

ELEKTRONISTEN JULKAISUJEN SÄILYTTÄMINEN JA IDENTIFIOINTI

Kansalliskirjasto julkisti vuonna 2021 elektronisten julkaisujen identifiointioppaan. Tämä artikkeli kertoo oppaan sisällöstä ja elektronisten julkaisujen identifiointin haasteista. Kuka vastaa tunnuksien toiminnallisuuden säilyttämisestä, jos kustantajan toiminta loppuu? Ja mistä tutkija saa toiminnallisen tunnuksen sellaista tarvitessaan?

Liike-elämässä kvartaali on kolme kuukautta, tieteessä 25 vuotta, mutta kansalliskirjastoissa se voi olla 250 vuotta tai enemmänkin. Kansalliskirjastoilla on kollektiivinen vastuu julkaisujen pysyvästä säilyttämisestä, mikä useissa maissa kattaa nykyään myös elektroniset aineistot. Ja pysyvä tarkoittaa yksittäisen ihmisen näkövinkkelistä jotakuinkin samaa kuin ikuisesti: kansalliskirjastojen kansalliskokoelmissa vanhimmat aineistot ovat jo ajalta ennen kirjapainotaidon syntyä.

Painettujen aineistojen osalta säilyttämisen keinot tunnetaan, mutta elektronisten aineistojen pitkäaikaissäilytys vaatii uusia ratkaisuja. Suomessa niistä merkittävin on CSC:n toteuttama muistiorganisaatioiden yhteinen pitkäaikaissäilytys- eli PAS-järjestelmä¹. Sen tavoite on elektronisten aineistojen ymmärrettävyyden säilyttäminen. Palveluun tallennetut tiedostot pitää muuntaa tarvittaessa modernimpaan tiedostomuotoon niiden käytettävyyden takaamiseksi. Jos muunnos on häviöllinen eli julkaisun sisältö ja/tai

ulkoasu muuttuu, PAS-järjestelmässä on tarpeen säilyttää myös alkuperäinen tiedosto. Muunnoksen eli migraation jälkeen uusi tiedosto siirretään PAS-järjestelmästä omistavan organisaation tuotantojärjestelmään, jossa se korvaa vanhentuneen tiedoston.

Muuttuva tekniikka, pysyvät tunnisteet

Tekniikan kehityksen vuoksi tiedostomuodot vanhenevat. Esimerkiksi 80-luvulla julkishallinnossa laajasti käytetyllä Teko-tekstinkäsittelyohjelmalla tuotettujen tekstien lukemiseen tarvittavia ohjelmistoja ja laitteita on enää tekniikan museoissa, ja harva enää tietää, miten Mikromikko2-tietokonetta ja Teko-ohjelmistoa käytetään. Teko-sovelluksella olevat tiedostot ovat nykyään helposti käytettävissä vain jos ne on muunnettu esimerkiksi HTML-teksteiksi.

Kun maali eli julkaisun käyttökelpoinen versio vaihtuu, sovellukset päivittyvät ja kun verkon perusteknologiatkin voivat muuttua, miten voimme taata tieteellisten julkaisujen

¹ <https://www.digitalpreservation.fi/>

käytettävyyden vuosikymmenien ja jopa vuosisatojen päähän? Miten voimme varmistaa sen, että julkaisujen lähdeviiteissä oleva linkit elektronisiin aineistoihin toimivat vielä vuosien päästä? Kun URL-osoitteiden varaan rakennetut ratkaisut seisovat savijaloilla, ratkaisu on pysyvien ja teknologiariippumattomien toiminnallisten tunnisteiden kuten DOI:n ja URN:n käyttäminen.

Tekniikkariippumattomaan pysyvään tunnisteeseen perustuva linkki voidaan kaukaisessa tulevaisuudessa-kin ohjata resolverisovelluksen avulla laskeutumissivuun, johon väitöskirjan alkuperäinen ja sitä uudenaikaisemmat, migraatioilla tuotetut manifestaatiot on linkitetty tuolloin käytetyllä verkotekniikalla. Laskeutumissivulla tulevaisuuden asiakas voi makunsa mukaan valita julkaisusta joko helppokäyttöisen modernin version, tai pitkäaikaissäilytysjärjestelmään tallennetun alkuperäisen version.

