

16  40

KANSALLISKIRJASTO  
NATIONALBIBLIOTEKET

# Elektronisten julkaisujen identifiointi

Nimeke: Elektronisten julkaisujen identifiointi. Versio 1.0

Julkaisuvuosi: 2021

Tekijä: Kansalliskirjasto

Julkaisija: Kansalliskirjasto

Lisenssi: CC-BY-NC

Kansilehden kuva: Pixabay / xresch

Pysyvä tunniste: URN:ISBN:978-951-51-7661-5

# Elektronisten julkaisujen identifiointi

## Sisältö

<b>1. Johdanto</b>	<b>5</b>
<b>2. Yleistä PID-tunnistejärjestelmistä</b>	<b>6</b>
<b>3. PIDit ja perinteiset tunnistejärjestelmät</b>	<b>7</b>
<b>4. Uniform Resource Identifier (URI)</b>	<b>9</b>
<b>5. Toiminnalliset tunnisteet – yleistä</b>	<b>11</b>
<b>6. ARK</b>	<b>13</b>
6.1 ARK-järjestelmän taustaja tukiorganisaatio	13
6.2 ARK-järjestelmän soveltamisalat ja keskeiset käyttäjät	13
6.3 ARK-tunnuksen rakenne	13
6.4 ARK julkaisujen tunnisteena	13
6.5 ARK-järjestelmän tulevaisuuden näkymät	15
<b>7. URN</b>	<b>16</b>
7.1 URN-järjestelmän tausta ja tukiorganisaatio	16
7.2 URN:n soveltamisalat ja keskeiset käyttäjät	16
7.3 URN-tunnuksen rakenne	17
7.4 URN-tunnistejärjestelmän palvelut ja niiden toteuttaminen	18
7.5 URN julkaisujen tunnisteena	19
7.6 National Bibliography Number (NBN)	19
7.7 URN:n tulevaisuudennäkymät	20
<b>8. The Handle System</b>	<b>21</b>
8.1 Handle-järjestelmän tausta ja tukiorganisaatio	21
8.2 Handlen soveltamisalat ja keskeiset käyttäjät	22
8.3 Handle-tunnuksen rakenne	22
8.4 Handle julkaisujen tunnisteena	22
8.5 Handlen tulevaisuuden näkymät	23

## **9. DOI 24**

9.1 DOI-järjestelmän tausta ja tukiorganisaatiot	24
9.2 DOI:n soveltamisalat ja keskeiset käyttäjät	24
9.3 DOI-tunnuksen rakenne	26
9.4 DOI kirjan tunnistena	26
9.5 DOI artikkeleiden tunnistena	27
9.6 DOI:n tulevaisuudennäkymät	27
9.7 DOI:n ja Handlen suhde	28

## **10. PID-tunnusten käyttö Suomessa 29**

10.1 URN	29
10.2 The Handle System	29
10.3 DOI	29
10.4 Yhteenveto	31

## **Sanasto 32**

# 1. Johdanto

Tämä opas antaa perustiedot elektronisten julkaisujen identifiointista toiminnallisten tunnuksien (Persistent IDentifiers, PID) avulla. Se on suunnattu ensisijaisesti kirjasto- ja kustannusalan ammattilaisille, mutta siitä on toivottavasti hyötyä myös muille asiasta kiinnostuneille.

Opas kuvaa Suomessa käytettävät PID-tunnistejärjestelmät, niiden sovellusalat ja vastuorganisaatiot. Opas pyrkii auttamaan lukijaa PID-tunnistejärjestelmän valinnassa ja PID-tunnusten soveltamisessa.

Kirjastot, kustantajat ja kirjakaupat ovat jo 1970-luvun alusta lähtien käyttäneet tunnistejärjestelmiä julkaisujen identifiointiin. Kirjojen tunniste ISBN ja jatkuvien julkaisujen tunniste ISSN ovat tehneet mahdolliseksi muun muassa elektronisten tilausjärjestelmien luonnin. Tämä on helpottanut oleellisesti esimerkiksi julkaisujen hankintaprosessia. Kirjastojen viitetietokannoissa nämä tunnisteet ovat helpottaneet tiedonhakua ja julkaisujen metatietojen käsittelyä.

Elektronisen julkaisemisen ja Internetin myötä perinteisten standarditunnusteiden rinnalle on kehitetty uusia järjestelmiä:

- Toimijoiden julkisten identiteettien tunnistejärjestelmät, kuten International Standard Name Identifier (ISNI), Open Researcher and Contributor ID (ORCID) ja Research Organization Registry (ROR).
- Teosten ja ekspressioiden tunnistejärjestelmät, kuten International Standard Audiovisual Number (ISAN).
- Toiminnalliset tunnistejärjestelmät (Persistent Identifier - eli PID-järjestelmät), eli ARK, DOI, Handle ja URN).

PID-järjestelmien avulla on mahdollista tarjota erilaisia Web-palveluita, kuten luoda identifioitujen objektien sijainnista riippumattomia hyperlinkkejä joko näihin objekteihin itseensä tai esim. objektien metatietoihin. Linkkien hallinnointi ja täydellinen riippumattomuus resurssien verkkosijainneista parantaa linkitetyn datan luotettavuutta.

PID-järjestelmät ovat luoneet myös uusia haasteita. Perinteisten standarditunnusteiden käytön periaatteet ovat tuttuja niin kustantajille kuin kirjastoillekin. Niiden soveltamisala on vuosikymmenien mittaan tarkasti rajattu ja perusteellisin manuaalein ohjeistettu. PID-järjestelmien tilanne ei ole yhtä selkeä. Esimerkiksi DOI-tunnuksen voi ISO 26324 -standardin mukaan antaa periaatteessa mille tahansa objektille. Mutta DOI-tunnuksien jakelua hallinnoivat organisaatiot kuten Crossref ovat asettaneet omien DOI-tunnustensa soveltamiselle ehtoja, joita ei saa rikkoa. Tämän oppaan DOI-osiossa kuvataan Crossrefin ja DataCiten DOI-ohjeita tarkemmin, koska Suomessa annetaan niiden DOI-tunnuksia muun muassa tieteellisille artikkeleille ja tutkimusdata-aineistoille. Näin sovellettuna DOI täydentää ISO:n perinteisiä tunnistejärjestelmiä, jotka eivät näitä aineistoja kata. Muiden PID-järjestelmien soveltamisaloja käsitellään lyhyesti niiden kuvauksissa.

PID-järjestelmät ovat kehittyneet viime vuosina nopeasti, ja muutoksia on odotettavissa jatkossakin. Siksi tästä oppaasta on tarkoitus julkaista uusia versioita tarpeen mukaan. Teksti on laadittu Kansalliskirjastossa, mutta DOI-järjestelmän kuvaukseen on saatu arvokasta palautetta Tieteen tietotekniikan keskus CSC:stä ja Tieteellisten seurain valtuuskunnasta, mistä lämpimät kiitokset asianosaisille.



## 2. Yleistä PID-tunnistejärjestelmistä

Toiminnallisia tunnisteita eli PID-tunnisteita ovat Archival Resource Key<sup>1</sup> (ARK), Uniform Resource Name<sup>2</sup> (URN), The Handle System<sup>3</sup> (Handle) ja Digital Object Identifier<sup>4</sup> (DOI) eli suomeksi Digitaalinen objektioiden tunnistejärjestelmä<sup>5</sup>. Uusia PID-tunnistejärjestelmiä tuskin syntyy aivan lähiaikoina, sillä kaikki nykyiset järjestelmät ovat jo noin 20 vuoden takaa.

Tunnisteen toiminnallisuus tarkoittaa sitä, että sen avulla voidaan tarjota identifioituun resurssiin liittyviä palveluita. Palvelut tarjotaan resoluvoimalla eli linkittämällä PID-tunnus yhteen tai useampaan URL-osoitteeseen. Linkin takana voi olla identifioitu resurssi itse, sen laskeutumissivu (landing page) tai esimerkiksi resurssin metatietoja. Linkitys toteutetaan resolveiksi (resoluutiopalvelimiksi) kutsutuilla ohjelmistoilla. Niiden palveluvalikoima on toistaiseksi suppea, esimerkiksi Kansalliskirjaston URN-resolverin ensimmäisellä versiolla ainoa palvelu oli URN-tunnuksen linkitys yhteen identifioitun objektin verkko-osoitteeseen. Tulevaisuudessa resoluutiopalvelimien palveluvalikoima kasvaa; esimerkiksi Kansalliskirjaston tätä kirjoitettaessa vielä kehitteillä olevaan uuteen resolveriin on jo lisätty mahdollisuus linkittää URN-tunnus samanaikaisesti useisiin URL-osoitteisiin. Rakenteilla on vanhentuneiden URL-osoitteiden tallentaminen resolverille siten, että linkki luodaan identifioitun dokumentin verkkoarkistoissa oleviin kopioihin..

Resolverin tarjoamaa palvelua ei pidä sekoittaa Internetin HTTP-protokollan tarjoamaan uudelleenohjaukseen (redirect). Uudelleenohjauksessa vanhentunut tai muuten toimimaton URL-osoite korvataan uudella

joko pysyvästi (HTTP 301) tai väliaikaisesti (HTTP 302), mutta linkitys yhdestä URL-osoitteesta samanaikaisesti useampaan tai URL-osoitteen valinta halutun palvelun kuten metatietojen haun nojalla ei onnistu. PID-resoluutiossa on mahdollista vaihtaa linkitys uuteen URL-osoitteeseen myös silloin, kun vanha osoite on ollut niin sanottu syvälinkki eli esimerkiksi tietokannassa olevan tietueen URL. Niiden HTTP-uudelleenohjaus on teknisesti vaikeaa tai mahdotonta. Resoluutiossa saatu URL-osoite on tietenkin uudelleenohjattavissa muiden URL-osoitteiden tapaan.

Tarjottavien resoluutiopalvelujen ylärajan määrittävät elektronisten aineistojen ja niiden metatietojen hallintoihin käytettävät sovellukset kuten asiakasliittymät, julkaisuarkistot ja pitkäaikaissäilytys- eli PAS-järjestelmät. Niiden avulla voidaan tulevaisuudessa toteuttaa elektronisten aineistojen käyttäjille uusia, resolveiden avulla tarjottavia palveluita.

Asiakas voi esimerkiksi pyytää nähtäväkseen pitkäaikaissäilytyksen metadattaa, ja arvioida sen avulla elektronisen julkaisun alkuperäisen ja myöhemmin luotujen versioiden sisällön eroja, ja valita sen perusteella parhaiten itselleen sopivan version julkaisusta. Autenttisuutta arvostava tutkija voi valita alkuperäisen tiedostomuodon, jonka avaaminen voi edellyttää digitaalista arkeologiaa eli tarkoitukseen soveltuvien ohjelmistojen ja laitteistojen etsimistä ja soveltamista.

PID-tunnukset voivat elää identifioimiaan elektronisia objekteja pidempään. Kun alkuperäinen objekti on poistettu tai vaikeakäyttöinen, PID-linkki voi ohjata käyttäjän objektin uudempaan versioon. Ja jos sellaista ei ole käytettävissä, PID-tunnus voi tarjota linkin muistokiveen (tombstone), jossa kerrotaan kadonneesta objektista ja / tai vaihtoehtoisista tavoista halutun informaation saamiseksi. Koska PID-tunnukset ja niiden tarjoama toiminnallisuus eivät ole riippuvaisia identifioitujen objektien elinkaaresta, PID-järjestelmät ovat hyviä apuvälineitä elektronisten aineistojen hallinnoinnissa.

1 [https://n2t.net/e/ark\\_ids.html](https://n2t.net/e/ark_ids.html)

2 [https://en.wikipedia.org/wiki/Uniform\\_Resource\\_Name](https://en.wikipedia.org/wiki/Uniform_Resource_Name)

3 [https://en.wikipedia.org/wiki/Handle\\_System](https://en.wikipedia.org/wiki/Handle_System)

4 [https://en.wikipedia.org/wiki/Digital\\_object\\_identifier](https://en.wikipedia.org/wiki/Digital_object_identifier)

5 Muilla PID-järjestelmillä ei tätä kirjoitettaessa ole suomalaista nimeä, ja niihin sekä DOI:hin viitataan jatkossa vain lyhenteellä.

# 3. PIDit ja perinteiset tunnistejärjestelmät

Perinteisiä julkaisujen tunnistejärjestelmiä voidaan ja tulee hyödyntää esimerkiksi elektronisten kausijulkaisujen ja monografioiden identifiointissa, mutta verkkoaineistojen osalta PID-tunnisteet täydentävät perinteisten tunnistejärjestelmien puutteita. Alla esimerkkinä jatkuvat julkaisut kuten tietokannat ja verkkosivustot kattava ISSN-standardi:

- Perinteiset tunnistejärjestelmät eivät mahdollista virheetöntä Internet-hakua ja resoluutiota vastaavaa linkitystä. ISSN-tunnuksen avulla voi toki hakea verkkosivustoa Googlestä, mutta ISSN voi silloin sotkeutua samalta näyttäviin numerosarjoihin. URN:ISSN-tunnuksella tätä vaaraa ei ole.
- Perinteiset tunnistejärjestelmät luotiin painettuja julkaisuja varten, eivätkä ne skaalaudu kunnolla verkkojulkaisemisen tarpeisiin. Esimerkiksi ISSN-järjestelmässä on tilaa vain 10 miljoonalle tunnukselle, joten jatkuvasti päivittyviksi julkaisuiksi tulkittavissa olevia satoja miljoonia WWW-sivustoja voidaan identifioida vain hyvin rajallisesti.
- Jos perinteistä tunnusta ei voi antaa kaikille relevanteille aineistoille, mitä rajausperusteita tulisi soveltaa? Mille verkkosivustoille esimerkiksi ISSN pitäisi antaa? Valintaa helpottaa tieto siitä, että ISSN-tunnuksen asemesta sivustolle voidaan antaa esim. Handle- tai URN:NBN-tunnus.
- Perinteisten tunnisteiden jakelu voi olla työvoimavaltaista. Jokainen ISSN-tunnuksen saava julkaisu on luetteloitava, ja metatiedot on lähetettävä Pariisiin kansainväliselle ISSN-keskukselle globaaliin ISSN-tietokantaan tallennettavaksi. Jo tunnuksien antaminen verkkolehdistä on lisännyt kansallisten ISSN-keskusten työn määrää merkittävästi. PID-tunnuksia voidaan tarvittaessa luoda koneellisesti miljoonia.

PID-tunnisteet tukevat perinteisiä tunnistejärjestelmiä, mutta voivat myös tunkeutua niiden tontille. Ja vaikka näin ei olisikaan, tiedonhakijoille voi olla epäselvää, mitä PID-tunnus oikein identifioi.

PID-järjestelmät suhtautuvat perinteisiin tunnistejärjestelmiin eri tavoin. URN-tunniste perustuu olemassa oleville tunnistejärjestelmille, ja se on sen vuoksi varmasti sopuosinnussa niiden kanssa. Eräänlainen vastakohta URN-tunnistelle on ARK, jonka kanssa ei ole tarkoitusta käyttäen perinteisiä standarditunnuksia lainkaan, eikä

ARK-määrittys<sup>6</sup> sen vuoksi edes kerro, miten mahdollinen yhteiskäyttö tulisi toteuttaa. ARK-tunnuksilla ei siis pidä identifioida mitään sellaista, jolle voitaisiin antaa esimerkiksi ISBN tai ISSN.

URN- ja ARK-tunnuksien välimaastossa ovat Handle ja DOI. Handle-määrittys ei ota mitään kantaa perinteisiin tunnistejärjestelmiin, joten niiden soveltaminen osana Handle-tunnusta on mahdollista, mutta vapaaehtoista. Handleen perustuva DOI-standardi ISO 26324 on konservatiivisempi, sillä se edellyttää ISO:n bibliografisten standarditunnuksien käyttöä aina kun mahdollista. Kun kirjalle annetaan DOI, sen pitää perustua kirjan ISBN-tunnukseen.

DOI-tunnuksen käyttö on yleistynyt voimakkaasti 2010-luvun aikana erityisesti tieteellisten artikkelien identifiointissa, mutta DOI-tunnuksia käytetään myös monografioiden ja kausijulkaisujen identifiointissa. Ensi näkemältä soveltamistapa näyttää poikkeavan DOI-standardin linjauksista. Esimerkiksi kirjoille on annettu DOI-tunnuksia jotka eivät perustu kirjan ISBN-tunnukseen. Ja jos ISBN on osa DOI:ta, sitä ei esitetä ISO 26324 -standardissa linjatulla tavalla (josta lisää DOI-alaluvussa).

Ristiriita on kuitenkin näennäinen. DOI-tunnus voi olla julkaisun kaikkien manifestaatioiden yhteinen, ja linkittyä kirjan tai artikkelin laskeutumissivulle (landing page), jolle eri manifestaatiot on linkitetty. Manifestaatioilla saattaa olla ISBN-tunnuksia, tai vain julkaisujärjestelmän sisäinen tunnus, jos kyse on artikkeleista. Julkaisujen ohella PID-tunnuksen linkittäminen laskeutumissivulle on yleinen käytäntö myös esim. data-aineistoilla (ks. DataCiten suositus)<sup>7</sup>.

PID-tunnuksia resoluoituvat tätä kirjoitettaessa yleensä vain yhteen URL-osoitteeseen. Siksi samalla julkaisulla voi olla monta PID-tunnusta. Jos esimerkiksi väitöskirja on saatavilla kustantajan maksullisessa palvelussa, yliopiston julkaisuarkistossa ja Kansalliskirjaston vapaakappalekokoelmassa, kustantajan kappaleella voi olla DOI-tunnus, julkaisuarkistossa olevalla kappaleella Handle ja Kansalliskirjaston kappaleella URN. Ja vaikka jatkossa olisi teknisesti mahdollista resoluoida yksi PID-tunnus kaikkiin kolmeen kappaleeseen, kukin osapuoli voi haluta hallinnoida itse omaa PID-järjestelmäänsä.

