

ELEKTRONISTEN JULKAISUJEN TUNNISTAMINEN

Juha Hakala

Hallinto- ja kehittämisspalvelut
Kansalliskirjasto
juha.hakala(at)helsinki.fi

Versio 4.0, 13.8.2007

URN:NBN:fi-fe20071780

Tämän artikkelin pysyvä osoite on <http://urn.fi/URN:NBN:fi-fe20071780>

Sisällysluettelo

- [1. Johdanto](#)
- [2. ISBN](#)
- [3. BICI](#)
- [4. ISMN](#)
- [5. ISSN](#)
- [6. SICI](#)
- [7. NBN](#)
- [8. Teosten tunnisteet](#)
- [9. URN](#)
- [10. DOI](#)
- [11. Tunnusten tallennus](#)

Tämä ohje antaa perustiedot elektronisten julkaisujen identifiointista ensisijaisesti kirjasto- ja kustannusalan ammattilaisille, mutta myös muille asiasta kiinnostuneille.

1. Johdanto

Kirjastot, kustantajat ja kirjakaupat ovat jo pitkään käyttäneet erilaisia tunnuksia painettujen julkaisujen identifiointiin. ISBN (International Standard Book Number) ja ISSN (International Standard Serial Number) ovat mahdollistaneet muun muassa elektronisten tilausjärjestelmien luonnin ja siten helpottaneet oleellisesti painettujen julkaisujen hankintaprosessia. Kirjastojen viitetietokannoissa nämä tunnukset mahdollistavat paitsi hyvät hakuominaisuudet, myös mahdollisuuden samaa julkaisua koskevien tuplaviitteiden yhdistämiseen, mikä on tärkeä piirre etenkin yhteisluetteloissa.

Internetissä tunnisteet ovat välttämättömiä paitsi perinteisistä syistä, myös sen vuoksi että ne ovat edellytys julkaisujen tehokkaalle paikallistamiselle. Esimerkiksi elektronisen lehden löytää helposti, jos sen ISSN on tallennettu lehden kotisivulle. Viitetietojen ja julkaisujen luotettava linkittäminen

vaatii paitsi perinteisiä tunnisteita, myös tunnistejärjestelmiä, joita ovat esimerkiksi Uniform Resource Name (URN) ja Digital Object Identifier (DOI). Ne täydentävät perinteisiä tunnisteita mahdollistamalla linkityspalveluiden rakentamisen. Näiden järjestelmien tavoitteena on URL-osoitetietoon perustuvan linkityksen korvaaminen pysyvillä linkeillä.

Elektronisen aineiston löytyvyyden parantamiseksi tunnisteet on otettava käyttöön mahdollisimman kattavasti. Tämä edellyttää käytännössä monentasoisia tunnisteita:

- Tekijöiden tunniste International Standard Name Identifier (ISNI) on vasta kehitteillä, mutta sen tarpeesta ei ole epäilystä: samasta tekijästä voidaan "luvan kanssa" käyttää monia erilaisia nimenmuotoja (Cehov, Tsehov ja niin edelleen), jotka voidaan sitoa yhteen vain standarditunnisteella.
- Teosten ja niiden ekspressioiden tunnisteita rakennetaan ISO:ssa. Niiden avulla voidaan liittää toisiinsa teoksen eri manifestaatiot (sidottu, nidottu ja elektroninen versio) sekä esimerkiksi käännökset.
- Julkaisujen (manifestaatioiden) tunnisteet. Perinteiset tunnisteet kuten ISBN ja ISSN kuuluvat tähän ryhmään.
- Julkaisujen osien tunnisteet. Serial Item and Contribution Identifier (SICI) sekä Book Item and Component Identifier (BICI) kuuluvat tähän ryhmään.

Muutamit tunnisteet ovat tätä kirjoitettaessa vasta kehitteillä. Muutamia perinteisiä järjestelmiä, kuten ISBN ja ISSN, on hiljan modernisoitu. Uudistustyöhön osallistuvat kirjastojen ohella koko muu kirja-ala ja tekijänoikeusjärjestöt. Eri osapuolten intressit eivät aina ole yhteneviä, kuten jatkossa ilmenee.