Kirjastotermein laskeutumissivulle resolvoituva pysyvä tunniste identifioi immateriaalisen teoksen, joka ei milloinkaan muutu. Siksi pysyviä tunnisteita on turvallista käyttää elektronisiin aineistoihin viitattaessa tai niitä luetoitaessa.

Koska kansalliskirjastojen on suunniteltava toimintaansa satojen vuosien

aikajänteellä, Kansalliskirjasto pyrkii aktiivisesti edistämään pysyvien tunnisteiden käyttöä Suomessa. Jos julkaisija antaa tuotteilleen pysyvät tunnukset,

**Laskeutumissivulla
tulevaisuuden
asiakas voi valita
julkaisusta modernin
tai järjestelmään
tallennetun
alkuperäisen version.**

Kansalliskirjaston ei tarvitse antaa ja ylläpitää niitä itse. Voimme toki sopia kustantajien kanssa yhteistyöstä tunnustien toiminnallisuuden takaamiseksi sen jälkeenkin, jos (tai kun) kustantajan toiminta syystä tai toisesta aikanaan

päätyy. Käytännössä tämä merkitsee sitä, että Kansalliskirjaston vastuulle tulee ennen pitkää kaikkia niitä pysyviä tunnisteita, joita Suomessa on sovellettu.

Elektronisten julkaisujen identifiointiopas

Pysyvien tunnisteiden soveltamista edistetään muun muassa Kansalliskirjaston syksyllä 2021 julkaisemalla Elektronisten julkaisujen identifiointioppaalla². Se tukee pysyvien tunnisteiden käyttöä ja käyttäjiä KAM-sektorilla ja kirja-alalla. Oppaasta on tarkoitus julkaista sekä vuosittain päivittyvä PDF-versio että ajantasaisemmin ylläpidettävä HTML-versio. Tätä kirjoitettaessa PDF-versioon perustuva HTML-versio on jo valmis, mutta se ei ole vielä ilmestynyt.

Opas kattaa ARK-, DOI-, Handle- ja URN-järjestelmät ja pyrkii esittele-

² <https://urn.fi/URN:NBN:-fi-fe2021041610764>

mään ne tasapuolisesti, vahvuuksineen ja mahdollisine rajoituksineen.

Opas kertoo perusasiat DOI- ja Handle-järjestelmien suhdetta. Niilähän on sama tekninen infrastruktuuri ja tietyin rajoituksin myös Handle- ja DOI-tunnuksien rakenne on sama. Jokainen DOI on aina myös Handle, ja ISO:n DOI-standardia päivitetään parhaillaan niin, että Handle-tunnuksien takautuva muuttaminen DOI-tunnukseksi mahdollistuu.

Teknisesti Handle ja DOI ovat veljeksiä, mutta teknisesti niitä erottaa toisistaan kontrolli, kontrolli ja kontrolli. Handle-tunnuksen voi periaatteessa kuka tahansa antaa mille vain objektille. Mutta DOI:n osalta jo ISO-standardi (suomennettuna SFS-ISO 26324:2019) kieltää DOI-tunnuksen antamisen niille objekteille, joille voidaan antaa muu ISO:n standarditunnus. Handlen voi siis antaa esimerkiksi monografialle, mutta DOI-tunnusta ei.

Merkittävimpiä DOI-tunnuksien käytön säätelijöitä ovat Crossrefin ja DataCiten kaltaiset DOI-rekisteriviranomaiset tai niiden yhteistyöorganisaatiot, joiden kanssa tunnuksia antavan organisaation on solmittava sopimus. Suomessa TSV:n DOI-tunnusten lähde on Crossref ja CSC:n DataCite. Edellisen tunnuksia voi sopimuksen nojalla antaa esimerkiksi tieteellisille artikkeleille, jälkimmäisen data-aineistoille. Sopimukset määräävät myös sen, miten tunnuksien pitää resolvoitua. Tunnuksen antaja on velvollinen korjaamaan hetimiten kaikki ongelmat tunnuksen

toiminnallisuudessa.

