

# Tunnisteet osana tutkimusaineistojen metatietoja

Juha Hakala  
Kirjastoverkkopalvelut  
2022-04-26

# Haaste

- Data-aineistoja pitää identifioida monella tasolla:
  - Kokoelmataso (esim. samaan hankkeeseen liittyvät data-aineistot)
  - Teostaso yhteinen kaikille aineiston manifestaatioille)
  - Manifestaatiotaso (osakohteineen)
  - Metatiedot (tietueiden tunnisteet)
- Data-aineistojen lisäksi meidän on identifioitava myös:
  - Toimijat (tutkijat, tutkimusorganisaatiot, tutkimushankkeet)
  - Tutkimusjulkaisut (jotka tulee linkittää tunnisteilla data-aineistoihin)
  - Data-aineistoja kuvaavat metatiedot
  - Tutkimuslaitteistot ja tutkimuksessa käytetyt ohjelmistot
- Tämä esitys keskittyy data-aineistojen identifiointiin

# Tunnistejärjestelmistä

- Perinteisesti tunnisteet kehitettiin yhteen tarkoitukseen
  - ISBN – monografiat; ISSN – kausijulkaisut
- Uusia ”perinteisiä” tunnistestandardreja kehitetään yhä
  - RAiD (Research Activity Identifier) tutkimushankkeille
- Perinteisten tunnisteiden rinnalle on tullut toiminnallisia tunnisteita (persistent identifiers, PID), jotka tarjoavat palveluita verkossa
- Palveluvalikoima on toistaiseksi suppea (linkki laskeutumissivulle tai identifioituun resurssiin), mutta tulee laajenemaan

# Nykytilanne

- Kaikkia tarvittavia tunnistejärjestelmiä ei ole vielä olemassa
  - Esim. tutkimushankkeiden RAiD (Research Activity Identifier) valmistuu syksyllä 2022; siirtyminen tuotantoon voi kestää vielä pitkään
- Vaikka sopiva tunnistejärjestelmä olisi olemassa, siitä ei aina ole apua
  - Tutkimusorganisaatioiden ROR-tunniste (Research Organization Registry) ei ole saanut tuulta purjeisiin, koska se edellyttää organisaation tekevän kuvailutyön itse
  - DOI-tunnusta voi olla vaikea hankkia ilman tunnisteiden käyttöä koskevaa sopimusta
- Toiminnallisten tunnisteiden tulevaisuus on hämärän peitossa
  - Kuka vastaa toiminnallisuuden ylläpidosta sitten kun nykyisiä vastuorganisaatioita (kustantaja, korkeakoulu) ei enää ole?

# DOI

- DOI (ISO 26324) on "Digitaalinen objektien tunnistejärjestelmä", mutta sen keskeiseksi soveltamisalaksi on muotoutunut tieteelliset julkaisut ja tutkimusaineistot
- Tunnisteen soveltamista ohjaa kansainvälinen rekisteriviranomainen, International DOI Foundation
- IDF on nimennyt 10 Registration Agencyä (RAG) eli rekisteriviranomaista, jotka tekevät DOI-tunnisteen käyttö sopimuksia kolmansien osapuolien kanssa
- Tutkimusjulkaisujen ja -aineistojen osalta keskeiset RAG-organisaatiot ovat DataCite ja Crossref

# DOI ja tutkimusaineistot Suomessa

- CSC välittää DataCiten DOI-tunnuksia
- Data-aineistoja identifioitaessa käyttäjän on noudatettava sekä ISO-standardin, DataCiten että CSC:n kanssa solmittavan DOI-käyttösopimuksen ehtoja
  - <https://support.datacite.org/docs/metadata-quality>
  - <https://support.datacite.org/docs/doi-registration-policy>
    - Tunnuksot linkittyvät laskeutumissivuille, joita koskevat ehdot on määritelty sivulla <https://support.datacite.org/docs/landing-pages>
- DataCiten DOI on teostason tunnus, eli data-aineiston manifestaatiot tai niiden osakohteet eivät voi saada omia tunnuksia, jotka ovat tarpeen esim. pitkäaikaissäilytyksen takia