Perinteisiä tunnisteita kuten ISBN:ää voidaan hyödyntää elektronisen aineiston identifioinnissa, mutta niiden käytössä on omat rajoituksensa. Jäljempänä tarkastelen tilannetta tarkemmin kunkin järjestelmän osalta erikseen. Seuraavat periaatteet pätevät kuitenkin kaikkiin julkaisujen tunnuksiin:

- Tunniste ei välttämättä skaalaudu siten, että kaikille asianomaisen järjestelmän alaan kuuluville elektronisille julkaisuille voitaisiin antaa tunnus. Esimerkiksi ISBN-järjestelmässä kustantajille varattu alue on niin pieni, että jokaiselle Internet-kustantajalle ei voida antaa omaa kustantajatunnusta. Lisäksi kansallisilla keskuksilla ei ole voimavaroja kaikkien verkkojulkaisijoiden tietojen lähettämiseen kansainväliselle ISBN-keskukselle, joka ylläpitää ISBN-järjestelmään rekisteröityjen kustantajien luetteloa.
- Jos tunnuksien jakelua ei voida automatisoida, toiminnan laajentaminen painetusta materiaalista verkkojulkaisuihin edellyttää paljon lisää työvoimaa jakeluorganisaatiolle. Ja vaikka jakelu voitaisiin automatisoida, tunnuksen mahdollisesti edellyttämä aineiston kuvailu vaatii edelleen ihmistyötä. Esimerkiksi jokainen ISSN-tunnuksen saava lehti on luetteloitava, ja uusien/päättyvien kuvailutiedot on lähetettävä Pariisiin kansainväliselle ISSN-keskukselle globaaliin ISSN-tietokantaan tallennettavaksi. Niinpä verkkolehtien ja mahdollinen muiden verkkoresurssien mukaan ottaminen lisää kansallisten ISSN-keskusten työn määrää oleellisesti.
- Tunnuksen soveltamisala verkkoaineiston osalta voi olla epäselvä. Esimerkiksi verkkolehden määrittely on vaikeaa, mikä hankaloittaa ISSN-jakelua. Tätä ongelmaa pyritään helpottamaan uusimalla tunnusten käytösääntöjä niin, että elektroninen aineisto on otettu huomioon. Selvää on, että suuri osa verkkojulkaisuista tulee kuitenkin jäämään perinteisten tunnisteiden sovellusalueen ulkopuolelle. Kattavaan verkkoaineiston identifiointiin esimerkiksi kansalliskirjastojen ylläpitämissä verkkoarkistoissa tarvitaan

erillisratkaisu. Kansallisbibliografian ID-numeron käyttöalueen laajentaminen tarjoaa tähän yhden mahdollisuuden, jota Suomen ja eräiden muiden maiden kansalliskirjastot soveltavat.

Tunnistejärjestelmät ja linkityspalvelut

Internet-verkossa elektroniset julkaisut ”identifioidaan” usein niiden sijaintitiedon eli URLn (Uniform Resource Location) avulla. URL ilmaisee kuitenkin vain dokumentin osoitteen (esimerkiksi www.kansalliskirjasto.fi/kirjastoala/tietolinja/0107/urn.html) sekä protokollan, jota käyttäen asianomaisessa osoitteessa olevan dokumentin voi noutaa. Jossakin osoitteessa oleva dokumentti voi muuttua sisällöltään, ja sama dokumentti voi löytyä useista eri osoitteista tai siirtyä osoitteesta toiseen. Osoitteen vaihtuessa käyttäjällä ei ole tunnisteiden puuttuessa mitään tehokasta keinoa löytää dokumentin uutta sijaintipaikkaa. Tämä ongelma voidaan ainakin periaatteessa ratkaista DOI:n ja URN:n kaltaisten tunnistejärjestelmien ja niiden tarjoamien linkityspalvelujen avulla.

Tunnistejärjestelmien avulla perinteiset tunnisteet alkavat ”toimia” verkossa. Linkityspalvelu on se osa tunnistejärjestelmää, joka tästä toiminnasta vastaa. Yksinkertaistaen linkityspalvelu on tietokantasovellus, johon tallennetaan julkaisun pysyvä tunniste ja sen (muuttuvat) sijaintipaikat verkossa. Tämän taulukon ajantasaisuus on palvelujen toiminnan kannalta kriittinen vaatimus, kuten sekin että tunnistejärjestelmä on toteutettu niin, että oikea linkityspalvelu voidaan löytää verkosta. URN ei toimi, jos se linkittyy väärään URL-osoitteeseen. Niinpä URN:n kaltaisia tunnistejärjestelmiä tulee soveltaa vain silloin, kun julkaisu on tarkoitus säilyttää joko pysyvästi tai ainakin hyvin kauan. Koska kustantajilla ja kirjakaupoilla ei tätä intressiä ole, tunnistejärjestelmien innokkaimpia soveltajia ovat olleet kirjastot ja ennen kaikkea kansalliskirjastot vapaakappalekokoelmineen.

