

5. Toiminnalliset tunnisteet – yleistä

Sisältöä päivitetty viimeksi maaliskuussa 2023.

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

Perinteiset tunnistejärjestelmät kuten ISBN ja ISSN ovat 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^[1] on suljettu organisaatio. ARK-järjestelmän standardisointi on ollut käynnissä jo yli 20 vuotta, mutta se on yhä kesken.

Selkein yhdistävä piirre on, että sekä perinteiset että PID-tunnukset ovat ainutkertaisia ja pysyviä. Samaa tunnusta ei saa antaa uudelleen millekään muulle objektille. Tunnuksen, olipa se perinteinen tai PID, pitää identifioida objekti ainakin koko sen elinkaaren ajan, ja mielellään vielä sen jälkeenkin. Myös PID-tunnukseen perustuvien hyperlinkkien pitää toimia periaatteessa pysyvästi. Jos tunnus on annettu manifestaatiolle ja identifioitu julkaisu poistetaan verkosta, sen PID-tunnus tulisi linkittää muistokiveen. Siihen voidaan tallentaa julkaisun metatietoja ja linkki mahdolliseen korvaavaan julkaisuun, kuten saman teoksen uuteen manifestaatioon modernimmassa tiedostomuodossa. Jos PID-tunnus on annettu teokselle, uusilla manifestaatioilla ei ole vaikutusta tarvittavaan toiminnallisuuteen, jos tunnus linkittyy uusienkin manifestaatioiden linkit sisältävään laskeutumissivuun. Ja jos viimeinenkin manifestaatio poistuu verkosta, laskeutumissivusta tulee muistokivi.

Sekä perinteisten että PID-tunnusten ainutkertaisuus ja pysyvyys perustuvat tekniikan sijasta tunnusten jakelun ja käytön hallinnointiin. Tämä tarkoittaa esimerkiksi sitä, että kirja saa maailmanlaajuisesti vain yhden ISBN-tunnuksen ja siihen perustuvan URN:ISBN-tunnuksen. Teoksena sama kirja saattaa saada Crossrefin DOI-tunnuksen. Kaikkien näiden tunnistejärjestelmien tunnusjakelu on huolellisesti organisoitu ja yksityiskohtaisten sääntöjen ohjaama, ja esimerkiksi Suomessa identifioidun kirjan pysyvästä säilyttämisestä vastaavat Kansalliskirjasto ja muut vapaakappalekirjastot Kulttuuriaineistolain valtuuttamana.

URN-järjestelmässä kirjojen manifestaatioiden identifiointi perustuu Suomessa ISBN-tunnuksiin. ISBN-tunnuksen omaava kirja saa kyseiseen ISBN-tunnukseen perustuvan URN:ISBN-tunnuksen. Vain jos kirjalla ei ole ISBN-tunnusta Kansalliskirjasto antaa sille NBN-tunnuksen ja siihen perustuvan URN:NBN-tunnuksen, kun kirja kuvaillaan kansallisbibliografiaan. Toisinkin on mahdollista toimia: Saksassa kansalliskirjasto antaa jokaiselle kirjastolle URN:NBN-tunnuksen, koska se ei pidä kustantajan antamaa ISBN-tunnusta riittävän luotettavana.

Tekniikan kehittyessä tiedostomuodot vanhenevat. URN:ISBN- ja URN:NBN-tunnuksilla identifioidujen elektronisten julkaisujen käytöstä tulee vuosien saatossa "antiikkia". Tällöin niiden lukeminen voi edellyttää käytöstä jo ammuin poistuneita sovelluksia ja niiden käyttämiseen tarvittavia vanhentuneita laitteita ja käyttöjärjestelmiä ja varusohjelmistoja. Hyvä kotimainen esimerkki vanhentuneesta tekstidokumentista on TEKO-tekstinkäsittelyohjelmalla kirjoitettu teksti, jota lukeakseen kirjaston asiakkaan pitäisi perehtyä edes kursorisesti sekä DOS-käyttöjärjestelmään että komentopohjaiseen TEKO-ohjelmistoon. Molemmat olivat erittäin laajasti käytettyjä vielä 1980-luvun lopulla.

Aineistojen käytön helpottamiseksi Kansalliskirjaston Doria-palvelun kaltaisissa tuotantojärjestelmissä vanhentuneet tiedostomuodot korvataan uusilla. Muunnoksen eli migraation tekee joko vastuuorganisaatio itse tai CSC:n PAS-palvelu, jonne kaikki e-aineisto siirretään säilytettäväksi. Käytöstä poistuneet tiedostot jäävät PAS-palveluun, jotta seuraavakin migraatio voidaan tarvittaessa tehdä alkuperäisestä tiedostosta. Se voidaan noutaa PAS-palvelusta asiakkaalle, joka haluaa käyttöönsä mahdollisimman autenttisen version mielenkiintonsa kohteena olevasta teoksesta. Tällöin asiakkaan on oltava valmis harjoittamaan digitaalista arkeologiaa, eli esimerkiksi TEKO-dokumentin kohdalla perehtymään DOS-käyttöjärjestelmään ja TEKO-sovellukseen.