# URN-tunnistejärjestelmä

- URN eli Uniform Resource Name on Internet Engineering Task Forcen (IETF) standardi RFC 8141 (<https://www.rfc-editor.org/info/rfc8141>)
- URN-tunniste on hallinnollisesti ja toiminnallisesti hajautettu; ei ole yhtä valvovaa tahoa eikä yhtä URN-sovellusta
  - URN:DOI-tunnukset resolvoituvat normaalin DOI-tunnuksen tapaan
- URN jakautuu noin 50 nimialueeseen, joiden kate vaihtelee
  - URN:ISBN - monografiat, URN:DEV – Internet of Things, URN:LEI – Legal entity identifier –tunnus, URN:UUID, jatkossa myös URN:DOI
  - <https://www.iana.org/assignments/urn-namespaces/urn-namespaces.xhtml>
- Joidenkin URN-nimialueiden tunnukset eivät ole toiminnallisia; esimerkiksi URN:DEV-tunnus on vain laitteen osoite

# Kansalliskirjastojen URN:NBN-tunniste

- NBN = National Bibliography Number
- URN-nimialuerekisteröinnin tavoite oli luoda kansalliskirjastoille ja niiden yhteistyökumppaneille tunnistejärjestelmä, jonka avulla voidaan identifioida kaikki se, jota perinteiset tunnistejärjestelmät eivät kata
- Tunnus jakautuu maakohtaisesti ja tarvittaessa edelleen toimijakohtaisesti (URN:NBN:FI:)
- Käytön ehdot ovat väljät verrattuna DOI-tunnisteeseen
  - Tunnistejakelua rajoittavat URN-standardi ja NBN-nimialuerekisteröinti (RFC 8458, <https://www.rfc-editor.org/rfc/rfc8458.html>)



# URN ja tutkimusaineistot

- URN:NBN-tunniste täydentää DataCiten DOI-tunnuksen katetta
  - URN-tunnus voidaan antaa tutkimusaineiston manifestaatioille, niiden osakohteille ja lisäksi myös tutkimusaineistoja kuvaaville metatiedoille ja metatiedon elementeille kuten aineistoa kuvaaville termeille
  - Ainoa merkittävä raja on se, että tunnusten pitäisi olla toiminnallisia
- Tunnukset ovat jatkossakin maksuttomia (koska URN-tunnusten on oltava ilmaisia), mutta Kansalliskirjaston URN-palveluille voi tulla jatkossa sopimuksin määritelty hinta (ja sisältö)
- URN:NBN-nimialueen lisäksi voidaan soveltaa muita URN-nimialueita tarpeen mukaan; esim. URN:DEV laitettunneena

# Mitä toiminnallista tunnistetta tulisi soveltaa?

- DOI ja URN täyttävät tunnetut tarpeet, mutta niitä täydentää Handle, koska sen soveltamisessa ei ole rajoituksia eikä mitään sopimuksia tarvita
- Data-aineistoille ei ole kehitteillä kansainvälistä standarditunnistetta a'la ISBN eikä sellaiselle ole enää tarvetta, koska DataCiten DOI on "vallannut markkinan"
  - Haasteena DataCite-sopimuksen asettamat rajoitukset
- Koska CSC:n Fairdata PAS –palvelu edellyttää manifestaatiotason tunnuksia, pitkäaikaissäilytettävät data-aineistot voivat saada sekä teostason DOI-tunnukset että manifestaatioiden URN:NBN-tunnukset

# Muuta huomionarvoista

- Perinteisissä tunnistejärjestelmissä tunnuksen antajan (julkaisijan) vastuut ovat selkeät ja rajalliset
  - Kun tunnus on kerran annettu, ei enää tarvitse tehdä mitään
- DOI- ja URN-tunnisteet vaativat tunnusten toiminnallisuuden (pysyvää) ylläpitämistä
- Sopimuksilla voidaan edellyttää tunnisteita koskevien vastuiden siirtymistä aineistojen uudelle omistajataholle, mutta mitä tehdään silloin, kun vastuuorganisaatio katoaa?
- Kulttuuriaineistolaki takaa sen että Kansalliskirjasto säilyttää e-julkaisut, mutta laki ei sano mitään näiden julkaisujen toiminnallisten tunnisteiden ylläpidosta eikä se kata tutkimusaineistoja