Eri tunnistejärjestelmät tarjoavat erilaisia palveluita. URN-järjestelmän avulla käyttäjä voi periaatteessa saada esimerkiksi dokumentin itsensä, sen kuvailutiedot tai vain listan URL-osoitteista joista dokumentin voi hakea, jos on siihen oikeus. Palveluja voidaan aikaa myöten määritellä lisää, kun Internet kehittyy ja siihen tulee uusia ominaisuuksia. Mutta kun verkossa oleva elektroninen kirja identifioidaan esimerkiksi ISBN-tunnuksella, tuo ISBN pitäisi jo nyt esittää URN- tai DOI-tunnuksena jo pelkästään staattisen linkityksen takaamiseksi.

Tunnistejärjestelmistä URN on Internet-yhteisön 90-luvun puolivälissä käynnistynyt aloite, DOI puolestaan on suurten kustantajien rakentama järjestelmä, jonka tavoitteena on tarjota kirja-alan tarvitsemat palvelut. Toimiakseen nämä järjestelmät edellyttävät hyvin erilaisen infrastruktuurin; tästä enemmän URN- ja DOI-järjestelmiä käsittelevissä luvuissa. Muita tunnistejärjestelmiä ovat OCLC:n (Online Computer Library Center, <http://www.oclc.org/>) kehittämä PURL (Persistent URL, <http://purl.oclc.org/>) sekä John Kunzen Kalifornian yliopiston digitaalista kirjastoa varten kehittämä ARK (http://www.sspnet.org/documents/149_Kunze.htm). Näiden suosio on DOI- ja URN-järjestelmiin verrattuna vähäinen, eikä niitä ole viime vuosina aktiivisesti kehitetty.

Keskustelua eri tunnistejärjestelmien teknisestä paremmuudesta on käyty paljon, mutta mitään yksimielisyyttä asiassa ei ole saavutettu. Joitakin eroja on helppo nähdä: PURL-järjestelmää ei ole koskaan yritettykään standardoida, ja järjestelmän nykyinen rakenne ei mahdollista PURL-linkityspalvelimien välistä kommunikointia.

Linkityspalvelun tärkein tekninen piirre on se, miten tunnisteiden ja dokumenttien linkittäminen verkossa käytännössä tapahtuu. DOI- ja URN-järjestelmien nykyiset toteutukset ovat teknisesti yksinkertaisia. DOI, URN ja PURL hoitavat linkityksen taulukon avulla; ARK-järjestelmän ratkaisu on periaatteessa sofistikoituneempi.

Linkityspalvelun laadun arvioinnissa keskeisin kriteeri on järjestelmän teknologiariippumattomuus. Jos tunnistejärjestelmä rakentuu jonkin olemassa olevan teknisen palvelun tai tietoliikenneprotokollan kuten http-standardin varaan, sen elinkaari jää todennäköisesti hyvin lyhyeksi, korkeintaan parin vuosikymmenen mittaiseksi. Tämä voi riittää kaupallisille yrityksille, mutta kansalliskirjastolle moinen ratkaisu ei sovellu, koska aineiston pitää olla käytettävissä periaatteessa ikuisesti.

Linkityspalveluista on sitä enemmän hyötyä, mitä suuremmalla osalla pitkään säilytettävistä dokumenteista on tunniste. Siksi elektronisten julkaisujen tunnisteiden jakelu kannattaa tehdä mahdollisimman nopeasti ja kattavasti; periaatteessa mikään tärkeä ja/tai pitkään säilytettävä verkkojulkaisu ei saisi jäädä ilman tunnisteita. Kansalliskirjasto pyrkii edesauttamaan tätä kehitystä muun muassa tarjoamalla tietoa tunnisteista ja tunnistejärjestelmistä, ylläpitämällä kansallista URN-linkityspalvelua sekä laajentamalla kansallisbibliografian ID-numeron käyttöaluetta sellaisiin verkkojulkaisuihin, joille ei muuta tunnusta voida antaa.

