

2. Yleistä PID-tunnistejärjestelmistä

Sisältöä päivitetty viimeksi maaliskuussa 2023.

Toiminnallisia tunnisteita eli PID-tunnisteita ovat Archival Resource Key^[i] (ARK), Uniform Resource Name^[ii] (URN), The Handle System^[iii] (Handle) ja Digital Object Identifier^[iv] (DOI) eli suomeksi Digitaalinen objektien tunnistejärjestelmä^[v]. Uusia PID-tunnistejärjestelmiä tuskin syntyy aivan lähiaikoina. Nykyiset tunnisteet on kehitetty jo 1990-luvulla, mutta edelleenkin ne täyttävät soveltajien tarpeet hyvin.

Tunnisteen toiminnallisuus tarkoittaa sitä, että sen avulla voidaan tarjota identifioituun resurssiin liittyviä palveluita. Palvelut tarjotaan resoluimalla eli linkittämällä PID-tunnus yhteen tai useampaan URL-osoitteeseen tai lähettämällä resolverilta tunnukseen liittyviä metatietoja. Linkin takana voi olla identifioitu resurssi itse, sen laskeutumissivu (landing page) tai esimerkiksi resurssin metatietoja kirjaston tietokannasta.

Linkitys toteutetaan resolveiksi (resoluutiopalvelimiksi) kutsutuilla ohjelmistoilla. Niihin kootaan toistaiseksi vain vähän identifioituja aineistoihin liittyviä metatietoja, eikä ole olemassa mitään yleisesti sovittua rajapintaa näiden metatietojen siirtämiseen. Siksi resolveiden palveluvalikoima on toistaiseksi suppea. Esimerkiksi Kansalliskirjaston URN-resolverin ensimmäisellä versiolla ainoa palvelu oli URN-tunnuksen linkitys yhteen identifioidun objektin verkko-osoitteeseen. Tulevaisuudessa resoluutiopalvelimien palveluvalikoima kasvaa; esimerkiksi Kansalliskirjaston URN-resolveriin on jo lisätty mahdollisuus linkittää URN-tunnus samanaikaisesti useisiin URL-osoitteisiin. Mietinnässä on myös vanhentuneiden URL-osoitteiden tallentaminen resolverille. Resolveri voi käyttää 404-virheen aiheuttavia osoitteita verkkoarkistohakuihin. Jos verkosta kadonnut aineisto löytyy yhdestä tai useammasta arkistosta, arkistoitujen kopioiden URL-osoitteet voidaan tallentaa resolverille.

Resolverien tarjoamia palveluita 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 linkitettävän URL-osoitteen valinta halutun palvelun kuten metatietojen haun nojalla ei onnistu. PID-resolverilla 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 HTTP-protokollan avulla kuten muutkin URL-osoitteet.

Käytettävissä olevat resoluutiopalvelut määrittävät paitsi resolveiden ja niiden metatietojen perusteella, myös elektronisten aineistojen ja niiden metatietojen hallinnointiin käytettävien sovellusten ja niiden ominaisuuksien pohjalta. Näitä sovelluksia ovat esimerkiksi asiakasliittymät, julkaisuarkistot ja pitkäaikaissäilytys- eli PAS-järjestelmät. Niihin rakennettavien uusien piirteiden avulla voidaan tulevaisuudessa toteuttaa uusia PID-tunnuksiin perustuvia resoluutiopalveluita.

Asiakasliittymän käyttäjä voi tulevaisuudessa pyytää nähtäväkseen esimerkiksi pitkäaikaissäilytyksen metadatan PAS-järjestelmästä, ja arvioida sen avulla elektronisen julkaisun alkuperäisen ja migraatioissa luotujen uudempien versioiden sisällön ja ulkoasun 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.

Uusien resoluutiopalveluiden kehittäminen edellyttää sekä resolveiden että PID-tunnistejärjestelmien kehittämistä. Standardisointityötä helpottaisi, jos kaikki PID-tunnisteet rekisteröisivät itselleen URN-nimialueen. Silloin uudistettuun URN-standardiin vuonna 2017 lisätyt ominaisuudet olisivat muidenkin järjestelmien käytettävissä ilman niiden päivittämistä.

Jos PID-tunnus on annettu teokselle, se voi säilyä toiminnallisena periaatteessa ikuisesti. Teoksen metatiedot eivät muutu, lukuun ottamatta linkkejä mahdollisten uusien manifestaatioiden kuvauksiin. Mutta myös elektronisen manifestaation kuten kirjan PDF-version PID-tunnus voi elää identifiomaansa objektia pidempään. Kun PDF-kirja on vuosikymmenien saatossa poistettu kirjaston julkaisuarkistosta ja on saatavissa vain PAS-palvelusta, PID voi ohjata käyttäjän teoksen uudempaan manifestaatioon ja tarjota lisäksi linkin PDF-version muistokiveen (tombstone), jossa kerrotaan tästä historian hämärään vajonneesta manifestaatiosta.

PID-tunnukset ja niiden tarjoama toiminnallisuus eivät ole riippuvaisia identifioidun objektin elinkaaresta tai verkossa sovelletuista teknologioista. Siksi PID-järjestelmät ovat hyviä tai suorastaan välttämättömiä apuvälineitä elektronisten aineistojen hallinnoinnissa ja pitkäaikaissäilytyksessä. On vaikea kuvitella verkko-osoitetta, joka olisi toiminnallinen vielä satojen vuosien päästä., mutta teknologiariippumaton PID-tunnus voi sitä hyvinkin olla.

^[i] https://n2t.net/e/ark_ids.html

^[ii] https://en.wikipedia.org/wiki/Uniform_Resource_Name

^[iii] https://en.wikipedia.org/wiki/Handle_System

^[iv] https://en.wikipedia.org/wiki/Digital_object_identifier

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