Toistaiseksi PID-tunnuksia linkittyvät yleensä joko identifioitun objektin kotisivulle / landing pagelle tai

6 <https://datatracker.ietf.org/doc/draft-kunze-ark/>

7 <https://support.datacite.org/docs/landing-pages>

suoraan objektiin. Jatkossa palveluvalikoima muuttuu monipuolisemmaksi, ja tiedon tarvitsija voi valita objektin sijaan esimerkiksi sen metatiedot. PID-standardit eivät anna resoluutiopalvelujen toteuttamisesta ohjeita, vaan mahdollisten linjausten teko jää palvelujen tarjoajien vastuulle.

Mitä haittaa PID-tunnisteiden ja perinteisten tunnisteiden päällekkäisyydestä on? Alun perin huoli on poliittinen: perinteisten tunnistestandardien vähittäinen syrjäytyminen elektronisten aineistojen kuvailussa. Tästä syystä kansainvälinen ISSN-keskus vastusti aikanaan DOI-tunnisteen standardisointia ISOssa. Jos esimerkiksi verkkolehden kotisivulle annetaan DOI, lehden ISSN-tunnuksen merkitys voi vähentyä, koska verkkolehden osoitetta käyttävä linkitetty data perustuu ISSN-tunnuksen asemesta DOI-tunnukseen. Käytännössä

DOI-tunnisteesta on kuitenkin tullut ISSN-järjestelmää täydentävä ratkaisu, koska se mahdollistaa artikkelien identifiointin. Ja kansainvälinen ISSN-keskus voi puolestaan toteuttaa URN:ISSN-pohjaisia palveluita, joiden avulla kansainvälisen ISSN-tietokannan palvelut voidaan saada nykyistä helpommin käyttäjien ulottuville.

Rinnakkaisten PID-tunnisteiden käyttö voi hankaloittaa julkaisujen hakua ja identifiointia. Esimerkiksi e-kirjan haku voi epäonnistua, jos käytetään toimintansa lopettaneen kustantajan antamaa, kuollutta DOI-tunnusta. Julkaisu voisi olla edelleen löydettävissä ISBN-tunnukseen perustuvalla URNillä tai julkaisuarkiston Handle-tunnuksella. Näiden ongelmien välttämiseksi tulisi sopia siitä, miten PID-tunnisteiden toiminnallisuus taataan silloin, kun organisaatioiden toiminta syystä tai toisesta päättyy.



## 4. Uniform Resource Identifier (URI)

Useimmilla verkkojulkaisuilla ei ole perinteistä tunnusta eikä myöskään PIDIä, vaan tunnuksen kaltainen verkko-osoite (Uniform Resource Identifier, URI), josta käytetään yleisesti nimitystä Cool URI. Se määriteltiin vuonna 2005 Internet-standardissa RFC 3986<sup>8</sup>. ”Cool” viittaa siihen, ettei URI koskaan muutu. Sen pysyvyys taataan uudelleenohjaamalla (redirect) se tarvittaessa julkaisun muuttuneeseen URL-osoitteeseen. Uudelleenohjausmahdollisuuden ansiosta on mahdollista väittää, että verkossa, toisin kuin kirjajhylyssä, sijainti ja tunnus ovat yhdistettävissä.

Internetin teknisistä standardeista vastaava Internet Engineering Task Force erotti alun perin verkko-osoitteet ja tunnisteet toisistaan. Tunnistejärjestelmä oli RFC 2141<sup>9</sup> -standardissa vuonna 1997 määritelty URN (Uniform Resource Names). Functional requirements for Uniform Resource Names (RFC 1737<sup>10</sup>) määritteli URN-tunnuksille joukon vaatimuksia, joita verkko-osoitteiden olisi mahdotonta tai ainakin vaikeaa täyttää. URN-tunnuksien tuli olla muun muassa pysyviä, ainutkertaisia, skaalautuvia, teknologiariippumattomia ja niiden piti olla yhteensopivia perinteisten tunnistejärjestelmien kanssa. URL-osoitteiden (Uniform Resource Locators) ei tarvinnut olla mitään näistä.

RFC 3986 yhdistää URN-tunnuksia ja URL-osoitteita URI-sateenvarjon alle. Sen mukaan jokainen URI on tunniste, joten sähköpostiosoite ja puhelinnumero, joista voidaan muodostaa URI-tunnuksia: tel:<puhelinnumero> ja mailto:<email-osoite>, ovat toimijan tunnuksia ISNI-, ORCID- ja HETU-tunnuksen ohella. Käytännössä näin ei kuitenkaan ole, koska puhelinnumero ja sähköpostiosoite eivät ole ainutkertaisia eivätkä pysyviä. Toimijalla voi olla useita muuttuvia sähköpostiosoitteita, ja sama osoite voidaan myöhemmin antaa jollekulle toiselle.

Jotta RFC 3986 olisi vakavasti otettava tunnistestandardi, siinä olisi pitänyt rajata tunnuksiksi kelpaavien URI-tunnuksien joukkoa, mutta tätä kautta olisi ennen pitkää päädytty takaisin alkuperäiseen URN-URL-jaotteluun.

World Wide Web Consortium (W3C) ei ole varoitannut käyttämästä esimerkiksi puhelinnumeroa toimijan tunnuksena, mutta se on RFC 3986:n julkistamisen jälkeen antanut ohjeita siitä, miten tunnuksena käytet-

tävät URI:t tulisi muodostaa. Vuonna 2007 julkistetun Cool URIs for the Semantic Web -ohjeen<sup>11</sup> mukaan niiden pitäisi olla lyhyitä sekä helposti muistettavia ja hallinnoitavia. Ohje paljastaa W3C:n käsityksen URI-tunnuksien elinkaaresta:

Once you set up a URI to identify a certain resource, it should remain this way as long as possible. Think about the next ten years. Maybe twenty.

10–20 vuotta on lyhyt ikä pysyvälle tunnisteelle, mutta pitkä URL-osoitteelle. W3C-ohjeen luvussa viisi on kahdeksan Cool URI-esimerkkiä, joista seitsemän tuotti jo 2018 HTTP 404 -virheen. Niiden elinkaari on siis ollut lyhyempi kuin yllä ohjenuoraksi asetettu 10–20 vuotta. Tätä kirjoitettaessa ainoa edelleen toimiva URI antaa tulokseksi identifioitun henkilön tietojen sijaan vain hänen tuolloisen työnantajaorganisaationsa julkaisuluettelon, joten sekään ei enää toimi oikein.

Kahdeksan URIn perusteella ei voi tehdä yleisiä päätelmiä URI-tunnuksien elinajanodotteesta. Mutta Herbert Van de Sompel ja hänen tutkimusryhmänsä ovat analysoineet tieteellisten julkaisujen viitteiden avulla Cool URIn pitkäikäisyyttä, ja osoittaneet että sekä sisältönyrjähdys eli viitatun dokumentin muuttuminen<sup>12</sup> että linkkimätä eli viitattujen dokumenttien katoaminen<sup>13</sup> ovat yleisiä ongelmia. Mitä vanhempi linkki, sitä todennäköisempää on, ettei se enää toimi.

Kirjastojen kannalta RFC 3986 -standardin merkittävin periaatteellinen haaste on, että Cool URIt eivät ota perinteisiä tunnistejärjestelmiä huomioon (niiden roolia verkkoaineistojen identifioimisessa ei edes mainita RFC 3986:ssa). Verkossa julkaistun kirjan ISBN saa verkko-osoitteesta kilpailevan tunnuksen, joita on yhtä monta kuin kirjasta on verkossa kopioita. Kirjastonhoitaja, joka tallentaa e-kirjan kirjastonsa julkaisuarkistoon, tuskin ajattelee antaneensa kirjalle samalla uuden tunnuksen, jonka tulisi olla ISBN:n tavoin pysyvä ja ainutkertainen. Mutta lukija, joka viittaa kirjaan, voi käyttää viitteessä ISBN:n sijasta URL-osoitetta. Tämä hankaloittaa sen selvittämistä, mitä julkaisua tai julkaisun versiota on käytetty lähteenä varsinkin sen jälkeen, kun URL on lakannut toimimasta linkkimädän tai sisältönyrjähdysten vuoksi.

Kenties keskeisin kaikkia luotettavia tunnistejärjestelmiä yhdistävä kriteeri on hallinnointi. Jos kuka tahansa

8 <https://tools.ietf.org/html/rfc3986>

9 <https://tools.ietf.org/html/rfc2141>

10 <https://tools.ietf.org/html/rfc1737>

11 <http://www.w3.org/2001/sw/sweo/public/2007/cooluris/>

12 <https://doi.org/10.1371/journal.pone.0167475>

13 <https://doi.org/10.1371/journal.pone.0115253>

saa antaa tunnuksia mille tahansa, syntyy kaaos. Tutkijan ORCID-tunnuksen korvaaminen sähköpostiosoitteella saattaa olla toimiva ratkaisu jonkin aikaa. Mutta pysyvän ja toiminnallisen tunnuksen vaatimuksia sähköpostiosoitte ei täytä.

Tunnusten ja niiden toiminnallisuuden pysyvyys ei riipu vain tunnistejärjestelmästä ja sen teknisestä toteutuksesta. Myös tunnuksia jakavan ja/tai niiden toiminnallisuudesta vastuun kantavan organisaation tulee olla pitkäikäinen ja sillä pitäisi olla pysyvä, sopimukseen tai lakiin perustuva vastuu säilyttää identifioitua julkaisut. Tällöin pysyvien tunnistejärjestelmien toiminnallisuuden takuumiehiksi tulevat julkaisijoiden rinnalle erilaiset keskitettyjen palvelujen tarjoajat ja kansalliskirjastojen ja kansallisarkistojen kaltaiset vakaat organisaatiot, joilla on vankka kokemus myös perinteisistä tunnistejärjestelmistä. Kaupallisten toimijoiden antamien tunnisteen pysyvyys ei riipu esimerkiksi kustantajan elinkaaresta, jos kustantajalla on yhteistyökumppanina kansalliskirjaston kaltainen organisaatio. Tällä tavoin on toimittu jo pitkään esimerkiksi kirjojen osalta: suomalaiset kustantajat antavat ISBN-tunnukset julkaisuilleen itse, mutta kirjojen pysyvistä säilyttämisestä vastaavat Kansalliskirjasto ja muut vapaakappalekirjastot.

Merkittävä pitkän aikavälin ongelma Cool URI:en pysyvyyden kannalta on se, että Internet-domaineja ei voi omistaa. Niitä voi ainoastaan vuokrata muutamaksi vuodeksi kerrallaan. Jos ja kun esimerkiksi nokia.com -domainin vuokrasuhde päättyy ja hallintaoikeus siirtyy toiselle toimijalle, kaikki tämän domainin Cool URI-tunnukset joko lakkaavat toimimasta tai linkittyvät väärin dokumentteihin.

Verkkoarkistojen URI-linkit vievät aina samaan dokumenttiin, mutta vain niin kauan kuin kyseinen arkisto on tuotannossa. Lisäksi arkistosta löytyvä verkkosivu ei välttämättä vastaa alkuperäistä, koska jos sivu muodostuu useista tiedostoista, ne on voitu haravoida eli noutaa arkistoon WWW-palvelimelta tai -palvelimilta eri aikoina. Tällöin arkiston tarjoama verkkosivu voi olla yhteensopimattomista palasista koottu tilkkutäkki, kuten sääennuste, jonka sääkarta ja teksti kuvaavat eri päiviä.

Internet-nimipalvelun (Domain Name System, DNS) tarjoaman perustan epäluotettavuudesta huolimatta Cool URI -tunnuksia voidaan käyttää menestyksekkäästi rajatuissa kohteissa, kuten paikkatietojen tunnisteenä, paikkatiedon yhteiskäyttöä tukevan INSPIRE-direktiivin<sup>14</sup> nojalla ja määräyksestä. Huolellisesti suunnitellen ja tunnuksia hallinnoiden nämä URI:t saadaan elämään vuosikymmeniä. Mutta Cool URI:t ovat teknologiariippuvaisia ja toimivat vain kunnes verkon perusteknologiat kuten HTTP-protokolla muuttuvat. Tähän voi tulla mennä kymmeniä vuosia tai pidempäänkin, mutta emme voi olla varmoja säilyykö edes HTTP:n kaltainen ydinteknologia tuettuna satoja vuosia. Sen kuitenkin tiedämme, että julkaisulle annetun tunnuksen pitäisi säilyä vähintään yhtä kauan kuin julkaisun, ja mieluiten pidempäänkin, eli pysyvästi.

Joissakin järjestelmissä URI-tunnuksien sijaan on lupa käyttää Internationalized Resource Identifier – eli IRI-tunnuksia. IRI-määrittäjä RFC 3987<sup>15</sup> julkaistiin heti RFC 3986:n jälkeen, ja se laajentaa URI-tunnuksissa käytettävissä olevan suppean merkkivalikoiman Unicodeksi. Jokainen URI on siis myös IRI, mutta jos tunnuksessa käytetään RFC 3986:ssa kiellettyjä merkkejä, se on IRI. IETF yritti päivittää RFC 3987:n ajan tasalle 2010-luvun alussa, mutta yritys epäonnistui, koska IRI-tunnuksiin liittyviä teknisiä ongelmia ei pystytty ratkaisemaan. Sen vuoksi on turvallisinta käyttää vain URI-tunnuksissa hyväksytyjä merkkejä, vaikka IRI-tunnusten käyttö olisi periaatteessa mahdollista. Esimerkiksi emoji-merkkeitä sisältävä IRI olisi tunnisteenä ihmiskäyttäjille hankala, koska emoji-symbolin ulkonäkö voi vaihdella käyttöjärjestelmästä riippuen. Vastakkainen ongelma tälle on ”Unicode confusables”, eli merkit, joiden ulkoasu on sama tai lähes sama. Niiden avulla verkon käyttäjä voidaan huijata väärin verkko-osoitteisiin<sup>16</sup>.

Cool URI -tunnukset eivät täytä kirjastojen tunnisteen asettamia pysyvyyden, ainutkertaisuuden ja hallinnoitavuuden vaatimuksia. Siksi niitä käsitellään tässä ohjeessa jatkossa vain kursorisesti.

14 INSPIRE mahdollistaa paikkatietojen yhteentoimivuuden yli organisaatioiden ja valtioiden rajojen. Esimerkiksi ilmanlaatuun ja vesistöihin liittyvät ympäristövaikutukset eivät tunne rajoja, joten tietojen tulee olla yhteentoimivia, laadukkaita ja helposti saatavilla, jotta päätöksiä voidaan perustaa niiden varaan. Kyseessä on EU-direktiivi, jonka tavoitteena on Euroopan yhteisön paikkatietoinfrastruktuuri. INSPIRE tulee sanoista Infrastructure for Spatial Information in the European Community. (Katsottu 4.1.2021, <https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/paikkatietojen-yhteentoimivuu/inspire/mika-inspire>)

15 <https://tools.ietf.org/html/rfc3987>

16 <http://websec.github.io/unicode-security-guide/visual-spoofing/>

# 5. Toiminnalliset tunnisteet – yleistä

Toiminnallisilla ja perinteisillä tunnistejärjestelmillä on useita yhteisiä piirteitä, mutta myös merkittäviä eroja.

Selkein yhdistävä piirre on, että sekä perinteiset että PID-tunnukset ovat ainutkertaisia ja pysyviä. Samaa tunnusta ei saa antaa uudelleen jollekin muulle julkaisulle. Tunnuksen, olipa se perinteinen tai PID, pitää identifioida julkaisu koko sen elinkaaren ajan, ja vielä sen jälkeenkin. Myös PID-tunnukseen perustuvien hyperlinkkien pitää toimia periaatteessa pysyvästi. Jos identifioitu julkaisu poistetaan verkosta, sen PID-tunnus tulisi linkittää muistokiveen eli julkaisun metatietoihin, jotka voivat sisältää linkin korvaavaan julkaisuun.

Sekä perinteisten että PID-tunnusten ainutkertaisuus ja pysyvyys perustuvat tekniikan sijasta tunnusjakelun hallinnointiin. Tämä tarkoittaa esimerkiksi sitä, että kirja saa maailmanlaajuisesti vain yhden ISBN-tunnuksen ja vain yhden ja ainoan URN:ISBN-tunnuksen. Näiden tunnistejärjestelmien tunnusjakelu on huolellisesti organisoitu ja yksityiskohtaisten sääntöjen ohjaama.

URN-järjestelmässä ISBN-tunnuksen omaava kirja voi saada vain tuohon ISBN-tunnukseen perustuvan URN:ISBN-tunnuksen. URN-järjestelmässä kirjojen identifiointi perustuu aina jos mahdollista ISBN-tunnuksiin ja niiden jakelujärjestelmään. Mutta jos kirjalla ei ole ISBN-tunnusta, Kansalliskirjasto voi antaa sille NBN-tunnuksen ja siihen perustuvan URN:NBN-tunnuksen, josta se vastaa samalla tapaa kuin URN:ISBN-tunnuksestakin.

ISBN- ja NBN-tunnistejärjestelmät eroavat toisistaan siinä, että ISBN-tunnuksen voi antaa vain kirjalle, mutta NBN-tunnuksilla kansalliskirjastot ja niiden valtuuttamat tahot voivat identifioida myös sellaisia julkaisuja, joille ei voida antaa ISBN:ää tai muutakaan ISO:n standarditunnusta. Esimerkiksi Suomessa Kansalliskirjasto antaa digitoimilleen artikkeleille ja vanhoille kirjoille NBN-tunnukset.