Tarkastelen seuraavaksi alla lueteltujen tunnisteiden ja tunnistejärjestelmien soveltuvuutta elektronisten julkaisujen identifiointiin. Selvitys koskee toisaalta teknistä hyödynnettävyyttä, toisaalta tunnuksen jakelukäytänteitä.

1. *International Standard Book Number (ISBN)*
2. *Book Item and Contribution Identifier (BICI)*
3. *International Standard Serial Number (ISSN)*
4. *International Standard Music Number (ISMN)*
5. *Serial Item and Contribution Identifier (SICI)*
6. *Kansallisbibliografian ID-numero (NBN, sanoista National Bibliography Number)*
7. *Teosten ID-tunnukset*
8. *Uniform Resource Name (URN)*
9. *Digital Object Identifier (DOI)*

2. ISBN

International Standard Book Number kehitettiin 1960-luvun lopulla kansainväliseksi kirjojen ja muun monografia-aineiston tunnisteeksi. Tavoite on saavutettu: ISBN-tunnuksia käytetään yli 160 maassa, eikä yksikään niin sanottu tärkeä valtio ole jäänyt järjestelmän ulkopuolelle. Tosin muutamissa maissa, kuten Meksikossa, ISBN-jakelu ei toimi kunnolla. Lisäksi ISBN-tunnusten maksullisuus esimerkiksi Englannissa ja Yhdysvalloissa on leikannut järjestelmän suosiota.

Viime vuosina ISBN-tunnusta on ryhdytty käyttämään uusien aineistoryhmien kuten elektronisten julkaisujen identifiointiin. Periaatteena on, että ISBN voidaan antaa kaikille julkaisuille, jotka sisällöltään vastaavat kirjaa. Sen vuoksi sillä, missä muodossa sisältö esitetään tai välitetään, ei ole merkitystä. Fyysinen muoto voi olla painettu kirja, audiovisuaalinen tallenne tai verkkojulkaisu.

ISBN-tunnus koostui alun perin kirjainlyhenteestä ISBN sekä kymmenestä merkistä, jotka on jaoteltu neljään ryhmään. Vuoden 2007 alusta käyttöön otetussa uudessa EAN-yhteensopivassa ISBN-tunnuksessa on 13 merkkiä viidessä ryhmässä.

Luettavuuden parantamiseksi nämä ryhmät erotetaan väliviivoin (tai tavuviivoin). Julkaisussa numerotunnuksen eteen pitää aina merkitä kirjaimet ISBN. Elektronisessa julkaisussa tämä takaa sen, että ISBN-tunnus löytyy myös automaattisessa indeksoinnissa ja tulkitaan oikein.

- Etuliite, joka on joko 978 tai 979.
- Maa- tai kieliryhmätunnus, joka identifioi joko maan (esimerkiksi 951 ja 952 = Suomi) tai kieliryhmän (3 = Saksa, Sveitsin saksankielinen alue sekä Itävalta). Maatunnuksen pituus on 1-5 merkkiä, riippuen asianomaisessa maassa tai kieliryhmässä julkaistujen dokumenttien määrästä. Tunnukset jakaa kansainvälinen ISBN-keskus. Maatunnuksen pituus vaikuttaa oleellisesti jaettavissa olevien kustantaja- ja julkaisutunnuksien määrään. Suomessa tunnus on kolminumeroinen toisin kuin esimerkiksi Ruotsissa, Tanskassa ja Norjassa, joissa se on kaksinumeroinen.
- Kustantajatunnus. Tämän tunnuksen antaa kustantajalle kansallinen ISBN-keskus. Kustantajatunnus voi olla 1-7 merkkiä pitkä. Maksimi riippuu maatunnuksen pituudesta; Suomessa se on viisi merkkiä. Suurilla kustantajilla on lyhyt tunnus (ja paljon julkaisutunnuksia) ja pienillä kustantajilla vastaavasti pitkä tunnus. Todella pienet kustantajat saavat ISBN:n yhteisestä kustantajatunnuksesta, joka on Suomessa toistaiseksi 952-91.
- Julkaisutunnus. Kustantaja jakaa teoksilleen tunnukset ISBN-keskuksen antamien ohjeiden mukaan. Esimerkiksi saman teoksen eri versiot (nidottu, sidottu yms.) saavat eri ISBN-tunnuksen, sen sijaan muuttamattomat lisäpainokset saman tunnuksen. Samaa ISBN-tunnusta ei saa koskaan käyttää uudestaan. Teoksen tunnus voi olla 1-7 merkkiä pitkä, joten yhden kustantajan käytettävissä olevien numeroiden määrä vaihtelee kymmenestä kymmeneen miljoonaan. Suomessa tunnisteiden maksimi on 100.000 per kustantaja, koska maantunnuksemme on kolmen numeron mittainen.
- Tarkistusmerkki lasketaan käyttäen Modulus 10 -algoritmia. Se voi olla numero 1-9. Tarkistusmerkin käyttö mahdollistaa sen, että kirjastojärjestelmät ja muut julkaisutietoja käsittelevät sovellukset kuten kirjakauppojen hankintajärjestelmät voivat tarkistaa ISBN-tunnuksen oikeellisuuden.