Merkittävien aineistojen kuten keskeisten tieteellisten lehtien artikkeleiden tunnisteenä DOI on luotettavuutensa ansiosta varmasti Handlea parempi. Samaa vertailukriteeriä on hankalaa soveltaa URN- ja DOI-tunnisteen välillä. Vertailu pitäisi tehdä URN-nimialuekohtaisesti, jolloin havaittaisiin että jotkin URN-nimialueet voivat olla yhtä luotettavia kuin DOI, mutta jotkut muut eivät pärjäisi edes Handlelle. Esimerkiksi URN:UUID-tunnisteiden luotettavuudesta ei ole järin mielekästä puhua, koska tunnistheet eivät tarjoa mitään resoluutiopalveluita. Ei siis ole mitään, mikä voisi lakata toimimasta. Ja URN:NBN-tunnuksien luotettavuus pitäisi arvioida maakohteisesti, koska kukin kansalliskirjasto hallinnoi omaa osuuttaan järjestelmäänsä itsenäisesti.

Identifointioppaasta on rajattu pois perinteisten tunnisteen kuten ISBN:n ja ISSN:n soveltaminen e-julkaisuihin, sekä toimijoiden tunnistheet. Näiden tunnisteen standardijulkaisut ja soveltamisohjeet sisältävät jo periaatteessa kaiken tarpeellisen ohjeistuksen, ja näiden tunnisteen käyttö niin painettuihin kuin elektronisiin aineistoihin on vakiintunutta.

Opas kuvaa kuitenkin perinteisten tunnisteen ja toiminnallisten tunnisteen suhteet periaatetasolla. Esimerkiksi ARK- tai DOI-tunnusta ei pitäisi käyttää, jos identifioitavalle objektille voidaan antaa perinteinen tunnus, kuten ISBN. E-kirjoille annetut DOI-tun-

nukset eivät tätä sääntöä riko, koska ne ovat kirjan eri manifestaatioille yhteisiä eli teoksen tunnisteita. Tekstimuotoisilla teoksilla oli omakin standarditunnus ISTC (International Standard Text Code), mutta DOI-järjestelmän kannalta onneksaasti ISTC-järjestelmän käyttöönotto epäonnistui, ja standardi kumottiin vuonna 2021.

Toimijoiden tunnisteet eli ISNI, ORCID, RAiD ja ROR rajattiin oppaan ulkopuolelle, koska Kansalliskirjaston on tarkoitus laatia erillinen toimijoiden identifiointipias. Rajaus on tarpeen myös sen vuoksi, että toimijoiden tunnukset eivät ole määritelmän mukaisia pysyviä tunnuksia vaan Cool URI-tyyppisiä. Ne eivät käytä erillistä resolisovellusta, vaan tunnukset linkittyvät http:n avulla suoraan identifioitun toimijan tietoihin keskitetyssä tietokannassa, mikä on sekä vahvuus että Akilleen kantapää. Teknisesti ratkaisu on yksinkertainen eikä vaadi pysyvää tunnustetta, mutta toiminnallisuus perustuu tunnistejärjestelmän edellyttämään keskitettyyn tietokantaan. Jos sen ylläpito onnahtelee, on koko järjestelmä vaarassa. Tämä painajaisskenaario on toteutunut jo kerran, ISTC-standardin kohdalla. ISNI- ja ORCID-järjestelmien käyttö on tätä kirjoitettaessa paljon aktiivisempaa kuin mihin ISTC parhaimmillaankin kykeni. Siksi näiden toimijoiden tunnisteiden tulevaisuus vaikuttaa turvatulta.