Perinteiset tunnistejärjestelmät kuten ISBN ja ISSN ovat kaikki avoimia kansainvälisiä standardeja, mutta PID-järjestelmistä vain DOI- ja URN-tunnisteilla on sama status. Tosin DOI on avoin vain osittain; sen tekninen perusta on The Handle System, jonka standardisoinnista vastaava The DONA Foundation<sup>17</sup> on suljettu organisaatio.

Perinteisiä julkaisujen tunnuksia hyödynnetään käytännöllisesti katsoen kaikissa kirja-alan ja kirjastojen järjestelmissä esimerkiksi tiedonhakuun. PID-tunnukset

voidaan indeksoida perinteisten tunnisteiden tavoin, mutta PID-tunnukseen perustuva palvelutarjonta eli esimerkiksi tunnuksen linkitys identifioidun julkaisun kannalta relevantteihin URL-osoitteisiin edellyttää tunnistejärjestelmäkohtaisia resoluutiopalvelimia eli resolveita. Ainakaan toistaiseksi niitä ei ole integroitu kirja-alan tai kirjastojen järjestelmiin, vaan ne ovat erillisiä ohjelmistoja.

Nykyiset resolverit ovat yksinkertaisia sovelluksia. Esimerkiksi Kansalliskirjaston alkuperäinen URN-resolveri tarjosi vain linkityksen URN-tunnuksesta yhteen URL-osoitteeseen. Linkin pysyvyys oli sen ainoa, käyttäjälle melko näkymätön lisäarvo. Ohjelmistoja kehittämällä tilanne kuitenkin muuttuu; Kansalliskirjaston resolverisovellukseen on lisätty mahdollisuus linkittää URN-tunnus samanaikaisesti useisiin URL-osoitteisiin siten, että käyttäjä näkee vain ne osoitteet, joihin hänellä on pääsy. Ilman resolveria tällaista toiminnallisuutta ei olisi mahdollista toteuttaa.

Resolverisovelluksia kehitetään edelleen. Esimerkiksi ARK-resolveri on jo nyt tietokanta, joka sisältää kuvailuvia ja hallinnollisia metatietoja tunnuksen saaneista objekteista. Näitä tietoja voi pyytää ARK-resolverilta objektin URL-osoitteen asemesta lisäämällä ARK-tunnukseen yksi tai kaksi kysymysmerkkiä. Valitettavasti tätä syntaksia ei olisi helppoa laajentaa ARK-tunnisteen palveluvalikoiman monipuolistamiseksi. Mutta uusi URN-syntaksi, joka määriteltiin RFC-julkaisussa RFC 8141<sup>18</sup>, mahdollistaa tehokkaiden resolveiden rakentamisen. Toistaiseksi muut PID-tunnisteet eivät ole ottaneet vastaavia teknisiä ratkaisuja käyttöön. URN-syntaksiin tehdyt parannukset eivät valitettavasti ole vielä siirtyneet URN-resolveriohjelmistoihin; tiettävästi ensimmäisenä RFC 8141:ssä määritellyt ominaisuuksia on ottamassa käyttöön Kansalliskirjasto resolverinsa uudessa versiossa.

URL voidaan uudelleenohjata toiseen verkko-osoitteeseen, kunhan webmaster huolehtii tästä siirtäessään dokumenttia paikasta toiseen. Mutta jos verkon domainin (esim. helsinki.fi) omistaja muuttuu, URL-osoitteen uudelleenohjaus ei yleensä ole enää mahdollista. Resoluutiopalvelimien ansiosta PID-tunnisteet ovat immuuneja domainien omistuksen ja verkkotekniikan muutoksille. Verkon näkökulmasta resolverit asettuvat kulloisenkin yhteysprotokollan kuten HTTP:n ja HTTPS:n yläpuolelle. PID-tunnuksella identifioitu

17 <https://www.dona.net/>

18 <https://tools.ietf.org/html/rfc8141>

dokumentti voidaan aina siirtää verkossa osoitteesta toiseen, eivätkä verkon infrastruktuurin merkittävään muutokseen vaikuta linkitykseen.

Jokainen julkaisu saa yleensä vain yhden perinteisen tunnuksen. PIDit ovat toista maata: samalla julkaisulla voi olla useita PID-tunnuksia. Jos julkaisusta on useita kopioita eri toimijoiden hallinnoimissa tietojärjestelmissä, niillä voi olla omat PID-tunnukset, joilla on kullakin on oma vastuunorganisointinsa. Jokaisen tunnuksen pitäminen pysyvästi toiminnallisena on helpompaa, jos tunnuksien antajatahot tekevät yhteistyötä esimerkiksi vaihtamalla resolverien sisältämiä metatietoja.

Toistaiseksi kaikki PID-tunnukset esitetään HTTP URI -muodossa lisäämällä tunnisteseen resolverin osoite. Esimerkiksi näin:

<http://urn.fi/URN:NBN:fi-fe201102171251>

Resolverin osoite, edellä <http://urn.fi>, ei ole osa URN-tunnusta. Tulevaisuudessa kaikkia PID-tunnuksia

pitää voida käyttää verkossa hyperlinkkeinä sellaisenaan, koska resolverin osoitteen lisääminen tekee PID-tunnuksesta teknologiariippuvaisen. Riittävän kaukaisessa tulevaisuudessa HTTP-protokolla korvautuu jollakin toisella teknologialla, ja jo kauan ennen tätä resolverin osoite voi muuttua. PID-tunnuksien käyttö linkkinä on mahdollista, jos Internetin Domain name service – eli DNS-nimipalvelujärjestelmä<sup>19</sup> tai sen tulevaisuudessa korvaava järjestelmä ”tietää” kaikkien resolverien osoitteet.

Seuraavissa luvuissa tarkastellaan PID-tunnistejärjestelmiä mm. seuraavista käytön kannalta merkittävistä näkökulmista:

- Järjestelmän tausta ja tukiorganisaatio
- Soveltamisalat ja keskeiset käyttäjät
- Tunnuksen rakenne
- PID julkaisujen tunnisteenä
- Tulevaisuuden näkymät

---

<sup>19</sup> <https://fi.wikipedia.org/wiki/DNS>

# 6. ARK

## 6.1 ARK-järjestelmän tausta ja tukiorganisaatio

Archival Resource Key eli ARK-tunnistejärjestelmän ja sen resolveriohjelmiston kehitti John Kunze The California Digital Librarystä (CDL), joka edelleen vastaa järjestelmän kehittämisestä. ARK julkistettiin hieman muiden PID-järjestelmien jälkeen vuonna 2001, ja siinä on pyritty korjaamaan muiden järjestelmien puutteita. ARK-tunnisteissa ei esimerkiksi saa käyttää muita kuin tulostettavia ASCII-merkkejä. Handle-tunnisteen merkkivalikoima on Unicode, mikä heikentää tunnusten ihmislueuttavuutta. ARK-tunnuksissa tätä ongelmaa ei ole.

ARK-järjestelmän kehittämiseen osallistuu tätä nykyä CDL:n ohella useita muitakin tunnisteen käyttäjätahoja.

## 6.2 ARK-järjestelmän soveltamisalat ja keskeiset käyttäjät

ARK-tunnisteella voidaan identifioida periaatteessa mitä tahansa. John Kunzen ja Emmanuelle Bermésin laatima ARK-määrittäminen<sup>20</sup> kuvaa asian seuraavasti:

ARK is well suited to long-term access and identification of any information resources that accommodate reasonably regular electronic description. This includes digital documents, databases, software, and websites, as well as physical objects (books, bones, statues, etc.) and intangible objects (chemicals, diseases, vocabulary terms, performances).

ARK-tunnisteella on satoja käyttäjäorganisaatioita. Merkittävimmät ARK-tunnuksia hyödyntävät organisaatiot lienevät Louvren taidemuseo, Ranskan kansallisarkisto sekä Ranskan ja Ison-Britannian kansalliskirjastot. Ranskassa ARK onkin erittäin suosittu keskeisten muistiorganisaatioiden tarjoaman vetoavun ansiosta. Siellä hienoin esimerkki on Louvre, jonka digitoiduille kokoelmille on annettu 483 000 ARK-tunnusta<sup>21</sup>. Suomessa käyttäjiä ei tiettävästi ole, minkä takia tunnistetta käsitellään tässä tekstikokonaisuudessa vain lyhyesti.

## 6.3 ARK-tunnuksen rakenne

ARK koostuu seuraavista osista, jotka erotetaan kauttaviivalla (/):

<sup>20</sup> <https://datatracker.ietf.org/doc/draft-kunze-ark/>

<sup>21</sup> <https://arks.org/blog/the-louvre-collection-goes-online-with-483000-arks/>

- Name Mapping Authority (NMA), ARK-resolverin osoite
- ARK Label ”ARK:”, ARK-tunnuksen nimiö
- Name Assigning Authority Number (NAAN), ARK-tunnuksen rekisteröijän tunnus
- Name, identifioidun objektin tunnus
- Qualifier, objektin nimen tarkenne

Tarkenne erotetaan nimestä joko kauttaviivalla, jos kyse on identifioidun objektin osakohteesta, tai pisteellä, kun identifioidaan alkuperäisen objektin versio. Esimerkiksi Baudelairen Pahan kukkia –teoksen ensipainoksen digitoidun version ARK on <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m>.

Runolla Au Lecteur (Lukijalle, kuva 1) on ARK <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19>, ja tämän runon alla olevan sivukuvan ARK on <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19.highres>.

Muista PID-järjestelmistä poiketen resolverin osoite on vapaavalintainen osa tunnusta. Tarvittaessa tunnukseen voidaan myös lisätä resolverille suunnattuja pyyntöjä. Vanhoissa ARK-luonnoksissa ”?” vaati resolveria toimittamaan identifioidun objektin suppeat metatiedot; ”??” kattavat metatiedot. Nykyinen ARK-versio on korvannut nämä edelleen tuetut kyselyt URI-syntaksin mukaisella kyselyllä ”?info”. Esim. URI <http://n2t.net/ark:67531/metadc107835?info> tuottaa identifioidun objektin resolverille tallennetut metatiedot.

## 6.4 ARK julkaisujen tunnisteena

ARK-määrittäminen ei kerro, miten ARK ja perinteiset tunnistejärjestelmät kuten ISBN sovitetaan yhteen. ARK-yhteisö on ratkaissut ongelman niin, että jos julkaisulla on jokin standarditunnus, sille ei anneta ARK-tunnusta. Tämä linja näyttää pitäneen – ARK-tunnuksia kokeillesani en ole löytänyt yhtään kirjaa, jolla olisi sekä ISBN että ARK.

ARK-tunnisteen soveltumattomuus julkaisuille joilla on perinteinen tunnus rajoittaa ARK:in käytettävyyttä. Sen rinnalla pitäisi käyttää jotakin muitakin PID-tunnistetta, jos halutaan identifioida esimerkiksi ISBN-tunnuksen omaavia e-kirjoja.

ARK-tunnusten käyttöä ei valvota keskitetysti. Järjestelmä ei vaadi tunnuksen antajaa korjaamaan tilannetta, jos hänen hallinnoimansa ARK lakkaa resoluvoitumasta. Tässä suhteessa ARK muistuttaa Handlea ja sellaisia



## AU LECTEUR

---

La sottise , l'erreur, le péché , la lésine  
Occupent nos esprits et travaillent nos corps ,  
Et nous alimentons nos aimables remords ,  
Comme les mendiants nourrissent leur vermine.

Nos péchés sont têtus , nos repentirs sont lâches ;  
Nous nous faisons payer grassement nos aveux ,  
Et nous rentrons gaîment dans le chemin bourbeux ,  
Croyant par de vils pleurs laver toutes nos taches.

Sur l'oreiller du mal c'est Satan Trismégiste  
Qui berce longuement notre esprit enchanté ,  
Et le riche métal de notre volonté  
Est tout vaporisé par ce savant chimiste.

Kuva 1. Runolla Au Lecteur (Lukijalle) on ARK <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19>, ja runon tämän oppaan kuvitukseksi valitun sivukuvan ARK on <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19>. highres. Lähde: gallica.bnf.fr / Médiathèque de la communauté urbaine d'Alençon.

URN-nimialueita, joissa tunnusten antamista ei valvota. Hallinnoinnin puute on ongelma siksi, että keskitetysti valvottu PID-tunnus on lähes varmasti pitkäikäisempi kuin valvomaton. Kun tunnuksen antaja saa reklamaation resoluution tuottamasta 404-virheestä, hän voi korjata ongelman. Hyvin hallinnoidussa järjestelmässä, josta DOI on paras esimerkki, hänen pitää korjata tilanne. ARK-järjestelmässä ongelmaa ei välttämättä edes huomata, elleivät käyttäjät valita siitä. Niinpä ARK-määrittelyynkin on livahtanut ARK-tunnus <https://444.berkeley.edu/ark:28722/x9t38rk45c> joka ei ainakaan tätä kirjoitettaessa toimi.

## 6.5 ARK-järjestelmän tulevaisuuden näkymät

ARK on teknisesti vakaa järjestelmä, mutta sen standardisointi on yhä kesken. Prosessi alkoi Internet Engineering Task Forcessa (IETF) jo vuonna 2001, ja helmikuuhun 2021 mennessä ARK-määrittelyksen Internet Draft -luonnoksia oli ilmestynyt jo 28<sup>22</sup>. Työ jatkuu yhä, mutta on mahdollista, ettei sitkeys tuota toivottua tulosta. IETF on jo ehtinyt hyväksyä URN-tunnisteen Internet-standardiksi, eikä se yleensä siunaa päällekkäisiä ratkaisuja samaan tarkoitukseen. Lisäksi ARK on niin sanottu *private contribution*, eli sitä ei ole laatinut

---

22 <https://datatracker.ietf.org/doc/html/draft-kunze-ark>

IETF:n asettama virallinen työryhmä. Kuka tahansa voi laatia Internet-standardiluonnoksen eli Internet draftin, mutta yksityishenkilöiden aloitteita ei läheskään aina edes käsitellä IETF:n johtelimityksessä, ja vaikka käsiteltäisiinkin, niistä ei yleensä tule varsinaisia standardeja. Parhaassa tapauksessa ARK julkaistaan Handlen tapaan Informational-statusella, jolloin siitä tulee RFC-julkaisu, mutta ei standardia.

ARK-järjestelmällä on tätä kirjoitettaessa noin 750 käyttäjäorganisaatiota, joilla on yhteistyötä edistävä allianssi<sup>23</sup>. Koska ARK-tunnus on maksuton, niitä on luotu yli 8 miljardia, eli paljon enemmän kuin esim. DOI-tunnuksia. ARK-sovelluksen ja laitteistoympäristön kehittäminen ja ylläpito lienevät vakaalla pohjalla. Varmuutta tästä on vaikeaa saada, koska esimerkiksi rahoitusjärjestelyt eivät ole julkisia.

ARK-tunnisteen ja perinteisten tunnistejärjestelmien yhteiskäyttö ei toistaiseksi ole mahdollista. Tämä rajoitus on mahdollista poistaa helposti, määrittelemällä miten perinteisiä tunnuksia voidaan käyttää ARK-tunnuksen osana. Toistaiseksi ARK-määrittelyä ei kuitenkaan ole laajennettu tällä tavoin.

---

23 <https://arks.org/>

# 7. URN

## 7.1 URN-järjestelmän tausta ja tukiorganisaatio

Uniform Resource Name -tunnistejärjestelmästä eli URN:stä vastaa Internet-standardeihin keskittyvä Internet Engineering Task Force. URN-tunnukset ovat kaikkea sitä, mitä Internetin URL-osoitteet eivät ole: pysyviä, ainutkertaisia ja yhteismitallisia perinteisten tunnistejärjestelmien kuten ISBN:n kanssa. Sitä, miksi nykyään puhutaan URL-osoitteiden ja URN-tunnusten sijaan Uniform Resource Identifier -tunnisteista (URI), käsitellään tarkemmin tuonnempana.

URN-tunnuksen standardisoinnilla on pitkä ja monipolvinen historia. URN-syntaksi (RFC 2141) julkaistiin toukokuussa 1997 vajaan vuoden kestäneen työryhmyöskentelyn jälkeen, mutta URN-tunnuksen vaatimukset oli kuvattu jo 1994, RFC 1737:ssä (Functional requirements for Uniform Resource Names, <https://tools.ietf.org/html/rfc1737>).

URL-osoitteiden ja URN-tunnusten rinnalle syntyi URI vuonna 2005, kun RFC 3986 (Uniform Resource Identifier (URI): Generic Syntax, <https://tools.ietf.org/html/rfc3986>) kuvasi URI-tunnusten rakenteen ja suhteen URL-osoitteisiin ja URN-tunnuksiin. URN-tunnistejärjestelmän kannalta merkittävää oli, että URI määritteli URI Fragment ja URI Query –ominaisuudet. Fragmentin avulla asiakasohjelma (esim. Web-selain) voi siirtyä noudetussa dokumentissa haluttuun kohtaan, ja Queryllä voidaan lähettää palvelimelle (esim. HTTP-serverille) komentoja ja niihin liittyviä parametreja.

URN-syntaksin uudistaminen käynnistyi 2010-luvun alussa. Monivaiheisen työn lopputulemana syntyi RFC 8141 (URN Syntax, <https://tools.ietf.org/html/rfc8141>), joka julkaistiin vuonna 2017.

URN-syntaksin uudistaminen oli vaativa prosessi. URI-syntaksin lisäksi oli varmistettava URN-tunnuksen yhteensopivuus myös muiden relevanttien Internet-standardien kuten UNICODEn kanssa. Työ jäi harmittavasti vähän kesken: R-komponentin eli URN-tunnukseen lisättävän komennon (ks. luku 6.3) sisäistä rakennetta ei määritelty, vaan se jätettiin standardin implementoijien vastuulle. Kansalliskirjasto on tätä kirjoitettaessa lisäämässä R-komponentin tukea URN-resolveriin, joten osana kehitystyötä joudumme myös tekemään täydentävää standardisointia.