Maatunnukset ja kustantajatunnukset on jaettu siten, että tunnukset eivät mene päällekkäin vaikka ne "tavutettaisiin" eri tavoin osiin. Suomessa WSOY:llä on kustantajatunnus "0" ja muita "978-951-0" -alkuisia kustantajatunnuksia ei ole. Vastaavasti Yhdysvalloissa on maan tunnus "0", ja muita merkillä 0 alkavia maakoodeja ei ole jaettu.

Suomessa on yli 4700 ISBN-järjestelmään rekisteröityä kustantajaa. Tämän lisäksi jaetaan yksittäisiä 978-952-92 -alkuisia tunnuksia noin 1500 vuodessa omakustantajille ja muille hyvin pienille kustantajille. Maailmanlaajuisesti ISBN-kustantajatunnuksia on jaettu jo yli puoli miljoonaa. Kansainvälinen ISBN-keskus, joka koordinoi ISBN-järjestelmän käyttöä ja tukee kansallisia ISBN-keskuksia, ylläpitää luetteloa tunnuksen omaavista kustantajista, johon kansalliset keskuskeskukset lähettävät tarvittavat tiedot.

Internet-käyttö

ISBN-tunnus voidaan antaa kaikille kirjan kaltaisille julkaisuille riippumatta julkaisun fyysisestä muodosta. Toistaiseksi ISBN-tunnuksia on annettu etupäässä kustantajille, jotka julkaisevat verkko-

tai muita elektronisia julkaisuja painettujen julkaisujen ohella. Saman teoksen erilaiset painetut ja elektroniset versiot saavat aina eri ISBN:n. Esimerkiksi kirjan nidotulla ja sidotulla versiolla tai samasta teoksesta tehdyllä äänikirjalla on eri ISBN, muun muassa sen vuoksi että kirjakaupan on kyettävä erottamaan nämä versiot toisistaan. Saman logiikan mukaan elektronisen julkaisun eri variantit - HTML-versio, PDF-versio, ja niin edelleen - saavat eri ISBN-tunnukset. Eri elektroniset versiot voivat poiketa ilmiänsultaan oleellisesti toisistaan, vaikka intellektuaalinen sisältö olisi sama. Siksi eri ISBN-tunnusten käyttö on hyvin perusteltua.

ISBN-käytösäännöt määrittelevät tarkoin sen, milloin julkaisulle on annettava uusi ISBN. Painovirheiden korjaaminen tai muuttamaton lisäpainos ei vielä edellytä uutta ISBN:ää. Julkaisun nimen muutos vaatii aina uuden ISBN-tunnuksen. Esimerkiksi Väinö Linnan Tuntematon sotilas ja sen alkuperäinen versio Sotakirja saavat eri ISBN-tunnuksen, vaikka kyseessä on (likimain) sama teos.

Moniosaisen teoksen kaikki osat saavat oman ISBN-tunnuksen, sen lisäksi että koko teos saa oman tunnuksen. Elektronisten julkaisujen osalta tämä periaate on ongelmallinen: jos verkkokirjan jokainen luku on itsenäinen kokonaisuus (tiedosto tai joukko yhteen linkattuja tiedostoja), pitäisikö joka luvulle antaa oma ISBN? Tai jos yksi luku koostuu monesta tiedostosta, jotka ovat erikseen kaupan, pitäisikö nekin identifioida erikseen? Periaatteessa näin pitäisi toimia, koska jokainen dokumentti on identifioitava. Tästä seuraisi kuitenkin ISBN-tunnusten kysynnän nopea kasvu.

