodaan wiki-sivuston lisäksi/päälle yleistajuisempi laskeutumissivusto, jolle kerätään käyttöoppaita ja muuta ajantasaista tietoa pysyvistä tunnisteista. Tällä PID:eistä kertovalla sivustolla tieto tulisi organisoida tunnistejärjestelmän kohtaisesti lukijan roolin mukaan ja sen sivuilla tulisi olla määriteltynä omistajat, jotka pitävät niitä ajan tasalla.
2. Järjestetään koulutusta. Seuraavat osaamistarpeet on tunnistettu:
 - a. toimijat ja konsortiot
 - b. tunnistejärjestelmien tuntemus
 - c. tarjolla olevien PID-palveluiden tuntemus
 - d. kansalliset ja tarpeellisin osin kansainväliset palvelut
3. Koulutuksen kohderyhmiä ovat ainakin:
 - a. Organisaatiot, palvelut ja tukipalvelut:
 - i. Pysyvien tunnisteiden käyttötapaukset
 - ii. Pysyvien tunnisteiden
 - iii. Pysyvien tunnisteiden palvelut
 - b. Tutkijat:
 - i. ORCID, mikä on ja miten toimii
 - ii. viittaaminen, paitsi data myös muut asiat
 - iii. palveluiden valitseminen
4. DVV ottaa osaksi valistustoimintaansa pysyvät tunnisteet

Tutkijoille laaditaan ohjeistus tutkimuksen elinkaaren eri vaiheisiin liittyen ja miten PID:ejä käytetään ja luodaan ja ylläpidetään. Tähän voidaan perustaa työryhmä Avoimen tieteen ja tutkimuksen koordinaation piiriin.