## 7.2 URN:n soveltamisalat ja keskeiset käyttäjät

URN on globaali tunnistejärjestelmä, jonka avulla voi identifoida periaatteessa mitä tahansa. Tämän mahdollistaa URN-järjestelmän jakautuminen tunnistejärjestelmäkohtaisiin nimialueisiin. URN voi toimia sateenvarjona, joka sisällyttää itseensä muut tunnistejärjestelmät ja soveltaa niiden toimintaperiaatteita. Missään muussa tunnistejärjestelmässä ei ole vastaavaa ominaisuutta.

URN tunnetaan kirjastojen ja erityisesti kansalliskirjastojen tunnisteena. Sitä sovelletaan esimerkiksi Suomessa, Ruotsissa, Norjassa, Saksassa ja Italiassa. Kansalliskirjastoilla on oma URN-nimialue, URN:NBN (national bibliography number), joka mahdollistaa kirjastojen käyttämien paikallisten tunnistejärjestelmien käytön URN-tunnuksina. URN-järjestelmään on rekisteröity nimialueet myös esimerkiksi ISBN- ja ISSN-tunnistejärjestelmille. Niiden nimialuerekisteröinneissä korostetaan sitä, että URN:ISBN- ja URN:ISSN-tunnuksia tulee soveltaa ISBN- ja ISSN-tunnuksien tapaan.

URN-tunnuksella ei kuitenkaan ole monopoliasemaa edes kansalliskirjastoissa. Ne käyttävät myös muita tässäkin esityksessä kuvattuja PID-tunnistejärjestelmiä. Muutamat kansalliskirjastot luottavat jopa Cool URI –tunnuksiin eli käytännössä verkko-osoitteisiin. Niitä sovellettaessa haasteeksi tulee esimerkiksi suhde perinteisiin tunnistejärjestelmiin sekä se, että Cool URI on FRBR-mallin näkökulmasta aina niteen tai tässä tapauksessa siis e-aineiston yksittäisen kopion tunnus. Sen soveltaminen teoksen tai manifestaation tunnuksena on vähintään hankalaa ellei mahdotonta.

URN-standardit kehittänyt IETF ei valvo eikä aktiivisesti edistä URN-tunnistejärjestelmän käyttöä. URN-markkinointi ei olisi edes mahdollista URN-tunnistejärjestelmän hajautetun luonteen vuoksi. URN-nimialueita ovat rekisteröineet hyvin erilaiset toimijat; yhteistä nimittäjää kuten yhteistä resolverisovellusta tai URN-sovellusohjetta ei ole. Tämä hajautettu organisoitumisen malli on yksi piirre, missä URN eroaa DOI:sta, jota hallinnoi ja valvoo International DOI Foundation.

Monet URN-nimialueet eivät tarjoa resoluutiopalveluja, eikä RFC 8141 tätä edellytäkään. URN-tunnusta voi käyttää perinteisen tunnisteiden tapaan. Esimerkkejä ei-toiminnallisista URN-nimialueista ovat URN:MRN ja URN:GDST. International Association of Marine Aids to Navigation and Lighthouse Authorities'n (IALA) rekisteröimä nimialue URN:MRN mahdollistaa

erilaisten merenkulkuun liittyvien fyysisten tai elektronisten objektien identifiointiin. URN:UVCI on EU:n koronarokotustodistuksen tunnisteiden nimialue. Sen rekisteröintiprosessi on tätä kirjoitettaessa vielä kesken. Global Dialogue on Seafood Traceability (GDST) on käynnistänyt kalastukseen liittyvän URN:GDST-nimialueen rekisteröinnin. Tavoitteena on minimoida laitton kalastus mahdollistamalla kuluttajille myyntiin tulevien kalaerien seuranta koko tuotantoketjun ajan.

Kulttuurialalla URN-tunnisteen aktiivisimpia käyttäjiä ovat kirjastot ja elokuvateollisuus. Niiden ”omat” nimialueet URN:NBN ja URN:EIDR ovat toiminnallisia ja hyvin hallinnoituja. URN-järjestelmässä valvonta on ylipäättään mahdollista korkeintaan nimialuekohtaisesti. Jotkut nimialueet, kuten URN:NBN, voidaan jakaa edelleen pienempiin osiin, joita kutakin hallinnoidaan erikseen. Esimerkiksi Kansalliskirjaston vastuulla on URN:NBN:FI- eli suomalaiset NBN-tunnukset. On myös URN-nimialueita, joilla ei lähtökohtaisesti ole mitään kontrollia. Esimerkiksi URN:UUID<sup>24</sup>-tunnuksia voi generoida kuka tahansa, koska tahansa ja mille tahansa elektroniselle objektille. Tämä täydellinen vapaus johtuu siitä, ettei UUID-tunnuksien luontiakaan rajoiteta millään tavalla.

Itsenäisiin tunnistekohtaisiin nimialueisiin pohjautuvan sateenvarjomallin ansiosta URN on sovellusalaan laajin tunnistejärjestelmä. Se voi sisällyttää itseensä paitsi kaikki perinteiset tunnistejärjestelmät, myös muut PID-tunnisteen, kuten Handlen ja DOI:n, ja sen avulla voidaan identifioida periaatteessa mitä tahansa.

## 7.3 URN-tunnuksen rakenne

URN-tunnuksessa on oltava kolme osaa, jotka erotetaan toisistaan kaksoispisteellä:

- Label ”URN:”, URN-tunnuksen nimiö. Jokainen URN-tunnus alkaa näin; tarkoituksena on helpottaa URN-tunnuksen löytämistä tekstin keskeltä, ARK-tunnuksen ARK:-nimiön tapaan.
- Namespace Identifier (NID), nimialueen tunnus. Esimerkiksi ”ISBN”.
- Namespace Specific String (NSS), resurssin tunnus. Esimerkiksi ISBN-numero

ISBN-tunnuksesta 951-8959-47-1 saadaan näin URN URN:ISBN:951-8959-47-1. URN-tunnukset on yleensä helppoa luoda ohjelmallisesti olemassa olevasta perinteisestä tunnuksesta. Mutta jos URN ei ole toiminnallinen, sen tarjoama lisäarvo voi olla vähäinen.

URN-tunnukseen voidaan lisätä vapaavalintaisia komponentteja. Ne ovat URI-syntaksin Fragment-ominaisuutta vastaava F-komponentti, URI Query -ominaisuutta vastaava Q-komponentti sekä R-komponentti,

<sup>24</sup> ks. [https://en.wikipedia.org/wiki/Universally\\_unique\\_identifier](https://en.wikipedia.org/wiki/Universally_unique_identifier)

josta lisää tuonnempana tässä luvussa.

Nimialueiden tunnukset ovat ainutkertaisia ja ne identifioivat pysyvästi URN-tunnuksessa käytetyn tunnistejärjestelmän. Kullakin nimialueella noudatetaan kyseisen tunnistejärjestelmän pelisääntöjä, jotka koskevat muun muassa tunnusten rakennetta ja identifiointin kohteita. Esimerkiksi URN:ISBN-nimialueella toimitaan ISBN-standardin ja käyttöoppaan mukaisesti. Heinäkuuhun 2021 mennessä nimialueita on rekisteröity yli 50<sup>25</sup>, ja useita hakemuksia on käsiteltävänä. Vain pieni osa näistä nimialueista kuuluu perinteisille bibliografisille tunnisteeille tai ylipäättään julkaisujen tunnisteeille.

URN-tunnuksen NSS-osioon tallennetaan varsinainen tunnus. Niiden tarvitsee olla ainutkertaisia vain oman nimialueensa sisällä. Jos URN-tunnuksen saavalla dokumentilla on jo olemassa standarditunniste, tulee käyttää sitä. Toisin sanoen: jos kirjalla on ISBN, URN-tunnuksen NSS-osion on oltava tuo ISBN-tunnus sellaisenaan, siihen ei saa lisätä mitään eikä siitä saa ottaa pois mitään. Jotkin nimialueet sallivat myös paikallisten tunnisteen käytön. Esimerkiksi kansalliskirjastojen antamista paikallisista tunnisteeista voidaan tehdä globaaleja ja ainutkertaisia URN:NBN-tunnisteita. Tämä edellyttää maakoodin ja mahdollisesti myös kirjastotunnuksen lisäämistä paikalliseen tunnukseen.

Esimerkiksi URN:NBN-tunnuksessa <http://urn.fi/URN:NBN:fi-fd2010-00003245> on Kansalliskirjaston sisäinen digitoidun aineiston tunnus fd2010-00003245, johon on lisätty Suomen maakoodi fi.

Tunnuksissa sallitut merkit määrittävät nimialuekohtaisesti. Esimerkiksi ISBN-nimialueella sallittuja ovat vain ISBN-standardissa määritellyt merkit. Jos URN-tunnus esitetään hyperlinkkinä, merkkejä pitää soveltaa URI-syntaksin (RFC 3986) edellyttämällä tavalla. URN-tunnuksissa ei esimerkiksi saa käyttää merkkejä ”?” ja ”#” sellaisenaan, koska niillä on URI-syntaksissa erityisrooli. Merkkiä ”?” seuraa aina query ja merkkiä ”#” fragment. URI-syntaksissa määritellyn erityisroolin omaavat merkit sekä periaatteessa kaikki ASCII-merkkivalikoiman ulkopuoliset merkit on esitettävä URI- ja URN-tunnuksissa koodattuna<sup>26</sup>. Tällöin esimerkiksi kysymysmerkestä tulee %3F.

Toisin kuin URI-syntaksin fragmentilla ja queryllä, URN-järjestelmän F- ja Q-komponenteilla ei ole roolia identifiointissa (mistä syystä niille piti antaa eri nimet). Käytännössä tämä merkitsee sitä, että kirjan lukuun viittaavan F-koodin lisääminen URN:ISBN-tunnukseen

<sup>25</sup> <https://www.iana.org/assignments/urn-namespaces/urn-namespaces.xhtml> Koska URN on Internet-standardi, URN-nimiavaruuksien rekisteröinnin vastuuorganisaatio on IANA (Internet Assigned Names Authority). Se vastaa myös rekisteröityjen nimialueiden luettelosta.

<sup>26</sup> <https://en.wikipedia.org/wiki/Percent-encoding>

ei muuta URN-tunnusta kyseisen luvun tunnukseksi. Siksi URN johon on lisätty f-komponentti, kuten <http://urn.fi/URN:NBN:fi-fd2010-00003245#Luku2> identifioi saman julkaisun kuin edellä oleva URN, mutta F-koodin avulla selain siirtää käyttäjän haluttuun kohtaan identifioidussa kirjassa sen jälkeen kun selain on noutanut julkaisun.

URI Queryn avulla voidaan lähettää komentoja suoraan kohdejärjestelmään kuten vaikkapa Melinda-tietokantaan. Tunnus on tällöin muotoa

<http://urn.fi/URN:NBN:fi-fd2010-00003245?=<Q-komponentti>>

Q-komponentti voi olla esimerkiksi tunnuksen perustuva SRU-haku, jolloin resolverin pitää tietää, minkä tietokannan SRU-palvelimelle haun voi lähettää. Jos tunnus on esimerkiksi suomalainen ISBN-tunnus, Melindan SRU-palvelin on sopiva kohde.

URI-syntaksi ei ota huomioon sitä, että asiakasohjelman ja kohdejärjestelmän välissä voi olla PID-tunnisteen resolverisovellus. Koska URI-syntaksiin ei sisälly mahdollisuutta lähettää komentoja resolverille, URN-tunnukseen lisättiin R-komponentti. Sen avulla voidaan resolverilta pyytää periaatteessa mitä vain, kuten lista identifioidun julkaisun nykyisistä tai vanhoista URL-osoitteista.

Esimerkkinä käytetty URN näyttää R-komponentin kera tältä:

<http://urn.fi/URN:NBN:fi-fd2010-00003245?+<R-komponentti>>

R-komponentin toteutus URN-syntaksissa on yhteismitallinen RFC 3986:n linjausten kanssa. Muilla PID-järjestelmillä on omat R-komponenttia vastaavat ratkaisunsa, mutta ne voivat olla yhteensopimattomia URI-syntaksin kanssa. Kun Handle- ja DOI-tunnistejärjestelmiin lisättiin R-komponenttia vastaava ominaisuus (10320/loc Handle type), tunnuksen ja lisäosan erotinmerkiksi valittiin @, koska se oli vapaa – sitä ei ollut hyödynnetty Handle-tunnuksissa. URI-syntaksin mukainen erotinmerkki ? oli jo käytössä.

PID-järjestelmiin ei olisi tarvinnut rakentaa näitä erillisratkaisuja, jos URI-syntaksi olisi ottanut resolverisovellusten tarpeet huomioon. Ne ohitettiin RFC 3986:tta kirjoitettaessa joko vahingossa tai tarkoituksella.

## 7.4 URN-tunnistejärjestelmän palvelut ja niiden toteuttaminen

URN:n ja muiden PID-järjestelmien edellyttämiä resolveita voidaan pitää ylimääräisenä askelena tunnuksen ja identifioidun resurssin välissä. Mutta resolverit voidaan nähdä myös keinona irrottaa resurssien tunnisteet ja sijaintipaikka toisistaan, samalla tavalla kuin kirjastot ovat perinteisesti erottaneet tunnisteet ja paikanmerkit

(signumit). Verkossa kuten myös kirjastossa on toki mahdollista pitää aineisto samassa paikassa, ja luostari-kirjastoissa kirjat olivat tätä varten ketteillä hyllyissä kiinni. Internet-ajan kettinki on URI, koska se edellyttää resurssille annetun osoitteen pysymistä toiminnallisena.

Tässä esityksessä käsitellyistä tunnistejärjestelmistä ARK, Handle ja DOI käyttävät omaa, keskitettyä ylläpidettyä avoimen lähdekoodin resolveriohjelmistoa. URN-järjestelmässä yhteistä resolverisovellusta ei ole, mikä on haitannut tunnisteen käytön yleistymistä, sillä käyttäjien on pitänyt laatia resolverisovelluksensa itse. Eri URN-käyttäjryhmien ja URN-nimialueiden tarpeiden erilaisuuden vuoksi ei ole varmaa, voisiko yhteinen ohjelmisto edes kelvata kaikille.

Kun Kansalliskirjaston resolveriohjelmiston uudistamishanke on päätynyt, ohjelmisto annetaan muiden kirjaston kehittämien sovellusten tapaan vapaasti käyttöön. Ellei joku muu toimija ehätä ensin, resolverista tulee ensimmäinen avoimen lähdekoodin URN-resolveri.

URN-järjestelmän tarjoamat palvelut ovat olleet ”kortilla”. Mutta Q- ja R-komponenttien avulla URN-tunnus voidaan linkittää paitsi identifioituun resurssiin itseensä, myös mihin tahansa kohteeseen joka jollakin tavoin liittyy kyseiseen resurssiin. Tämä luo mielenkiintoisia mahdollisuuksia linkitetyn datan kehittäjille. URL-osoitteiden korvaaminen tietueissa URN-tunnuksilla tulee myös helpottamaan linkkien ylläpitoa, koska resolverilla linkkien ylläpito on keskitettyä ja se voidaan myös automatisoida. Tämä on mahdollista, jos identifioitujen dokumenttien URL-osoitetiedot voidaan haravoida resolverille julkaisuarkistoista ja muista tallennusjärjestelmistä.

Q-komponentin tukeminen on teknisesti yksinkertaista. Asiakasohjelmassa suomalaisen kirjan URN:ISBN-tunnukseen voidaan lisätä Q-komponenttina SRU-standardiin perustuva hakulause. Resolverin tarvitsee vain lähettää Q-komponentti sellaisenaan Melindan SRU-palvelimelle, joka toimittaa halutun kirjan metatiedot asiakkaalle.

R-komponentin tukeminen on resolverille hieman hankalampaa. Edellisen esimerkin URN:ISBN-tunnukseen voidaan lisätä pyyntö noutaa identifioidun resurssin metatiedot MARCXML-formaatissa. URN-resolverin pitää muuntaa tämä R-komponentti Melindan SRU-palvelimelle lähetettäväksi hauksi.

R-komponentin tarjoama toiminnallisuus vastaa Q-komponenttia. Niiden ero on se, että jos kohdejärjestelmän tukema rajapinta muuttuu esimerkiksi SRU:sta joksikin muuksi, R-komponenttia käytettäessä vain resolverilla tehtävää muunnosta R-komponentista hakulauseeksi pitää päivittää. Mutta jos käytetään Q-komponenttia, jonka resolveri aina toimittaa sellaisenaan suoraan kohdejärjestelmään, hakulauseeseen päivitys



pitää tehdä jokaisessa resolveria käyttävässä asiakasohjelmassa erikseen.

URN-resolverilla voidaan tukea myös ominaisuuksia, joita HTTP tai muut laajassa käytössä olevat Internet-protokollat eivät tue. Resolveri voi tallentaa saman resurssin kaikki URL-osoitteet, ja linkittää asiakkaan sellaiseen kappaleeseen, johon hänellä on käyttöoikeus. Resolverille voidaan tallentaa myös resurssin vanhat URL-osoitteet, joilla on merkitystä silloin kun aineisto on saatavissa enää verkkoarkistoista. Ja kun resurssi häviää verkosta, resolveri voi linkittää URN-tunnuksen resurssin pitkäaikaissäilytysjärjestelmään tallennettuun kopioon.

## 7.5 URN julkaisujen tunnistena

URN perustuu olemassa oleviin tunnistejärjestelmiin, ja tämä erottaa sen muista PID-tunnistejärjestelmistä. Jos kirjalla on ISBN ja kirjalle annetaan URN, sen on oltava URN:ISBN. URN-tunnuksen osana olevaa ISBN:ää ei tarvitse muokata eikä päivittää. Toisin kuin ISBN-A -tunnuksessa, myös vanha 10 merkin mittainen ISBN-tunnus kelpaa.