Uuden ISBN-standardin mukaan myös (elektronisten) kirjojen osille voi antaa oman ISBN-tunnuksen. Käytännössä tämä on tarpeen silloin kun esimerkiksi kirjan luku on erikseen ostettavissa. Tunnusten käyttö tuskin riistäytyy käsistä, koska sitä tasapainottaa vaatimus toimittaa kansalliselle ISBN-keskukselle identifioidun kirjan tiedot ONIX-formaatissa. Suomen kirja-ala laatii tätä kirjoitettaessa kansallista ONIX-formaattia sekä välineitä ja ohjeistusta, jotka helpottavat formaatin käyttöä.

ISBN:n tulevaisuus

Vaikka kustantajien on toimitettava kirjojen kuvailutiedot kansallisille ISBN-keskuksille ONIX-muodossa, jo lähitulevaisuudessa ISBN-tunnuksia tarvitaan selvästi enemmän kuin nyt. Suuri osa uusista tietokirjoista ja ennen pitkää myös romaaneista tulee ilmestymään sekä painettuna että elektronisena versiona. Tekijänoikeuksista vapaat kirjat digitoidaan lähivuosien mittaan kattavasti verkkoon koko maailman käyttöön. Klassikoista laaditaan kriittisiä laitoksia, jotka on erotettava alkuperäisistä versioista. Teoksen jokainen elektroninen manifestaatio voi koostua monista itsenäisistä osista, jotka voidaan identifioida (ja myydä) erikseen.

Uusi ISBN-standardi ei lisännyt tunnuksen kapasiteettia merkittävästi. 978-etuliitteen lisäksi ISBN:lle on varattu myös etuliite 979, mikä kaksinkertaistaa tarjolla olevien tunnusten määrän. Jatkossa EAN-järjestelmästä voidaan irrottaa kirjoille uusia kolmen numeron etuliitteitä, joskaan ei rajattomasti. Tämä ratkaisu on käytännöllinen; se takaa ISBN:n käyttökelpoisuuden vuosikymmeniksi, ellei tunnusten käyttö kasva räjähdysmäisesti. Jos näin kävisi, kirjan osille voidaan kehittää oma tunniste.

Tällä hetkellä 13-merkkinen ISBN-tunnus voi alkaa vain etuliitteellä 978, mutta vuoden 2008 alkupuolella otetaan käyttöön asteittain etuliite 979 niissä maissa, joissa vanhat kustantajatunnukset ovat loppuneet. Suomessa vanhat kustantajatunnukset riittävät vielä useammaksi vuodeksi, mutta muualla maailmassa 979-alkuisia tunnuksia tulee vähitellen jakeluun. Niillä ei ole enää 10-

merkkistä vastinetta ja näissä uusissa tunnuksissa muuttuvat myös maa- ja kieliryhmätunnukset sekä kustantajatunnukset.

ISBN:lle on ehdotettu vaihtoehtoisia, pidemmälle meneviä laajennusmenetelmiä. ISBN-13 – järjestelmää kehitettäessä jotkut asiantuntijat olivat sitä mieltä, että tarkistusmerkki on tarpeeton, koska ISBN-tunnuksia ei enää tallenneta käsin. Uuden ISBN:n kapasiteettia olisi voitu laajentaa 10-kertaiseksi luopumalla tarkistusmerkistä. Mutta käytännössä ISBN-tunnuksia tallennetaan edelleen käsin. Kustantajilta tulevien virheellisten tunnusten määrä ei ole suuri mutta kuitenkin niin merkittävä, että tarkistusmerkistä ei haluttu luopua. Asiaan palattaneen kun standardia seuraavan kerran uudistetaan.

13 merkin mittaisen ISBN:n asemesta esitettiin siirtymistä 16 tai jopa 25 merkin mittaiseen tunnukseen. Näin mittava laajentaminen olisi taannut sen, että ISBN-tunnuksen rakennetta ei olisi tarvinnut pitkään aikaan muuttaa uudelleen ainakaan kapasiteetin puutteen vuoksi. Etenkin 25 merkin mittainen ISBN olisi riittänyt todella kauan. Mutta uuden ISBN-järjestelmän piti olla yhteensopiva (13-merkkisen) EAN-koodin kanssa, mistä syystä pitkä ISBN-tunnus piti kuopata.