Tunnistejärjestelmän valinnasta

Tunnistejärjestelmien käyttöönottoa ja soveltamista 20 vuotta seuranneena olen nähnyt tilanteita, joissa pysyvien tunnisteiden käyttöönottopäätöksiä on tehty ilman laajaa organisaation sisäistä valmistelua. Elektronisten julkaisujen tunnisteratkaisuksi saattoi kansalliskirjastoissakin tulla Cool URI, jos päätösvalta asiassa delegoitiin tai delegoitiin IT-henkilöstölle. Sen käsitys kvartaalin pituudesta kun saattaa poiketa jonkin verran perinteisemmän kirjastohenkilöstön näkemyksestä.

Nykyään on käytettävissä tunnistejärjestelmän valintaa helpottavia välineitä³⁴, mutta silloin kun niitä ei vielä ollut, päätökset saattoivat perustua melko vähäiseen tietoon. Esimerkiksi ARK-tunnisteen vahva asema Ranskassa perustuu siihen, että järjestelmän kehittäjä John Kunze kutsuttiin Ranskan kansalliskirjastoon kertomaan keksinnöstään. Hänen esitelmänsä vakuutti kirjaston päättäjät. Valittuaan ARK-tunnisteen Ranskan kansalliskirjasto on menestyksekkäästi esitellyt samaa ratkaisua muillekin keskeisille organisaatioille, kuten Ranskan kansallisarkistolle ja Louvren taidemuseolle. Ranskassa ARK-tunnisteella onkin vahva jalansija, mutta muualla Euroopassa sitä käytetään hyvin vähän.

Tunnistejärjestelmän valinta on tärkeä päätös muun muassa sen takia, että valitusta järjestelmästä on erittäin vaikeaa päästä eroon. Näillä näkymin tä-

3 <https://doi.org/10.5281/zenodo.4192174>

4 <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>



mänkaltaisiin ratkaisuihin ei onneksi ole minkään pysyvän tunnisteiden osalta tarvetta. OCLC:n Persistent URL (PURL) -tunniste ajautui jo vaaravyöhykkeelle, koska sen tekninen infrastruktuuri vanheni. Onneksi uudeksi ylläpitäjäksi tuli The Internet Archive, joka on modernisoinut tunnisteiden teknisen alustan, ja on sitoutunut pitämään sitä yllä.


Pysyvän tunnisteiden soveltamisen haasteista tavallisin lienee se, ettei valitua tunnistetta voida soveltaa kaikkeen tarvittavaan. Tämä on yksi keskeinen syy sille, miksi kansalliskirjastoille on rekisteröity URN:NBN-nimialue; sen avulla ne voivat identifioida kaiken haluamansa. Kun päätöstä valittavasta tunnistejärjestelmästä yli 20 vuotta sitten, toinen tärkeä peruste oli DOI-tunnisteiden silloinen maksullisuusjärjestely. Tunnuskohtainen vuosimaksu on hankala toimijalle, jolla on kymmeniä miljoonia identifioitavia dokumentteja.

DOI-tunnusten käyttö on sopimuksin tarkoin säädeltyä. Esimerkiksi Crossref-tunnus on julkaisun eri manifestaatioiden yhteinen ja sen pitää linkittyä niiden laskeutumissivuun. Linkki suoraan PDF-dokumenttiin on kielletty. HTML-dokumenttiin linkittäminen on sallittu, mutta vain jos teksti sisältää kaiken sen mitä laskeutumissivulta vaaditaan, kuten linkit toisiin manifestaatioihin.

Manifestaatioiden yhteinen DOI on linkitetyn datan kannalta mainio ratkaisu, koska julkaisun tulevaisuudessa luotavat uudet manifestaatiot jakavat vanhan DOI-tunnuksen edeltäjiensä kanssa. Lähdeviitteissä olevat linkit julkaisuun eivät vanhene manifestaatioiden päivittyessä. Viitteessä annetut sivunumerot toki voivat vanheta, ja sen vuoksi e-aineistoihin viitattaessa olisi-kin parempi viitata lukuun ja kappaaleeseen kuin sivuun.