Kirjastoalalla käytettyjen tunnistejärjestelmien soveltuvuutta URN-tunnuksiksi arvioitiin jo yli 20 vuotta sitten (Using Existing Bibliographic Identifiers as Uniform Resource Names, RFC 2288<sup>27</sup>). Mitään erityisiä ongelmia ei tässä selvityksessä havaittu. Siirtyminen teoriasta käytäntöön ISBN- ja ISSN-tunnisteilla paljasti kuitenkin mielenkiintoisia eroja eri tunnistejärjestelmien välillä.

Semanttinen PID-tunnus kertoo, mistä sen resolveri löytyy. Ei-semanttinen ("tyhmä") PID-tunnus voidaan resolvoida Internet-verkossa vain yhdessä paikassa, koska tunnus ei anna mitään vinkkiä siitä, mistä resolveri löytyy. URN:ISSN-tunnuksille tämä ei ole ongelma, koska kansainvälinen ISSN-keskus voi tarjota keskitetyn palvelun ISSN-tietokannastaan. ISBN-tunnisteella mitään vastaavaa tietokantaa ei ole, mutta koska tunnus on semanttinen, oikea resolveri on mahdollista löytää. Esimerkiksi 978-951- ja 978-952 -alkuiset URN:ISBN-tunnukset voidaan ohjata Kansalliskirjaston resolverille, ja 978-3 -alkuiset Saksan kansalliskirjaston palveluun.

Edellä esitetty raja koskee vain tulevaisuutta, jolloin URN-tunnukset voidaan esittää hyperlinkkeinä ilman resolverin osoitetta. Vain tällöin tunnukset ovat täysin teknologiariippumattomia, mikä takaa niiden toiminnallisuuden hyvin pitkälle tulevaisuuteen. Tämän tavoitteen vuoksi tulisi pidättäytyä ratkaisuihin, jotka eivät salli resolverin osoitteen poistamista. Esimerkiksi

<sup>27</sup> <http://www.ietf.org/rfc/rfc2288.txt>

URN:ISSN-tunnukset olisi mahdollista käsitellä Kansalliskirjaston resolverilla, jos tunnukset tallennettaisiin muodossa <http://urn.fi/urn:issn:<issn>>. Mutta kun resolverin osoite poistetaan, ja URN:ISSN-tunnuksen resolverin osoite noudetaan Internetin nimipalvelusta tai vastaavasta tulevaisuuden järjestelmästä, resolveriksi tulee kansainvälisen ISSN-keskuksen ylläpitämä järjestelmä.

Resolverin osoitteen poiston jälkeen URN-järjestelmä ei pysty käsittelemään ei-semanttisia tunnistejärjestelmiä, joilla olisi monta resoluutiopalvelua. Jos esimerkiksi ISBN-tunnuksesta tehtäisiin ISSN:n kaltainen "tyhmä" numerosarja, URN:ISBN-tunnustien toiminnallisuutta ei enää voisi taata. Tästä syystä URN-tunnisteen käyttäjien on tarpeen seurata tarkoin esimerkiksi ISBN-standardin kehitystä ja varmistaa URN:ISBN-tunnusten toiminnallisuus myös jatkossa. Handle- ja DOI-järjestelmille rakenteeton ISBN ei olisi ongelma, koska näiden tunnuksien etuliite osoittaa resolverin sijainnin. Esimerkiksi ISSN-tunnukseen perustuva kausijulkaisun DOI resolvoidaan kustantajan omalla resolverilla, jos kustantaja käyttää DOI-tunnuksessa omaa etuliitettään.

URN-nimialueet rekisteröitiin eniten käytetyille perinteisille tunnisteille pian URN-syntaksin määrittelyn jälkeen: ISSN-rekisteröinti vuonna 2000 ja ISBN sekä NBN vuonna 2001. Kaikki edellä mainitut nimialueet on sittemmin päivitetty uuden URN-syntaksin mukaisiksi. NBN-nimialueen rekisteröinti Using National Bibliography Numbers as Uniform Resource Names (RFC 8458, <https://tools.ietf.org/html/rfc8458>) on samalla myös NBN-tunnistejärjestelmän virallinen kuvaus.

URN:ISBN- ja URN:NBN-tunnuksia on käytetty nimialueiden rekisteröinnistä lähtien. ISSN-nimialueen tuotantokäyttö on alkamassa vuonna 2021. Kansainvälinen ISSN-keskus käyttää resolverina omaa versiotaan Kansalliskirjaston ohjelmistosta. Tavoitteena on tarjota ISSN-portaalin palveluja kuten linkitys lehden kotisivulle tai saman lehden eri manifestaatioihin URN-resolverin avulla.

## 7.6 National Bibliography Number (NBN)

Kirjastojen kannalta URN-nimialueista ehkä mielenkiintoisin on kansallisbibliografian ID-tunnus eli National bibliography number (NBN). Sen avulla on mahdollista käyttää kansalliskirjastojen ja niiden kanssa yhteistyötä tekevien organisaatioiden sisäisiä tunnistejärjestelmiä globaalisti ainutkertaisina toiminnallisina tunnisteinä. Kansalliskirjastolla näitä yhteistyökumppaneita on muutamia kymmeniä, joista aktiivisia oli kesäkuussa

2021 yhdeksän.

NBN-tunnuksilla identifioidiin perinteisesti sellaista kansallisbibliografiaan luetteloivaa aineistoa, jolla ei ollut muuta yksilöivää tunnistetta. URN-käytön myötä soveltamisala on laajentunut periaatteessa kaikkeen, mihin kansalliskirjastot ja niiden yhteistyökumppanit tarvitsevat pysyviä ja toiminnallisia tunnisteita, sanastojen käsitteistä metadatatietueisiin ja digitoituihin aineistoihin.

Jokainen kansalliskirjasto päättää itse, käyttääkö URN:NBN:ää ja huolehtii itse omien tunnustensa pysyvyydestä ja paikallisesta ainutkertaisuudesta. Kansalliskirjastossa NBN-tunnuksen osia olivat aiemmin f-kirjain, vuosiluku kahden tai neljän numeron tarkkuudella ja juokseva numero. Tunnuksen soveltamisalueen laajentamisen myötä f-kirjaimen jälkeen on alettu lisätä muita kirjaimia. Esimerkiksi merkintä ”fe” kertoo, että Kansalliskirjaston NBN-tunnus kuuluu syntyjään elektroniselle julkaisulle. Vastaavasti ”fd” tunnuksesta fd2010-00003245 osoittaa, että kyseinen julkaisu on digitoitu.

Kansalliskirjastojen antamien NBN-tunnusten ainutkertaisuus on taattu vain kansallisella tasolla. Koska mitään kansainvälistä NBN-tunnusten muodostamismenetelmää tai keskitettyä valvontaa ei ole, eri kirjastot voivat antaa aineistoilleen samoja tunnisteita. Esimerkiksi Saksan kansalliskirjasto voisi antaa tunnuksen fd2010-00003245 jollekin kokoelmassaan olevalle julkaisulle. URN:NBN-tunnuksiin lisätään sen vuoksi ainutkertaisuuden varmistava komponentti, joka on ISO-standardin mukainen maakoodi. Esimerkiksi Kansalliskirjaston hallinnoima Suomen NBN-aliavaruus on URN:NBN:FI, ja Saksan kansalliskirjasto hallinnoi NBN-aliavaruutta URN:NBN:DE.

Koska NBN-tunnuksen syntaksi voi olla melkein mitä vain, tunnuksien luonti on helppo automatisoida. Kansalliskirjasto käyttää esimerkiksi digitoinnissa hyväksi juoksevaa numerointia NBN-tunnusten luonnissa, ja niitä onkin generoitu miljoonittain.

Tunnusjakelun automatisointi on ainoa realistinen vaihtoehto verkkoarkistoinnin ja massadigitoinnin kaltaisissa palveluissa, joissa tunnuksia voidaan tarvita satoja miljoonia. Kansalliskirjaston ylläpitämässä suomalaisessa verkkoarkistossa<sup>28</sup> tiedostojen NBN-tunnuksena käytetään niiden sisällöstä koneellisesti tuotettuja MD5-tarkistussummia<sup>29</sup>, joiden eteen lisätään etuliite ”fea”. Tunnukset luo sovellus, jolla aineisto haravoidaan arkistoon. MD5 soveltuu tunnukseksi, koska kaksi eri tiedostoa ei periaatteessa milloinkaan saa samaa tarkistussummaa, jos niissä on pienikin ero. Ja toisaalta, identtiset tiedostot saavat aina saman tarkistussumman.

Perinteisten tunnusteiden soveltamisaloja ei voi

laajentaa niin, että ne kattaisivat verkkoaineistot. Käytettävissä olevia tunnuksia ei ole riittävästi, eivätkä henkilöresurssit riitä tunnuksien manuaaliseen jakeluun tai aineistojen kuvailuun. Siksi URN:NBN-tunniste ja muut automaattisen tunnusjakelun mahdollistavat järjestelmät ovat välttämättömiä, jotta esimerkiksi kaikki pitkäaikaissäilytettävät aineistot saavat pysyvän tunnusteen.

## 7.7 URN:n tulevaisuuden näkymät

URN-järjestelmällä on muita PID-järjestelmiä paremmat mahdollisuudet tulla kiinteäksi osaksi Internetin perusrakennetta. Se on maksuton, joten esimerkiksi esineiden Internetin vaatimat miljardit tunnuksot on mahdollista luoda kesäkuussa 2021 julkaistussa Internet-standardissa RFC 9039 kuvatulla tavalla<sup>30</sup>. URN on Internet-standardi, eikä siinä ole URI-syntaksin vastaisia linjauksia. URN ei myöskään tarjoa vaihtoehtoja ratkaisua Internetin nimipalvelulle ja Webille, vaan tukee niitä. Mutta tietyillä erikoisaloilla, kuten tieteellisten artikkeleiden tai tutkimusaineistojen tunnisteena, DOI on URN-tunnusta vahvempi, koska sillä on tieteellisten kustantajien tuki.

URN-järjestelmä perustuu kymmeneen itsenäisiin, toisistaan riippumattomiin nimialueisiin, joilla ei ole yhteistä nimittäjää. Tämä vaikeuttaa yhteistyötä esimerkiksi resolverisovelluksen kehittämisessä. Kun jokaisella nimialueella voi omat linjauksensa esimerkiksi tunnusteiden käytön ja tarvittavien palvelujen suhteen, sama resolveriohjelmisto ei välttämättä edes kelpaa niille, jotka resolveria ylipäätään tarvitsevat. Monilla nimialueilla resolverisovellus ei ole edes tarpeen. Hajautettu rakenne saattaa lisäksi vaikeuttaa kehittämistyön resursointia. Yhdelläkään URN-tunnistetta käyttävällä taholla tuskin on resursseja sellaisen resolverisovelluksen kehittämiseen, joka kelpaisi kaikille muille.

URN-tunnusteen hajautettu rakenne voi olla myös etu. Nimialueisiin perustuvan rakenteen ansioista muut pysyvät tunnisteet voivat käyttää URN-järjestelmää hyväkseen. Ensimmäiset askeleet tähän suuntaan on jo otettu: DOI-tunnus voidaan jo nyt esittää myös URN-tunnuksena, ja DOI:n käyttämä Handle.Net-sovellus pystyy resolvoimaan molemmat tunnuksot. Esimerkiksi DOI-tunnuksesta <https://doi.org/10.23978/inf.65188> tulee URN-tunnus <https://doi.org/urn:doi:10.23978/inf.65188>.

Tekniikka on edennyt nopeammin kuin standardisointi, sillä URN-nimialueen rekisteröinti DOI-tunnistelle ei ole vielä valmistunut, eikä URN:Handle-nimialueen rekisteröintiä ole edes aloitettu.

28 <http://verkkoarkisto.kansalliskirjasto.fi/va/>

29 <https://fi.wikipedia.org/wiki/MD5>

30 <https://www.rfc-editor.org/info/rfc9039>

# 8. The Handle System

## 8.1 Handle-järjestelmän tausta ja tukiorganisaatio

1990-luvun alussa Internet oli kasvanut niin, että alettiin kaivata verkkoon tallennettujen digitaalisten objektien yleistä hallinnointiratkaisua. Alan toimijoista tärkeimpiin kuulunut Defence Advanced Research Project (DARPA) rahoitti aikavälillä 1992–1996 Corporation for National Research Initiativesin (CNRI) tuohon päämäärään tähdännyttä työskentelyä. CNRI:n hankkeen tuotoksista on jäänyt elämään 1994 julkistettu tunnistejärjestelmä The Handle System ja sen taustalla oleva digitaalisten objektien hallintamalli. Digitaalisten aineistojen yleiseksi hallinnointijärjestelmäksi vakiintui kuitenkin Tim Berners-Leen CERN:issä kehittämä World Wide Web (WWW).

2020-luvulla WWW on selvästi tunnetuin ja käytetyin digitaalisten aineistojen hallinnointijärjestelmä, mutta myös CNRI:n ratkaisulla on omat kannattajansa. Siihen sisältyy kaksi standardia, Digital Object Interface Protocol (DOIP, versio 2 vuonna 2018) ja Handle Systemin seuraaja, Identifier/Resolution Protocol (IRP), joka on tätä kirjoitettaessa vielä kehitteillä. Vastuuorganisaatio on Sveitsiin rekisteröity The DONA Foundation, joka kertoo itsestään seuraavaa<sup>31</sup>:

In the late 2000s, following the many years of discussion about “Internet Governance,” it was clear that a system for managing information in digital form that was rooted in the U.S. would have limited global acceptance. Specifically, many countries and organizations interested in developing and deploying components of the Digital Object Architecture (DOA), including in particular the identifier/resolution mechanism were basically reluctant to do so.

Edellä viitataan Internetin nimipalveluun. Se on elektronisten julkaisujen hallinnoinnin kannalta ongelmallinen paitsi poliittisesti Yhdysvaltain vaikutusvallan takia, myös siksi ettei domain-nimiä voi omistaa. Niitä voi vain vuokrata määräajaksi. Handle Systemin digitaalisten aineistojen hallintamalli nojaa siihen, että jokainen digitaalinen objekti saa järjestelmässä heti pysyvän tunnuksen, joka ei muutu, vaikka aineistoa hallinnoiva organisaatio ja/tai sen domain-nimi muuttuisivat. Handle- ja samaan tekniikkaan perustuvasta DOI-järjestelmästä onkin tullut suosittuja eritoten tieteellisten julkaisujen ja data-aineistojen tunnistejärjestelminä.

<sup>31</sup> <https://www.dona.net/aboutus>

The DONA Foundation on vastannut Handle Systemiin liittyvistä standardeista vuodesta 2015. Sen itselleen nimeämiä tukiorganisaatioita ovat tämän tekstin kirjoittamisen hetkellä alkuperäinen vastuuorganisaatio CNRI ja DOI-tunnuksen vastuutaho DOI Foundation. Tulevaisuudessa tukiorganisaatioita saattaa järjestelmän käytön laajetessa tulla lisää. The DONA Foundation tekee tiivistä yhteistyötä myös International Telecommunications Unionin eli ITU:n<sup>32</sup> kanssa. ITU on esimerkiksi julkaissut DOIP-standardin version 1.0 standardina ITU-T X.1255 vuonna 2013.

The DONA Foundationin välit Internet Engineering Task Forceen (IETF) ja W3C:hen ovat etäisemmät. Jälkimmäisen osalta tämä on helppo ymmärtää, IETF:n kohdalla asiaa selittää historia. CNRI yritti tehdä The Handle Systemistä Internet-standardin, mutta IETF ei siihen suostunut, koska Handle System loi Internetin nimipalvelulle ja URN-tunnisteelle vaihtoehtoiset teknisen ratkaisun. Tämä on vastoin IETF:n periaatteita, organisaatiolla ei ole tapana hyväksyä toistensa kanssa kilpailevia määräytyksiä.

Handle-dokumentit ovat ilmestyneet RFC-julkaisuina RFC 3650–3652 vuonna 2003, mutta ne eivät ole Internet-standardeja. Niiden status on vain Informational, ei Standards Track<sup>33</sup>. IETF:n johtoelein Internet Engineering Steering Group (IESG) alleviivasi varautuneisuuttaan lisäämällä Handlen RFC-julkaisuihin seuraavan poikkeuksellisen huomautuksen (alleviivaus tekijän):

Several groups within the IETF and IRTF have discussed the Handle System and its relationship to existing systems of identifiers. The IESG wishes to point out that these discussions have not resulted in IETF consensus on the described Handle System, nor on how it might fit into the IETF architecture for identifiers. Though there has been discussion of handles as a form of URI, specifically as a URN, these documents describe an alternate view of how namespaces and identifiers might work on the Internet and include characterizations of existing systems which may not match the IETF consensus view.

”Alternate view” edellä viittaa siihen, että Handle-järjestelmässä Internetin nimipalvelun domain-nimen korvaa Handle-tunnisteen etuliite. Tunnuksissa oleva osoite (<http://hdl.handle.net>, <https://doi.org>) on vain välityspalvelimen eli proxyn osoite, Handle-järjestelmän sisällä oikea resolveri löytyy Handle-tunnuksen etuliitteen avulla. Tällä ratkaisulla on se merkittävä

<sup>32</sup> <https://www.itu.int/>

<sup>33</sup> [https://en.wikipedia.org/wiki/Request\\_for\\_Comments#Status](https://en.wikipedia.org/wiki/Request_for_Comments#Status)

etu, että toisin kuin domain-nimet, Handle-etuliitteet ovat pysyviä ja ainutkertaisia. Kertaalleen myönnettyä Handle-etuliitettä ei koskaan anneta toisen organisaation käyttöön.