Useimmat asiakkaat todennäköisesti kelpuuttavat julkaisujen uusimmat ja helppokäyttöisimmät versiot, jos niiden tietosisältö on täysin tai lähes sama kuin alkuperäisten. Kirjoille ja muille e-aineistoille pitää antaa manifestaatioiden tunnusten lisäksi myös teostason PID-tunnukset, koska niiden avulla asiakas saa tiedon julkaisun kaikista manifestaatioista. URN:ISBN-tunnukset eivät tähän tarkoitukseen sovi, mutta URN:NBN-tunnuksia voidaan antaa myös teoksille. DOI ja URN:NBN ovat korvanneet ISON kehittämän tekstimuotoisten teosten International Standard Text Code (ISTC) -standardin, joka kumottiin vuonna 2021.

Perinteisiä julkaisujen tunnuksia hyödynnetään käytännöllisesti katsoen kaikissa kirja-alan ja kirjastojen järjestelmissä esimerkiksi tiedonhakuun. PID-tunnukset ovat saaneet saman aseman e-aineistojen käsittelyssä, ja ne voidaan indeksoida perinteisten tunnisteiden tavoin. PID-tunnuksiin perustuva palvelutarjonta edellyttää kuitenkin tunnistejärjestelmäkohtaisia resoluutiopalvelimia eli resolveireita, ja niiden kehittäminen on vielä kesken.

Kehtyksensä alkuvaiheessa resolverit olivat yksinkertaisia sovelluksia. Esimerkiksi Kansalliskirjaston URN-resolveri toimi käyttäjän näkökulmasta kuten HTTP: se tarjosi linkityksen URN-tunnuksesta yhteen URL-osoitteeseen. Linkin luotettavuus on merkittävä lisäarvo, mutta se näkyy vain pitkällä aikavälillä. Resolverisovelluksia kehittämällä tilanne kuitenkin muuttuu. Esimerkiksi Kansalliskirjaston ja Saksan kansalliskirjaston URN-resolverisovelluksiin ja Handle. Net-resolveriin on lisätty mahdollisuus linkittää tunnus samanaikaisesti useisiin URL-osoitteisiin. Tarjolla olevat osoitteet voidaan esittää loppukäyttäjälle HTML-sivuna, tai ne voidaan lähettää toiselle resolverille tiedoksi erikseen sovittavan rajapinnan kautta. Pelkän HTTP:n avulla tällaista toiminnallisuutta ei voi toteuttaa. Seuraava askel Kansalliskirjaston resolverin kehittämisessä voi olla esimerkiksi toimimattomien URL-osoitteiden tallentaminen. Resolveri ei käyttäisi niitä toimivien osoitteiden tapaan linkitykseen, vaan verkkoarkistohakuihin. Tähänkään ei HTTP pysty, koska se "tietää" mitkä verkko-osoitteet eivät toimi vasta sitten kun vahinko on jo tapahtunut.

Resolverien palveluvalikoimassa on varmasti jatkossakin rajoituksia. Ne voisivat tarjota käyttäjille DSpace-sovelluksesta haravoimia metatietoja, mutta MARC-kuvailutiedot pitänee jatkossakin noutaa kirjastojärjestelmästä. Palvelupyynnöt pitää voida kohdentaa joko resolverille tai suoraan kohdejärjestelmään, jolloin resolveri vain välittää identifioituun objektiin liittyvän pyynnön sellaisenaan eteenpäin. Jokaisella PID-tunnisteella on erilainen tapa koodata resolverille suunnatut pyynnöt, mutta vain URN-tunnisteessa on myös menetelmä kohdejärjestelmään osoitettaville pyynnöille, sekä tekniikka sille, miten ne erotetaan resolveripyynnöistä tai normaalista URI Query -kyselystä.

URL voidaan uudelleenohjata toiseen verkko-osoitteeseen, kunhan webmaster huolehtii tästä dokumenttia siirrettäessä. Uudelleenohjauksessa on kuitenkin teknisiä rajoituksia. Esimerkiksi syvälinkit tietokannoissa oleviin objekteihin (kuten kirjaston tietokannassa oleviin bibliografisiin tietueisiin) eivät yleensä ole uudelleenohjattavissa. Ja jos verkon domainin (esim. helsinki.fi) omistus muuttuu, sen vanhojen URL-osoitteiden uudelleenohjaus on domainin uuden omistajan armoilla.