Elektronisten aineistojen pitkäaikaiskäilytys edellyttää kuitenkin manifestaatiokohtaisia tunnuksia, eikä siihen voida sopimusteknisistä syistä käyttää Crossrefin DOI-tunnistetta. Kansalliskirjasto selvittääkin parhailaan TSV:n kanssa Journal.fi:n DOI-tunnuksien täydentämistä URN-tunnuksin siten, että palvelun PDF-artikkelit ovat tunnisteiden osalta PAS-kelpoisia heti julkaisemisensa jälkeen.

Pysyvien tunnisteiden käyttö Suomessa

 heisessa taulukossa on tietoja Suomessa sovelletuista pysyvistä tunnisteista ja niiden tarjoamista palveluista. Taulukko ei ole kattava, ja toivomme että tunnisteita käyttävät organisaatiot täydentävät sitä. Lisätiedot voi toimittaa Kansalliskirjastoon osoitteeseen identifointiopas@helsinki.fi.

Tunnistejärjestelmä	Jakelija / Käyttäjä	Aineisto	Sovellus	Palvelut
DOI / Crossref	Tieteellisten seurain valtuuskunta Lehtien kustantajat	Journal.fi -palvelun artikkelit Silva Fennica	HANDLE.NET	Linkki laskeutumissivuun
DOI / Crossref	Korkeakoulut (JY)	Korkeakoulun omat julkaisut ja opinnäytteet	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-Clarintutkimusaineistot (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	Kv. ISSN-keskuksen resolveri	Linkit kausijulkaisun metatietoihin ISSN-tietokannassa sekä julkaisun edeltäjä- ja seuraaja-tietoihin

Vastuista

Perinteisissä tunnistejärjestelmissä tunnuksen antajan taakka on melko kevyt. Kaikki on hyvin, kunhan tunnusta ei anneta väärän tyyppiselle aineistolle, eikä samaa tunnusta anneta uudestaan. Virheitä toki sattuu, tahallisiakin etenkin maisa joissa tunnukset ovat maksullisia.

Toiminnallisissa tunnisteissa tunnuksen antajalla on kaikki perinteisen tunnistejärjestelmän antajan vastuut, mutta sen lisäksi hänen on huolehdittava toiminnallisuuden ylläpidosta. Ja jos vastuutaho on kustantaja, tunnisteiden toiminta voi vaarantua esimerkiksi konkurssin tai lehden lakkauttamisen vuoksi. ISBN- ja ISSN-järjestelmiin kustantajan katoaminen ei juurikaan vaikuta, mutta toiminnallisista tunnisteista voi kadota se tärkein, eli toiminnallisuus.

Kustantajilla on velvollisuus luovuttaa elektronisia aineistoja vapaakappaleina kansalliskirjastoille, mutta

siihen ei sisälly velvollisuutta tarjota kustantajien ylläpitämiin resoluutio-palveluihin linkit kirjastojen vapaakappaleisiin. Kansalliskirjastoilla taas on velvoite säilyttää julkaisut, mutta nykyiset vapaakappalelait eivät sano mitään niiden alkuperäisten resoluutio-palveluiden ylläpidosta. Ja ilman erivapautta maksuista kansalliskirjastojen voi olla jo kustannussyistä mahdotonta ottaa vastuuta esimerkiksi DOI-tunnusten toiminnallisuudesta.

Toiminnallisten tunnisteiden heikoin lenkki on tunnuksien hallinnointi silloin kun vastuut siirtyvät tai niiden pitäisi siirtyä toimijalta toiselle. Kansalliskirjastojen kannalta on tärkeää sopia tunnuksien ylläpidon periaatteista silloin kun vapaakappaleesta tulee julkaisun viimeinen kappale. Tämä on tärkeää myös kaikille tunnistejärjestelmille. Jokainen toimimattomaksi muuttunut toimiva tunnus syö luottamusta niihin jokaiseen. ♦

Kirjoittaja

JUHA HAKALA
Kansalliskirjasto

ISNI <https://isni.org/isni/0000000416625064>