Kalsea vastaanotto IETF:ssä selittää sen, miksi The Handle Systemin määrittelyt valmistellaan nykyään The DONA Foundationissa ja julkaistaan International Telecommunication Unionin standardeina. ITU:n kannalta merkittävää oli se, että Handle on hyvin toimiva tekninen järjestelmä, ei sen suhde URN-tunnisteeseen tai Internetin nimipalveluun. Lähes 20 vuotta vanhoja RFC-dokumentteja ollaan korvaamassa Identifier and Resolution Protocol -määrittelyllä<sup>34</sup>, mutta luonnosversiot eivät ole julkisia eikä uudistusprosessin tilasta ei ole saatavissa julkista tietoa. Tietävästi mitään merkittäviä muutoksia ei kuitenkaan ole tulossa.

Pesäero IETF:ään ja W3C:hen mahdollistaa Handle-standardien itsenäisen kehittämisen irrallaan muusta Internetin standardisoinnista, mutta voi myös johtaa marginalisoitumiseen.

## 8.2 Handlen soveltamisalat ja keskeiset käyttäjät

The Handle System -järjestelmällä on keskitetysti ylläpidetty resolverisovellus, Handle.Net -palvelin, jota käyttää myös DOI. Tätä kirjoitettaessa palvelimen viimeisin versio on heinäkuussa 2020 julkaistu 9.3.0<sup>35</sup>. Java-pohjaisena resolveri toimii monissa käyttöjärjestelmäympäristöissä. Muiden PID-sovellusten tapaan ohjelmisto on tehokas eikä vaadi paljoa laitteistokapasiteettia. Ohjelmiston käyttöönotto edellyttää Handle-etuliitteen eli rekisteröijätunnuksen hankkimista CNRI:ltä. Kaikki mikä näyttää Handlelta, ei ole sitä – Dspace-julkaisuarkistosovellus luo tallennettaville julkaisuille Handlelta näyttäviä tunnuksia, mutta niistä tulee virallisia ja toiminnallisia vasta, kun Handle-etuliite on rekisteröity CNRI:n ylläpitämään kansainväliseen Global Handle Registryyn<sup>36</sup>.

CNRI ylläpitää Global Handle Registry -tietokantaa, mutta se ei seuraa annettujen tunnusten määrää. The Handle System on kuitenkin tunnusten ja käyttäjien määrällä mitaten maailmanlaajuisesti eniten käytetty PID-järjestelmä. Tunnusteiden jakelun edellyttämä Handle-etuliitteen rekisteröinti<sup>37</sup> maksaa vain 50 Yhdysvaltain dollaria, ja etuliitteen vuotuinen ylläpitomaksu on sama. CNRI:n ohella kahdeksalla muulla organisaatiolla on oikeus Handle-etuliitteiden antamiseen<sup>38</sup>.

34 <https://www.dona.net/specsandsoftware#block-identifierandresolutionprotocolspec>

35 [http://handle.net/download\\_hnr.htm](http://handle.net/download_hnr.htm)

36 <https://www.dona.net/prefix/resolve>

37 <https://www.handle.net/prefix.html>

38 <https://www.dona.net/mpas>

## 8.3 Handle-tunnuksen rakenne

Tässä luvussa kuvatut Handle-tunnuksen ominaisuudet koskevat myös DOI-tunnuksia.

Handle-tunnus koostuu kahdesta osasta: etuliitteestä (prefix) sekä takaliitteestä (suffix), joita erottaa kauttaviviiva. Etuliite on tunnuksen rekisteröijän tunnus ja samalla organisaation ylläpitämän resolverin osoite, takaliite julkaisun tunnus. Tunnus esitetään yleensä HTTP URI -muodossa, jolloin sen eteen lisätään välityspalvelimen osoite. Esimerkki: <https://hdl.handle.net/10138/300023> <https://hdl.handle.net> on Handlen proxy- eli välityspalvelin, etuliite 10138 on Helsingin yliopiston tunnuksen rekisteröijän tunnus ja sen resolverin osoite Handle-järjestelmän sisällä, ja 300023 julkaisun tunnus. Koska DOI ja Handle jakavat saman infrastruktuurin, Handle-proxy on korvattavissa DOI-tunnusten proxy-osoitteella <https://doi.org>. doi.org. Mutta tämä ei vielä muuta Handle-tunnuksia DOI-tunnuksiksi, vaikka siltä näyttäisikin.

Handle-tunnusten on oltava pysyviä ja ainutkertaisia, minkä lisäksi niiltä edellytetään semanttisuuden karsimista. Tunnus ei siis saisi kertoa identifioidusta objektista mitään. Handle-tunnuksien pitäisi olla myös sekä koneiden että ihmisten luettavissa ja niiden tulisi noudattaa URI-syntaksia. Handle-määrittelyt julkistettiin kuitenkin jo ennen RFC 3986 -standardia, minkä vuoksi URI-vaatimusten ja Handlen välillä on esimerkiksi merkkivalikoimaa koskevia eroja. Tämä rajoittaa koneluettavuutta: Handle-tunnuksen sisältävän HTTP URI:n voi varmuudella tulkita oikein vain Handle-resolverisovellus.

Handle-tunnuksissa voi käyttää mitä tahansa Unicode-merkkejä. Siksi kaikki ne merkit, joilla on vuonna 2005 julkistetun URI-syntaksin mukaan erikoismerkitys, ovat olleet vapaasti käytettävissä Handle-tunnuksissa. Tällaisia ovat esimerkiksi ? ja #, joita ei voi enää soveltaa Handle-tunnuksissa RFC 3986 -standardin edellyttämällä tavalla. Ja koska on olemassa Handle-tunnuksia joissa näitä merkkejä käytetään ”väärin”, Handle-standardeja ei voi takautuvasti korjata RFC 3986:n mukaiseksi.

## 8.4 Handle julkaisujen tunnisteena

Koska tunnuksen käyttöä ei valvota keskitetysti, kyllään ei ole kattavaa tietoa siitä, miten Handlea sovelletaan. Suomessa se on suosittu esimerkiksi opinnäytteiden ja muiden korkeakoulujen julkaisuarkistoihin tallennettujen aineistojen tunnisteena. Tunnusta käytetään paljon



myös kansainvälisesti, esimerkiksi Kongressin kirjastossa ja monissa muissa organisaatioissa, joilla on laajat digitoitujen tai digitaalisenä syntyneiden aineistojen kokoelmat.

Handle-määrittäminen ei kiellä perinteisten standarditunnisteiden käyttämistä Handle-tunnuksen osana. Käytännössä esimerkiksi tästä on vaikea löytää. Kun ISBN-tunnuksen omaavalle e-kirjalle annetaan Handle-tunnus, se ei yleensä perustu kirjan ISBN-tunnukseen (ks. esim. <http://hdl.handle.net/10138/152270>). ISBN-tunnistejärjestelmän kannalta tämä käytäntö on ongelmallinen, koska esimerkiksi viitattaessa Handle voi tällöin korvata ISBN-tunnuksen. Jos Handle linkittyy laskeutumissivulle kuten yllä olevassa esimerkissä, se on tulkittavissa teoksen tunnukseksi, ja manifestaatioilla on edelleen omat ISBN-tunnuksensa. ISBN-tunnistejärjestelmän kannalta tämäkin käytäntö on silti haasteellinen, koska esimerkiksi viitattaessa on parempi käyttää Handlea kuin ISBN-tunnusta (tai manifestaation URN:ISBN-tunnusta, jos sellainen on olemassa).

## 8.5 Handlen tulevaisuuden näkymät

Standardisoinnin haasteista huolimatta on varmaa, että The Handle Systemiä käytetään vielä pitkään. Resolverisovellusta kehitetään aktiivisesti (ohjelmiston nykyinen versio on jo yhdeksäs) ja myös määrittäydokumentteja ollaan uudistamassa.

PID-tunnisteen valintaa harkitsevalle resolverisovelluksen saatavuus sekä järjestelmän levinneisyys ja tuki ovat valinnan kannalta tärkeitä asioita. Mutta jos niiden lisäksi tarvitaan hyvin hallinnoitu järjestelmä, Handlea parempi vaihtoehto on samaa teknologiaa käyttävä Digital Object Identifier eli DOI. Näiden järjestelmien suhdetta voi kuvata siten, että jokainen DOI on myös Handle, mutta kaikki Handlet eivät ole DOI-tunnuksia. DOI-standardin käynnissä oleva uudistus mahdollistaa kuitenkin sen, että Handle-tunnuksista tehdään DOI-tunnuksia takautuvasti.



# 9. DOI

## 9.1 DOI-järjestelmän tausta ja tukiorganisaatiot

Digitaalisen objektien tunnistejärjestelmän eli DOI:n<sup>39</sup> kehittäminen alkoi syyskuussa 1996, ja toimiva järjestelmä esiteltiin jo Frankfurtin kirjamesseilla lokakuussa 1997. Tuotantoon DOI otettiin vuonna 2000. Kahdessa vuosikymmenessä järjestelmä on vakiintunut ja laajasti käytössä. Käyttäjörganisaatioita on yli 5000, DOI-etu-  
liitteitä yli 150.000, ja tunnuksia 230 miljoonaa. Tunnuksia käytetään (resolvoidaan) 5 miljardia kertaa vuodessa<sup>40</sup>.

DOI standardisoitiin amerikkalaisessa kansallisessa standardissa ANSI/NISO Z39.84 jo vuonna 2000. Kansainvälinen ISO-standardi ISO 26324 Digital object identifier system<sup>41</sup> DOIsta tuli vasta melkoisen kiistelyn jälkeen vuonna 2012. Riidat eivät aiheutuneet niinkään DOI-järjestelmästä tai sen taustalla olevasta tekniikasta sinänsä, vaan DOI-tunnisteen ja perinteisten tunnistejärjestelmien suhteesta. Pelättiin, että DOI-tunnukset korvaavat ISBN:n ja ISSN:n kaltaiset perinteiset tunnistejärjestelmät digitaalisten aineistojen tunnisteena. ISO 26324-standardiin kirjattiin siksi linjaus, että DOI-tunnusta ei saa käyttää, jos aineistoille voidaan antaa perinteinen tunnus.

Tätä rajausta on noudatettu, joskin hieman tulkinnanvaraisesti. Esimerkiksi Crossref rohkaisee käyttämään DOI-tunnuksia monografoiden identifiointiin<sup>42</sup>, mutta tunnuksen pitää resolvoitua laskeutumissivuun, johon voidaan linkittää yksi tai useampia kirjan manifestaatioita. Crossrefin DOI ei siis tarkkaan ottaen ole manifestaation vaan teoksen tunnus, ja ISBN:n sijaan tekstimuotoisten teosten ISTC-tunnuksen (International Standard Text Code) tontilla. ISTC-tunnistejärjestelmän romahduksen vuoksi tästä ei toistaiseksi ole huolta.

DOI on toistaiseksi ainoa ISO:n standardoima PID-järjestelmä. ISO 26324 suomennettiin ja julkaistiin kansallisena SFS-standardina SFS-ISO 26324 Digitaalinen objektien tunnistejärjestelmä<sup>43</sup> vuonna 2019.

DOI-järjestelmästä vastaa ja sen kehittämistä koordinoi ISON nimittämä kansainvälinen rekisteriviranomainen, International DOI Foundation. Se ylläpitää

39 <http://www.doi.org>

40 <https://www.doi.org/factsheets/DOIKeyFacts.html>

41 <https://www.iso.org/standard/43506.html>

42 <https://www.slideshare.net/CrossRef/getting-started-with-books-at-crossref>

43 <https://sales.sfs.fi/fi/index/tuotteet/SFS/ISO/ID2/2/763307.html.stx>

myös DOI Handbook -käsikirjaa, joka on perusteellinen johdatus tunnistejärjestelmään<sup>44</sup>.

IDF:n rinnalla palvelua kehittävät edellä mainitut Handle-vastuutahot sekä DOI Registration Agency (RAG) eli DOI-rekisteriviranomaiset, joita on tätä kirjoitettaessa 10. Ne vastaavat jakamiensa tunnusten hallinnoinnista ja määräävät sen, mihin tunnuksia saa käyttää.

## 9.2 DOI:n soveltamisalat ja keskeiset käyttäjät

Handlesta poiketen DOI on hallinnoitu tunnistejärjestelmä. Annetut tunnuksien ja identifioitua aineistoa koskevat metatiedot kerätään tietokantaan. DOI-rekisteriviranomaiset reklaamoivat tunnuksen omistajatahoille, jos niiden antamat tunnuksien lakkaavat resolvoitumasta.

Taulukossa 1 esitetään nykyiset DOI-rekisteriviranomaiset, ja aineistot, joille niiden tunnisteita saa antaa. Kuvaus perustuu tilanteeseen syyskuussa 2021<sup>45</sup>. Tiedot voivat muuttua toimialojen muutosten tai uusien DOI-rekisteriviranomaisten myötä.

DOI-standardin mukaan tunnuksen voi antaa mille tahansa objektille, mutta DOI-rekisteriviranomaiset rajaavat tunnuksien käyttöä paljon tiukemmin. DOI-tunnusta ei toistaiseksi voi antaa esimerkiksi tutkijalle tai ontologian käsitteelle, koska yksikään nykyinen DOI-rekisteriviranomainen ei sitä salli. On toki mahdollista, että jokin nykyisistä rekisteriviranomaisista lisää esimerkiksi sanastot ja niiden termit vastuualueeseensa. Tätä kirjoitettaessa DOI-tunnuksen ominta alaa ovat kuitenkin tieteelliset artikkelit ja tutkimusaineistot, ja niiden keskeinen asema säilynee vielä vuosia.

Identifioitavissa olevan aineiston lisäksi DOI-rekisteriviranomaiset voivat määrätä myös tunnustensa toiminnallisuudesta. Esimerkiksi Crossrefin ja DataCiten tunnuksien pitää resolvoitua laskeutumissivuille, ei suoraan julkaisuihin tai tutkimusaineistoihin. Jos laskeutumissivuun on linkitetty kaksi tai useampia julkaisun manifestaatioita, DOI on siten niiden yhteinen teostason tunnus. Ja jos laskeutumissivuun linkitetyn manifestaation tiedostomuoto muuttuu, DOI pysyy samana. Siksi tieteellisten julkaisujen tunnuksena Crossrefin DOI on tulkittavissa teostason tunnisteeksi, ja manifestaatioille pitää löytää jokin muu tunnus.

Suomen kannalta merkittäviä DOI-rekisteriviranomaisia on kaksi: CSC:n yhteistyökumppani DataCite

44 <https://doi.org/10.1000/182>

45 [https://www.doi.org/RA\\_Coverage.html](https://www.doi.org/RA_Coverage.html)

Taulukko 1. DOI-rekisteriviranomaiset

Rekisteriviranomainen	Aineistot
Airiti, Inc.	Perinteiset kiinalaiset julkaisut
China National Knowledge Infrastructure (CNKI)	4000 kiinalaisen tieteellisen aikakauslehden artikkelit
Crossref	Monografiat ja niiden luvut, artikkelit, tutkimusaineistot, väitöskirjat, tutkimusraportit, apurahat, preprintit, standardit  <a href="https://www.crossref.org/education/content-registration/">https://www.crossref.org/education/content-registration/</a>
DataCite	Tutkimusaineistot sekä tieteelliset julkaisut jotka asetetaan tarjolle julkaisuarkistoissa yms. ei-kaupallisissa kanavissa.
Entertainment Identifier Registry (EIDR)	Elokuvat, TV-sarjat ja muut kaupalliset AV-aineistot
The Institute of Scientific and Technical Information in China (ISTIC)	Kiinassa julkaistavat tieteelliset julkaisut, ml. monografiat, väitöskirjat, konferenssijulkaisut sekä tutkimusaineistot
Japan Link Center (JaLC)	Japanilaiset tieteelliset julkaisut ja aineistot
Korea Institute of Scientific and Technical Information (KISTI)	Korealaiset tieteelliset julkaisut, virallisjulkaisut, patentit, tutkimusaineistot
Multilingual European DOI Registration Agency (mEDRA)	ISBN-A -tunnisteet monografioille
Publications Office of the European Union (OP)	Euroopan Unionin julkaisut

sekä Crossref, jolta ovat peräisin Tieteellisten seurain valtuuskunnan DOI-tunnukset. CSC ja Tieteellisten seurain valtuuskunta voivat antaa DOI-tunnusten luontioikeuksia kolmansille osapuolille. Niidenkin pitää solmia sopimus ja noudattaa siihen kirjattuja määräyksiä, joiden rikkomisesta voidaan langettaa sanktioita.

Toistaiseksi ei ole mitään suunnitelmia perustaa Suomeen Japanin tai Korean mallin mukaista kansallista DOI-rekisteriviranomaista. Tarve tähän voisi syntyä siten, että DOI-tunnuksia halutaan antaa Suomessa objekteille, joita minkään nykyisen DOI-rekisteriviranomaisen mandaatti ei kata, kuten vaikkapa ontologioiden termeille.

E-kirjojen julkaisujärjestelmät voivat mahdollistaa tai jopa vaatia DOI-tunnuksen antamista niissä julkaistaville aineistoille. Tällöin julkaisujärjestelmän toimittanut taho on solminut sopimuksen jonkin DOI-rekisteriviranomaisen kanssa siitä, että sen sovelluksessa voidaan käyttää DOI-tunnuksia. Asiakasorganisaation kannattaa julkaisujärjestelmäsopumusta laadittaessa tarkistaa, voiko se käyttää omaa DOI-etuliitettä. Jos voi, organisaatio voi siirtää aineistonsa DOI-tunnuksineen johonkin muuhun julkaisujärjestelmään. Jos tunnuksien etuliite on sidottu julkaisujärjestelmän toimittajaan, julkaisujärjestelmän vaihto aiheuttaa todennäköisesti myös julkaisuille annettujen DOI-tunnusten vaihtumisen. Esimerkiksi Journal.fi:ssä jokaisella lehdellä on oma DOI-etuliite artikkelien DOI-tunnusten pysyvyyden varmistamiseksi

siinäkin tapauksessa, että lehti päättää erota Journal.fi-palvelusta.