PID-tunnuksella identifioitu ja sen avulla linkitetty objekti voidaan siirtää verkossa paikasta toiseen rajoituksetta. Verkkotason tekniset muutokset tai syvälinkin muuttuminen eivät aiheuta ongelmia, koska linkitys PID-tunnuksista (uusiin) verkko-osoitteisiin hoidetaan resolveireilla. URL-osoitteiden muutokset tehdään yleensä ohjelmallisesti, haravoimalla PID-tunnukset ja päivittyneet URL-osoitteet kohdejärjestelmistä kuten DSpace-julkaisuarkistoista. Jos identifioitu objekti siirtyy esimerkiksi toiseen DSpace-arkistoon, sen muuttunut osoite haravoidaan uudesta kohdejärjestelmästä. Resolverille voidaan tallentaa muutakin haravoinnin yhteydessä saatua tai resolverin itsensä generoimaa kuvailevaa ja hallinnollista metatietoa identifioiduista objekteista. Tämän metatiedon avulla resolverit voivat tarjota uusia palveluita.

Mitään yhteisesti sovittua resolverien metatietoformaattia ei toistaiseksi ole olemassa, mutta Research Data Alliance laati PID Kernel Information -suosituksen vuonna 2018 ([10.15497/rda00031](https://doi.org/10.15497/rda00031)). Sen käytännön vaikutus resolverisovellusten kehittämiseen on ainakin toistaiseksi ollut vähäinen. Suositus perustuu Handle-teknologiaan, joten sen soveltaminen on helpompaa DOI- ja Handle-tunnisteissa kuin muissa PID-järjestelmissä. Tulevaisuudessa resolverikehitys voi edetä niin, että uusia resoluutiopalveluja ja niiden edellyttämiä metatietoja määritellään "käsi kädessä".

Toistaiseksi kaikki PID-tunnisteet ovat teknologiariippuvaisia, koska PID-tunnukset esitetään hyperlinkkeinä HTTP URI -muodossa lisäämällä tunnukseen resolverin osoite, esimerkiksi näin:

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

Osoite, edellä <http://urn.fi>, ei ole osa URN-tunnusta, kuten <https://doi.org> ole osa DOI-tunnusta. Tavoitteena on, että kaikkia PID-tunnuksia voidaan käyttää verkossa hyperlinkkeinä sellaisenaan. Tämä on mahdollista, jos a) Internetin Domain name service – eli DNS-nimipalvelujärjestelmä^[iii] "tietää" resolverien osoitteet, ja/tai b) Web-sovellukset kuten selaimet hyväksyvät PID-tunnukset hyperlinkkeiksi. Toistaiseksi pelkän PID-tunnuksen antaminen selaimessa käynnistää Google-haun, joka usein tuottaa toivotun tuloksen, joskin mutkan kautta.

ARK-tunnukset ja perinteinen DOI-tunnus olisi helppoa muuttaa hyperlinkkeiksi, koska osoiteriville kirjoitettu tunnus on helposti tunnistettavissa ja resolverien osoitteita on vain yksi (<http://n2t.net/> ja <https://doi.org>). URN:ISSN-tunniste on yhtä yksinkertainen tapaus, mutta URN:NBN-tunnisteella resolvereita on vähintään yksi jokaisessa tunnistetta käyttävässä maassa. Maaliskuussa 2023 niitä oli kolmetoista. Myös URN:ISBN-tunniste edellyttää maakohtaisia resolvereita, mutta URN:DOI-tunniste tarvitsee vain yhden, koska ne resoluoidaan DOI-tunnisteen tapaan.

Tarvittavien resolvereiden määrä havainnollistaa URN-tunnisteen ja muiden PID-tunnisteiden merkittävä eroa: nimialueisiin perustuvan rakenteensa ansiosta URN on sateenvarjo, joka voi sisällyttää itseensä perinteisten tunnisteiden lisäksi myös kaikki muut PID-tunnisteet. Maaliskuuhun 2023 mennessä URN-nimialuekisteröinnin on tehnyt vain DOI-yhteisö, mutta keskustelu asiasta on avattu myös ARK- ja Handle-yhteisöjen hallinnoissa. Yhtenä kannustimena on se, että jos jokin muu PID-tunnus esitetään URN-tunnuksena, siinä voidaan käyttää hyväksi URN-standardin ominaisuuksia. Muihin PID-tunnuksiin ei esimerkiksi ole luvallista lisätä URI-fragmenttia, mutta jos ne esitetään URN-tunnuksina, fragmenttia voi soveltaa,

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

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

^[i] <https://www.dona.net/>

^[ii] <https://tools.ietf.org/html/rfc8141>

^[iii] <https://fi.wikipedia.org/wiki/DNS>