DOI-rekisteriviranomaisten toimialojen päällekkäisyydet ovat vielä vähäisempiä kuin taulukon perusteella vaikuttaa. Crossrefin ja DataCiten tunnuksia voi periaatteessa käyttää samoille aineistoryhmille, mutta valinta niiden kesken perustuu tällöin julkaisuprosessiin. Tieteellisen kausijulkaisun artikkelien tunnuksia kannattaa hankkia Crossrefiltä, mutta jos artikkelit asetetaan tarjolle korkeakoulun julkaisuarkistossa, DOI-tunnukset voi hankkia DataCiteltä<sup>46</sup>. Tästä voi seurata se ongelma, että samalla artikkelilla on kaksi DOI-tunnusta, jotka resolvoituvat artikkelin eri palveluissa oleviin kopioihin. Tältä voidaan välttyä linkittämällä Crossrefin DOI molempiin URL-osoitteisiin<sup>47</sup>. Tämä vaatii yhteistyötä julkaisijoiden kesken.

DOI-järjestelmää ylläpidetään ja kehitetään DOI-rekisteriviranomaisten International DOI Foundationille maksamalla vuosimaksuilla. DOI-rekisteriviranomaiset puolestaan veloittavat maksuja asiakkailtaan, jotka tilaavat niiltä DOI-etuliitteitä tunnusten luontia ja jakelua varten. Tunnistejakelun CSC:n ja Tieteellisten seurain valtuuskunnan kaltaiset asiakasorganisaatiot organisoivat itse, samoin kuin antamiensa tunnisteiden ja identifioitujen objektien metatietojen lähetyksen DOI-järjestelmään.

<sup>46</sup> <https://support.datacite.org/docs/datacite-or-crossref>

<sup>47</sup> <https://www.crossref.org/get-started/multiple-resolution/>

## 9.3 DOI-tunnuksen rakenne

Koska DOI on Handle-tunnus, se koostuu kahdesta osasta: etuliitteestä (prefix) sekä takaliitteestä (suffix), joita erottaa kauttaviiva. Esimerkiksi DOI-tunnus

10.1530/34571

sisältää etuliitteen ”10.1530” ja takaliitteen ”34571”. Etuliite identifioi DOI-tunnuksen antaneen tahon, jälkimmäinen on identifioidun objektin tunnus. Etu- ja takaliitteillä ei ole pituusrajoituksia, joten tunnuksen kokonaispituuden ylärajan määrää käytännössä URI-syntaksi eli RFC 3986.

Etuliite jakautuu DOI-standardin nykyisessä laitoksessa The Handle Systemin Global Handle Registry -tietokantaan tallennettuun hakemisto-osoittimeen (Directory Code) sekä tunnuksen rekisteröijän tunnukseen (Registrant Code). Ne erotetaan pisteellä. Toistaiseksi ainoa sallittu hakemisto-osoitin on ”10”, mutta DOI-standardi on tarkoitus uudistaa niin, että DOI-tunnuksen etuliitteelle ei aseteta mitään ehtoja, mikä mahdollistaa Handle-tunnusten takautuvan muuttamisen DOI-tunnuksiksi.

Nykyisen DOI-syntaksin mukaiset tunnuksen rekisteröijän tunnuksot alkavat numerosta 1000. DOI-rekisteriviranomaiset antavat niitä yhteistyökumppaneilleen; esimerkiksi Crossref Tieteellisten seurain valtuuskunnalle ja DataCite CSC:lle, jotka voivat jakaa niitä edelleen kolmansille osapuolille.

DOI-järjestelmässä on tilaa käytännössä rajattomalle määrälle tunnuksen rekisteröijäorganisaatioita ja identifioituja objekteja. Identifioitujen julkaisujen ja resoluutiotahtumien määrä on jatkuvasti kasvanut, mutta Handle.Net-ohjelmiston suorituskyky ja käytettävissä oleva tekninen infrastruktuuri eli DOI:n tapauksessa Amazonin Elastic Beanstalk –pilvipalvelu, ovat skaalautuneet hyvin. DOI-resolverit muodostavat suuren, Handle-etuliitteitä resolvereiden osoitteina käyttävän virtuaalipalvelimien verkoston, joka on osa laajemmasta The Handle System -resolvereiden verkostosta.

Teknisesti DOI on käytännössä lähes riippumaton Internetin nimipalvelusta; tarvitaan vain proxy-osoitteen <http://doi.org>. Oikean resolverin löytäminen DOI-verkoston sisällä perustuu rekisteröijän tunnukseen, joka on ainutkertainen ja pysyvä. Sitä ei siis domain-nimen tapaan voida antaa jollekin toiselle toimijalle. Jos kustantaja myy julkaisujensa jakeluoikeudet eteenpäin, niiden DOI-tunnuksot voidaan ohjata resoluutimaan uuden jakelijan DOI-resolverille. Samoin voidaan toimia myös silloin, jos kustantajan toiminta päättyy ja vastuu aineistosta siirtyy esimerkiksi kansalliskirjastolle.

DOI-tunnuksot esitetään HTTP URI –muodossa siten, että niihin liitetään DOI-resolverin proxy-osoite. Tätä kirjoitettaessa se on <https://doi.org>, aiemmin <http://dx.doi.org>. Vanhakin osoite toimii tuki edelleen.

Ja koska Handle ja DOI jakavat saman teknisen infrastruktuurin, DOI-tunnusten kanssa voisi käyttää myös Handle-tunnusten proxy-osoitetta ja päinvastoin, jolloin Handle-tunnus näyttäisi ainakin maallikon silmissä DOI-tunnukselta.

DOI-tunnus voidaan jo nyt muuntaa myös URN-tunnukseksi. DOI Handbookilla on siis myös URN <https://doi.org/urn:doi:10.1000/182>, vaikka URN-nimialuetta ”DOI” ei ole vielä virallisesti rekisteröity eikä URN-muodon soveltamismahdollisuudesta ole kerrottu virallisesti. Anomus ”DOI”-nimialueen rekisteröinnistä lähetettiin IETF:lle lokakuussa 2020.

Proxy-osoite on DOI:n kannalta merkittävä siksi, että sen avulla merkkijono on tunnistettavissa DOI-tunnukseksi. Mutta jos DOI-tunnus esitetään URN-tunnuksena, proxy-osoitteesta on mahdollista luopua sen jälkeen kun Internetin nimipalvelu tietää URN-resolvereiden sijainnit.

## 9.4 DOI kirjan tunnisteena

DOI-standardi ISO 26324 määrää, että DOI-tunnusta ei saa käyttää, jos julkaisulle voidaan antaa perinteinen ISO:n standarditunnus. Kuten edellä on kuvattu, tämä pullonkaula voidaan välttää esimerkiksi antamalla DOI manifestaation sijaan teokselle.

DOI-tunnuksen voi kuitenkin antaa myös kirjan manifestaatiolle, jos tunnus perustuu sen ISBN:ään. Silloin tunnus on molemmissa standardeissa kuvattu ISBN-A eli Actionable (toiminnallinen) ISBN.

ISBN-A rakennetaan kirjan ISBN-13 –tunnuksesta siten, että DOI-etuliitteeseen tulee ISBN-tunnuksen GS-1 -osa, maa- ja kieliryhmätunnus sekä rekisteröijätunnus, takaliitteeksi julkaisutunnus ja tarkistusmerkki, Esimerkiksi ISBN:stä 978-952-254-157-4 tulee ISBN-A 10.978952254/1574, URI-muodossa <https://doi.org/10.978952254/1574> tai <https://doi.org/urn:doi:10.978952254/1574>.

ISBN-A -tunnus muodostetaan vanhasta ISBN-10 -tunnuksesta muuntamalla se ensin ISBN-13 -tunnukseksi. ISBN-A tunnuksien toiminnallistaminen edellyttää niiden DOI-etuliitteiden rekisteröintiä ja kirjojen metatietojen tallentamista DOI-järjestelmään. Jos ISBN-A -tunnusten jakelu Suomessa alkaa, toimijat ovat kansallinen ISBN-keskus sekä mEDRA, joka koordinoi ISBN-A -toimintaa maailmanlaajuisesti.

Jokainen kustantaja voi halutessaan ylläpitää omaa DOI-resolveria ISBN-A -tunnuksilleen, koska ISBN-A:n DOI-etuliitteet ovat kustantajakohkaisia. Muutkin vaihtoehdot ovat mahdollisia. Esimerkiksi suomalaisten ISBN-A tunnuksien resoluutio voitaisiin hoitaa myös yhteisellä palvelimella, samaan tapaan kuin Tieteellisten seurain valtuuskunnan ylläpitämässä Journal.fi –palvelussa, joka soveltaa Crossrefin ylläpitämää resolveria.

Jokaisella lehdellä on eri DOI-etuliite ja siten mahdollisuus oman DOI-resolverin käyttöönottoon.

Muutamia yksittäisiä poikkeuksia (Etelä-Korea, Italia) lukuun ottamatta ISBN-A -tunnuksen käyttö ei ole yleistynyt. Muutamat kustantajat ovat tehneet omia ISBN-pohjaisia ratkaisuja, mistä hyvä esimerkki on Cambridge University Pressin 10.1017/CBO[ISBN-13]; esimerkiksi <https://doi.org/10.1017/CBO9780511542428>. Useimmiten e-kirjalle annetaan DOI-tunnus, joihin ei sisälly kirjan ISBN-tunnusta. Tällöin ISBN:n ja DOI-tunnuksen perusteella ei voi päätellä, onko kyse samasta julkaisusta. Mutta jos DOI on yhteinen esimerkiksi kirjan EPUB- ja PDF-versioille, sen ei tulisi perustua jommankumman ISBN-tunnukseen. Niiden pitäisi olla saatavissa laskeutumissivulta, jonka perusteella käyttäjä voi varmistaa löytäneensä haluamansa version julkaisusta.

## 9.5 DOI artikkeleiden tunnisteena

DOI on tätä kirjoitettaessa ylivoimaisesti suosituin tieteellisten artikkelien tunnistejärjestelmä. Artikkeleille luotiin 90-luvulla amerikkalainen kansallinen tunniste-standardi ANSI/NISO Z39.56 Serial Item and Contribution Identifier (SICI)<sup>48</sup>. Tunnukset oli tarkoitus luoda rakenteisista elektronisista artikkeleista ohjelmallisesti. Hieno ajatus ei valitettavasti toiminut käytännössä.

DOI-standardi ei määrittele takaliitteen rakennetta, eli artikkeleiden DOI-tunnisteet voivat olla muodoltaan melkein millaisia vain. Taylor & Francisin Globalizations-lehden artikkeleiden DOI-tunnuksiin sisältyy ISSN ja vuosiluku, esim. <https://doi.org/10.1080/14747731.2014.993535>. Tieteellisten seurain valtuuskunnan Journal.fi-palvelussa takaliite koostuu lehden nimestä tai sen lyhenteestä ja juoksevasta numerosta; esim. <https://doi.org/10.23989/gerontologia.87315>.

Journal.fi:n DOI-tunnukset ovat peräisin Crossrefiltä. Tunnukset resolvoituvat Journal.fi:ssä laskeutumissivuun, Silva Fennica –palvelussakin DOI-tunnusten lähde on sama, mutta toimintamalli hieman erilainen. Tunnukset resolvoituvat suoraan artikkelin HTML-versioon, joka sisältää laskeutumissivulta edellytettävät metatiedot ja linkin PDF-versioon. Molemmassa järjestelmässä DOI voidaan tulkita tekstimuotoisen teoksen tunnukseksi, vaikka Journal.fi:ssä manifestaatioita on yleensä vain yksi.

## 9.6 DOI:n tulevaisuuden näkymät

48 <https://webstore.ansi.org/standards/niso/ansini-soz39561996r2002>

DOI:n soveltama Handle system tarjoaa vaihtoehdon Internetin nimipalvelulle (ja Webille). DONA Foundationin Digital Object Architecture eli DO Architecture (DOA)<sup>49</sup> on tämän vaihtoehdoisen verkkoarkkitehtuurin kuvaus. Riippumattomuus nimipalvelusta on tilanteesta ja kuulijasta riippuen etu tai haitta. Nimipalvelun korvaavan infrastruktuurin ylläpitäminen vaatii resursseja, mutta toteutustapansa ansiosta DOI-järjestelmä on immuuni DNS-järjestelmässä hallinnoitavien nimi-alueiden omistuksen muutoksille. Laitteistoympäristönä DOI:lla on Amazonin pilvipalvelu, joka on skaalautunut hyvin järjestelmän käytön myötä.

DOI-järjestelmän käyttäjien kannalta sen taustalla vaikuttavat periaatelinjaukset voivat olla täysin toissijaisia. Tunnuksen joustavuus on suuri etu: jos takaliite voi olla mitä vain, kaikki organisaation sisäisetkin tunnistejärjestelmät voidaan uusiokäyttää DOI-tunnuksina. Myös muiden toimijoiden antama esimerkki ja järjestelmän tarjoamat palvelut ovat tärkeitä. Jos kaikki muutkin käyttävät tieteellisten artikkeleiden tunnisteena DOI:ta, on helppo seurata perässä. Myös palvelujen osalta linja on selkeä. DOI-järjestelmän on tulevaisuudessa tarkoitus kattaa paljon muutakin kuin vain tunnuksen pysyvä linkitys identifioituun julkaisuun. Priscilla Caplanin mukaan<sup>50</sup>:

The DOI initiative, of course, is intended to do more than simply provide persistence. If that was all they wanted, publishers could have implemented a PURL server with a lot less trouble. The International DOI Foundation hopes to build a comprehensive system for managing permissions and has working groups actively addressing several aspects of this, including policy, applications, descriptive metadata, and metadata for rights management.

DOI-järjestelmän varaan on alusta pitäen ollut tarkoitus rakentaa (tieteellisten) elektronisten julkaisujen jakelujärjestelmä. Tunnuksen käyttöoikeudesta maksaessaan kustantajat maksavat infrastruktuurista, joka mahdollistaa julkaisujen välittämisen, maksullisesti tai ilmaiseksi. Tämä infrastruktuuri on vasta rakenteilla, mutta todennäköisesti kehitys on jatkossakin nopeaa. International DOI Foundationin omat voimavarat ovat vähäiset, mutta DOI-rekisteriviranomaiset sekä järjestelmää käyttävät kustantajat ovat investoineet merkittävästi kehittämistyöhön.

## 9.7 DOI:n ja Handlen suhde

Merkittävin ero DOI- ja Handle-järjestelmien välillä on hallinnointi. Handle-järjestelmässä ei ole International

49 <https://www.dona.net/digitalobjectarchitecture>

50 <http://info.lib.uh.edu/pr/v9/n1/cap19n1.html>

DOI Foundationia ja DOI-rekisteriviranomaisia vastavia tahoja. Kukaan ei valvo Handle-tunnuksien käyttöä, mutta DOI-tunnuksien käyttöä valvotaan tarkemmin kuin muiden PID-tunnuksien.

Vaikka järjestelmillä on yhteinen tekninen infrastruktuuri, niiden välillä on myös eroja. Handle-tunnukset voivat olla etuliitteestä riippuen joko merkkikokoriippumattomia (case insensitive) tai merkkikokoriippuvaisia (case sensitive), mutta DOI-tunnukset ovat ASCII-merkkien osalta aina merkkikokoriippumattomia. Kaikkien muiden merkkien osalta myös DOI-tunnukset ovat merkkikokoriippuvaisia. Handle-tunnuksissa voi

käyttää mitä tahansa merkkejä, mutta DOI-tunnuksissa standardi suosittelee ja DOI-rekisteriviranomaiset voivat vaatia käytettäväksi vain tulostettavissa olevia ASCII-merkkejä, ja näin on tiettävästi myös toimittu.

Handlen ja DOI:n tiivis yhteys hyödyttää molempia järjestelmiä. Handle-tunnuksen käyttäjiltä saatavat maksut eivät välttämättä riittäisi Handle-resolverisovelluksen ylläpitoon, mutta DOI Foundation voi tukea ohjelmiston kehittämistä omista varoistaan. Yhteisiä varoja voidaan ohjata myös yhteisen Global Handle Registry –tietokannan ylläpitoon ja kehittämiseen.



# 10. PID-tunnusten käyttö Suomessa

## 10.1 URN

Kansalliskirjasto jakaa URN-tunnuksia nimialueilta URN:ISBN ja URN:NBN. URN:NBN-tunnusjakelu perustuu yhteistyökumppanikohtaisiin aliavaruuksiin (esim. URN:NBN:FI:BOF), joita ne voivat käyttää omien tarpeidensa mukaisesti yhteisten sääntöjen asettamissa rajoissa. Aktiivisia tunnusteen käyttäjiä olivat kesällä 2021 seuraavat organisaatiot:

- bof – Suomen Pankki
- csc – CSC
- hulib – Helsingin yliopiston kirjasto
- jyu – Jyväskylän yliopisto
- lb – Kielipankki (CSC)
- oulu – Oulun yliopisto
- uel – Itä-Suomen yliopisto
- ula – Lapin yliopisto
- vtv – Valtiontalouden tarkastusvirasto

URN-tunnukset ovat maksuttomia suomalaisten ISBN- ja ISSN-tunnusten tapaan.

Kansalliskirjaston palvelua on haitannut se, että kirjaston URN-resolveri on kyennyt linkittämään URN-tunnuksen vain yhteen URL-osoitteeseen kerrallaan. Kun sama julkaisu on ollut saatavissa sekä yliopiston omasta julkaisuarkistosta että Kansalliskirjaston e-vapaakappalekokoelmasta, on voinut käydä niin, että URN linkittyy vapaakappaleeseen, johon asiakkaalla ei ole pääsyä. Tämä ongelma on ratkaistu resolverin uudessa versiossa, jonka tuotantokäyttö alkoi vaiheittain keväällä 2021. Mutta vaikka resolveri pystyy linkittämään URN-tunnuksen samanaikaisesti useisiin URL-osoitteisiin, haasteeksi jää se, että resolverille ei aina saada haravoiduksi tai muuten kerätyksi identifioitua dokumentin kaikkia URL-osoitteita.

## 10.2 The Handle System

Handle-tunnuksen käytöstä ei ole olemassa kansallisia eikä kansainvälisiä tilastoja. Tunnusta sovelletaan kaikissa DSpace-julkaisuarkistoissa, mutta niiden käyttämiä Handle-etuliitteitä ei ole pakko rekisteröidä The DONA Foundationin ylläpitämään Global Handle Registryyn. Mitään tietoa ei ole siitäkään, kuinka laajalti Handle-tunnusta käytetään muualla kuin julkaisuarkistoissa, yhtä poikkeusta lukuun ottamatta. CSC on Handle-tunnuksia tarjoavan Persistent identifiers for eResearch

(ePIC) -yhteisön<sup>51</sup> jäsen. Näitä tunnuksia käytetään CSC:n palveluissa tutkimusaineistojen identifiointiin.

## 10.3 DOI

Tieteellisten seurain valtuuskunnan Journal.fi-palvelussa käytetään DOI-tunnuksia. Niiden rekisteröijätunnusten lähde on Crossref. Tieteellisten seurain valtuuskunnalla on Crossrefin kanssa ns. Sponsoring Membership (<https://www.crossref.org/community/sponsors/>) -sopimus, jonka nojalla jokaiselle palveluun liittyvälle lehdelle voidaan antaa oma DOI-rekisteröijätunnus (DOI-etuliite). Tieteellisten seurain valtuuskunta hoitaa asiaan liittyvät sopimus- ja maksujärjestelyt Crossrefin kanssa ja ratkaisee myös mahdolliset tunnuksiin liittyvät ongelmat. Oman rekisteröijätunnuksen ansiosta lehden kustantaja voi halutessaan tehdä oman sopimuksen Crossrefin kanssa, irtautua Journal.fi:stä ja Tieteellisten seurain valtuuskunnan DOI-sopimuksesta ja siirtää tunnuksiensa käsittelyn joko omalle DOI-resolverilleen tai jonkin muun toimijan ylläpitämälle resoluutiopalvelulle.

Tieteellisten seurain valtuuskunta antaa DOI-tunnuksia Journal.fi-palvelun tieteellisten lehtien ja vuosikirjojen artikkeleille. Testivaiheessa olevassa Edition.fi -palvelussa DOI-tunnuksien sovellusala laajenee monografioiden identifiointiin. Edition.fi:ssä DOI-talukin on tultaneen käyttämään Open Monograph Press eli OMP-sovelluksen<sup>52</sup> luomaa ID-tunnusta, jonka muoto on

DOI-etuliite/julkaisijan\_lyhenne.järjestelmän\_antama\_id

Jos kirjassa olevalle osakohteelle kuten erillisenä luku- na ilmestyneelle artikkelille annetaan oma DOI-tunnus, kirjan tunnuksen lisäksi ohjelmallisesti luvun tunnus.

OMP-järjestelmässä sisäisten DOI-tunnusten käyttöä puoltaa se, että nämä DOI-tunnukset generoidaan automaattisesti, mutta ISBN-pohjaiset DOI-tunnuksia pitäisi tallentaa manuaalisesti. Mutta kaikkiin julkaisuarkistoihin soveltuva ja sen vuoksi tärkeämpi peruste on se, että Crossrefin myöntämä DOI-tunnus on yhteinen julkaisun laskeutumissivuun linkitetyille manifestaatioille, joilla on eri ISBN-tunnukset. Olisi käyttäjien kannalta harhaanjohtavaa valita niistä jokin DOI-tunnuksen pohjaksi.

51 <https://www.pidconsortium.eu/>

52 <https://pkp.sfu.ca/omp/>

Luetteloinnin kannalta tämä merkitsee sitä, että Crossrefin DOI-tunnusta ei pitäisi tallentaa esimerkiksi kirjan manifestaation tunnuksiksi vaan vain linkiksi kunkin manifestaation MARC-tietueen 856-kenttään sekä teostason tunnuksiksi. Kirjan manifestaatioiden tunnus olisi edelleen ISBN. Ja jos viittauksessa käytetään linkkinä Crossrefin DOI-tunnusta, on tärkeää täsmentää viitteessä esim. ISBN-tunnuksen avulla, mitä manifestaatiota viittaus koskee.

Jos DOI-tunnus ei ole Crossrefin, vaan mEDRA:lta hankittu, ISBN-tunnukseen perustuva ISBN-A, se ei identifioi laskeutumissivua vaan alkuperäisen ISBN-tunnuksen tapaan kirjan yksittäisen manifestaation. ISBN-A -tunnusta voi siten käyttää luetteloinnissa ISBN-tunnuksen tapaan. Kaikki DOI-tunnukset eivät siis kuvailun kannalta ole samanarvoisia.

CSC jakaa DataCite-rekisteröijätunnuksia korkeakouluille, jotka voivat soveltaa DOI-tunnistetta muun muassa tutkimusdata-aineistojen ja artikkeleiden identifiointiin. DataCiten DOI-tunnusten käyttö oli alun perin rajattu tutkimusdataan, mutta soveltamisalue on laajentunut ja on nykyisin lähellä Crossrefiä. Esimerkiksi OJS-järjestelmässä on lisäosa, jolla voi rekisteröidä artikkeleille DataCiten DOI-tunnuksia. Crossrefin tunnusten tapaan ne linkittyvät laskeutumissivulle ja ovat kuvailun kannalta teostason tunnuksia.

Suomalaisen kirjallisuuden seuran julkaisuarkisto käyttää DOI-tunnuksia. Esim. DOI <https://doi.org/10.21435/skst.1423> johtaa laskeutumissivulle, josta on linkit Kaupunki tapahtumien näyttämönä -kirjan PDF- ja EPUB- versioihin. Niillä on eri ISBN-tunnukset, mutta sama DOI. Journal.fi-palvelun tapaan DOI on teoksen ja laskeutumissivun tunnus, ISBN:t manifestaatioiden tunnuksia.

Joillakin korkeakoulujen DSpace-julkaisuarkistoihin tallennetuilla e-kirjoilla on sekä DOI-tunnus että Handle, mutta ei ISBN-tunnusta. Esimerkiksi kirjan Osallistava viestintä Heldaan tallennetulla PDF-versiolla on DOI <http://doi.org/10.31885/2019.00008> ja Handle <http://hdl.handle.net/10138/302465>, jotka linkittyvät kirjan laskeutumissivulle. Painetun version ISBN on 978-952-68576-2-6, mutta PDF-versiolta tunnus puuttuu. Manifestaation tunnus on kuitenkin tarpeen esimerkiksi pitkäaikaissäilytyksessä, ja siksi kustantajien pitäisi antaa teostason DOI-tunnuksen lisäksi myös ISBN-tunnus kaikille manifestaatioille.

## 10.4 Yhteenveto

Taulukko 2 esittää tiivistettynä julkaisujen PID-tunnusteiden soveltamismahdollisuudet Suomessa. Cool URI -tunnistetta ei ole otettu huomioon, ja vaikeaa se olisikin: niitä voi kuka tahansa antaa mille tahansa.

Kaikki elektroniset julkaisut pitää kulttuurineistolain perusteella säilyttää pysyvästi. Kun ne tallennetaan CSC:n ylläpitämään PAS-palveluun, ne tarvitsevat manifestaatiotason pysyvän tunnuksen ja mieluusti myös teostason tunnuksen, jonka avulla manifestaatiot voidaan linkittää toisiinsa.. Taulukosta 2 on helppo nähdä, että PID-tunnusten saatavuudessa ja käytössä on puutteita. Esimerkiksi kotimaiselle tieteelliselle artikkelille ei saa DOI-tunnusta, ellei sitä julkaista Journal.fi:ssä tai Silva Fennicassa tai jossakin muussa DOI-tunnistetta soveltavassa palvelussa. Saadakseen DOI-tunnukset itse julkaisemansa kausijulkaisun tieteellisille artikkeleille organisaatio voi solmia DOI-tunnisteen käytöstä sopimuksen Crossrefin kanssa. Mutta jos PID-tunnuksia halutaan antaa artikkeleiden lisäksi myös muille objekteille, kuten vaikkapa artikkeleita kuvaaville metatiedoille, Kansalliskirjaston hallinnoima URN-tunniste on sopivin ratkaisu.

Suomessa on vielä paljon julkaisuja, joiden ainoa tunnus on niiden verkko-osoite eli Cool URI. Lyhyellä aikavälillä ne ovat helppo vaihtoehto koska erillistä resolverisovellusta ei tarvita, mutta riittävän pitkällä aikavälillä niiden ylläpidosta tulee vaikeaa. Kannattaa myös muistaa, että CSC:n ylläpitämässä PAS-palvelussa julkaisun alkuperäinen verkko-osoite ei sovellu tunnuksiksi muun muassa siksi, että samasta verkko-osoitteesta voidaan kerätä PAS-palveluun useita eri julkaisuja, ja sama julkaisu voidaan siirtää palveluun useista osoitteista.

Cool URI -tunnusten toimintaa on vaikea taata esimerkiksi silloin, kun domain-nimet vaihtavat omistajaa tai kun domain-nimi muuttuu organisaation nimenmuutoksen tai organisaatioiden yhdistymisen vuoksi. Ja kaikkein suurimpana haasteena: mitä tapahtuu, kun organisaatio lakkaa olemasta? Silloin domain-nimen voi vuokrata jokin muu taho, ja kaikki kyseisen domainin Cool URI -tunnukset menevät rikki.

PID-tunnistejärjestelmien käyttö on jo nyt yleistynyt CSC:n, Kansalliskirjaston, korkeakoulujen ja Tieteellisten seurain valtuuskunnan tarjoamien palvelujen ansiosta. PID-tunnusteiden käyttöä tulee kuitenkin edelleen lisätä. Tämän edellyttää konkreettisia toimintaohjeita: mihin voi ottaa yhteyttä, kun tarvitaan pysyviä tunnuksia esimerkiksi tieteellisille julkaisuille tai tutkimusaineistoille? Ja miten voimme huolehtia siitä, että saman elektronisen julkaisun kopiot eri järjestelmissä saavat saman PID-tunnuksen? Tai siitä, että saman julkaisun eri manifestaatioilla on yhteisen teostason tunnuksen lisäksi myös pitkäaikaissäilytyksen edellyttämät manifestaatiokohtaiset tunnuksukset?

Näihin kysymyksiin vastaaminen edellyttää eri PID-vastuutahojen yhteistyötä, johon turvauduttiin myös tätä ohjetta laadittaessa.

Taulukko 2. PID-tunnisteiden soveltamismahdollisuudet Suomessa

Tunnistjärjestelmä	Jakelija	Aineisto	Sovellus	Palvelut
DOI / Crossref	Tieteellisten seurain valtuuskunta	Journal.fi -palvelun artikkelit Edition.fi -palvelun monografiat Silva Fennica	HANDLE.NET	Linkki laskeutumissivuun
DOI / DataCite	CSC	CSC:n ja sen yhteistyökumppaneiden hallinnoimat data-aineistot	HANDLE.NET	Linkki laskeutumissivuun
Handle	Korkeakoulut Kielipankki	DSpace-julkaisuarkistot FIN-Clarín tutkimusaineistot(ePIC)	HANDLE.NET	Linkki laskeutumissivuun tai aineistoon
URN:ISBN	Kansalliskirjasto	e-kirjat, joilla ISBN	KK:n resolveri	Linkki julkaisuun tai metatietoihin
URN:NBN	Kansalliskirjasto	e-kirjat, joilla ei ole muuta standarditunnusta Elektra-palvelun lehtien artikkelit Kansalliskirjaston konservointi- ja digitointiyksikössä digitoidut aineistot osakohteineen FINTO-palvelun käsitteet ja toimijat	KK:n resolveri	Linkki julkaisuun tai metatietoihin  Linkki käsitteen tai toimijan tietoihin
URN:ISSN	Kansalliskirjasto Kv. ISSN-keskus	kausijulkaisut	ISSN-keskuksen resolveri	Linkit kausijulkaisun metatietoihin ISSN-tietokannassa sekä julkaisun edeltäjä- ja seuraajatietoihin

# Sanasto

## asiakas

resolveria hyödyntävä ihminen tai ohjelmisto (esim. WWW-selain)

## DNS

### Domain Name System

#### Nimipalvelu

Internetin nimipalvelujärjestelmä, joka muuttaa verkossa olevien laitteiden nimet numeerisiksi IP-osoitteiksi. Laitteet kommunikoivat keskenään IP-osoitteiden avulla, mutta ihmisille nimien muistaminen on helpompaa. Laitteiden nimet ovat yleensä IP- eli Internet Protocol -osoitteita pitkäikäisempiä.

ESIMERKKI: Verkon nimeä helsinki.fi vastaa IP-osoite 128.214.

## HTTP-protokolla

asiakas–palvelinsuhteinen viestintäprotokolla, jota käytetään tiedon siirtämiseen Internetissä

## protokolla

joukko ohjeita, joilla hallinnoidaan digitaalisten viestien vaihtoa

## resoluutio

### resoluutiopalvelu

identifioituun resurssiin liittyvien palvelujen tarjoaminen toiminnallisen tunnuksen avulla

HUOM: Resoluutio on yksinkertaisimmillaan tunnisten linkittämistä identifioitujen objektin yhteen tai useampaan URL-osoitteeseen, mutta resolveri voi tarjota myös muita palveluita, kuten linkittää toiminnallisen tunnuksen identifioitujen objektin metatietoihin, tai lähettää asiakkaalle objektiin liittyvää, resolverilla olevaa informaatiota. Kyse on kuitenkin aina uudelleenohjauksen kaltaisesta toiminnasta, jossa resolveri muuntaa pysyvän tunnuksen (muuttuvaksi) URL-osoitteeksi.

## resolveri

### resoluutiopalvelin

resoluutiopalveluja tarjoava sovellus, joko itsenäinen tai laajemman ohjelmiston moduli

HUOM: Resolverit voivat tarjota palveluita joihin verkon protokollat kuten HTTP eivät pysty, kuten vaikkapa linkin identifioitua resurssia kuvaavaan metadataan.

## tunniste

### tunnistejärjestelmä

järjestelmä, josta tunnuksia jaetaan

HUOM Tunnistejärjestelmät voivat perustua standardiin (esim. ISBN) tai ne voivat olla sisäisiä (organisaation soveltama diaarinumero).

## tunnus

uniikki merkkijono joka joko yksinään tai yhdessä muiden metatietoelementtien kanssa pysyvästi identifioi resurssin, organisaation, henkilön, paikan tai muun entiteetin

## toiminnallinen tunnus

### (actionable) persistent identifier

#### PID

pysyvänä hyperlinkkinä Internetissä käytettävä tunnistus

HUOM: Toiminnallisia tunnisteita ovat Cool URI:t ja pysyvät tunnistukset (Persistent identifiers, PID). Ne eroavat toisistaan siten, että jälkimmäiset edellyttävät erillisen palvelimen, joka muuttaa PID-tunnuksen yhdeksi tai useammaksi verkko-osoitteeksi. Cool URI -tunnusten toiminnallisuus perustuu verkon yhteysprotokolle kuten HTTP:hen, ja ne voidaan tarvittaessa uudelleenohjata dokumentin uuteen osoitteeseen joko väliaikaisesti tai pysyvästi.

ARK-, DOI- ja Handle-tunnukset ovat periaatteessa aina toiminnallisia. URN-järjestelmä on liberaalimpi, tunnuksien voi olla joko toiminnallisia tai sitten eivät. On olemassa nimialueita kuten URN:UUID, jotka ovat ei-toiminnallisia, eli tavoitteena on perinteisten tunnistusjärjestelmien tapaan vain resurssien identifiointi.

Cool URI -tunnuksia ei voi teknisesti erottaa niistä URL-osoitteista, joita ei ole tarkoitettu pysyviksi.

## URI

### verkkoaineiston tunnistus

#### Uniform Resource Identifier

jono merkkejä, jonka avulla tunnistetaan abstrakti tai fyysinen resurssi

## URL

### verkko-osoite

mekanismi, jonka avulla tunnistetaan Internet-resurssia määrittelemällä resurssin osoite ja protokolla, jota on käytettävä siihen pääsemiseksi

## URN

### URN-tunnus

#### uniform resource name

pysyvä, verkkosijainnista riippumaton resurssin tunnistus, jota käytetään pääsyyn resurssin piirteisiin tai resurssiin itseensä

### uudelleenohjaus

#### dereferencing

#### redirection

HTTP-protokollalla tai muulla verkon yhteystavalla toteutettu siirto URL:sta toiseen

HUOM: Uudelleenohjausta ja resoluutiota ei pidä sotkea toisiinsa. Uudelleenohjauksen tarjoama toiminnallisuus riippuu Internetin teknisestä infrastruktuurista eli esimerkiksi käytettävissä olevista yhteysprotokollista kuten HTTP:stä. Resoluutiopalvelun toiminnallisuus määräytyy sen mukaan, mitä resolveri kykenee tarjoamaan. PID-tunnus voidaan esimerkiksi linkittää samanaikaisesti kahteen tai useampaan URL-osoitteeseen, mihin esim. HTTP ei pysty.

PID-tunnuksen resoluution antama URL-osoite voidaan uudelleenohjata HTTP:n avulla kuten mikä tahansa muukin URL-osoite.