Eräät asiantuntijat suosittelivat numeroiden käytön asemesta heksadesimaalilukujen käyttöä. Tämä ehdotus hylättiin ISBN-revisiossa sen vuoksi, että EAN-järjestelmä ei hyväksy heksadesimaaleja.

Kustantajien edustajat ISBN-standardia uusineessa työryhmässä esittivät, että nykyisestä ISBN-rakenteesta luovuttaisiin, jolloin ISBN:stä olisi tullut ISSN:n kaltainen "tyhmä" numero. ISBN:n tyhmistäminen olisi kaksinkertaistanut käytettävissä olevien tunnusten määrän. Ehdotuksen taustalla oli se tosiasia, että Yhdysvalloissa ja Englannissa ISBN ei enää kerro totuutta julkaisumaasta. Esimerkiksi New Yorkissa iso osa julkaisu- ja tiedustelu- ja amerikkalaisten, New Yorkissa julkaistujen teosten tunnisteena käytetään 3:lla (Saksan maatunnisteella) alkavia ISBN-tunnuksia. Kirjastojen kannalta tästä ei välttämättä seuraa mitään ongelmaa; jos kustantaja on saksalainen, julkaisu luetteloidaan Saksan kansallisbibliografiaan.

Useimmissa muissa maissa maatunnukset ja kustantajatunnukset ovat edelleen varsin luotettavia, ja enemmistö ISBN-työryhmän jäsenistä – etenkin kirjastojen ja kirjakauppojen edustajat - olivat vastahakoisia luopumaan vanhasta ISBN:n rakenteesta. Rakenteinen ISBN soveltuu hyvin myös URN-linkityspalveluille, toisin kuin ehdotettu "tyhmä" ISBN. Jos ISBN alkaa 978-951:llä, voidaan päätellä että kirjan tiedot löytyvät Suomen kansallisbibliografiasta. Jos ISBN:n rakenne joskus tulevaisuudessa muutetaan ISSN:n kaltaiseksi, 978951-alkuisia ISBN-tunnuksia voi jakaa periaatteessa kuka tahansa. Tunnusta vastaavan julkaisun tiedot voidaan löytää helposti vain jos rakennetaan ISSN-tietokantaa vastaava globaali ISBN-tietokanta.

Periaatteessa ISBN-järjestelmä toimii vuoden 2007 alusta ISSN-verkoston tapaan, eli kansallisten keskustusten pitäisi koota kirjoja koskeva bibliografinen tieto ja luovuttaa se ONIX-muodossa (<http://www.editeur.org/onix.html>) edelleen globaalia ISBN-tietokantaa ylläpitävälle kansainväliselle ISBN-keskukselle.

Tässä ehdotuksessa on kaksi ongelmaa. Ensinnäkin ONIX-vaatimus on kirjastoalan näkökulmasta erikoinen, koska a) ONIX ei ole ISO-standardi, ja b) yksikään kansalliskirjasto ei toistaiseksi tuota bibliografista tietoa ONIX-muodossa vaan MARC-formaatissa. Suomessa MARC säilyy jatkossakin kirjastojen ja kansallisbibliografian formaattina. Mutta kansalliskirjasto selvittää yhdessä kirja-alan kanssa mahdollisuudet MARC – ONIX –konversioon. Edellytyksiä konversion

onnistumiselle parantaa oleellisesti se, että kirja-ala on yhteistyössä kansalliskirjaston kanssa laatimassa suomalaista versiota ONIX-formaatista.

Toinen pääsärky on, että kansainvälisen ISBN-tietokannan luonti ja ylläpitäminen on suuri työ ja edellyttää kansainvälisen ISBN-keskuksen valtuuksien ja taloudellisten resurssien oleellista lisäämistä. Teknisesti tietokanta on nykyisillä palvelimilla toteutettavissa varsin helposti; OCLC:n WorldCat-tietokannassa oli kesällä 2007 jo 85 miljoonaa tietuetta ja vie pitkään ennen kuin ISBN-tietokanta on lähimainkaan näin suuri. Suurimmat ongelmat ovat hallinnollisia; datan kerääminen yli 150 maasta on vaikeaa. ISSN-järjestelmässä on mukana noin 70 maata, ja jo niistäkin osa jättää tietonsa lähettämättä globaaliin ISSN-tietokantaan.

Kirjastojen, kustantajien ja kirjakauppojen atk-järjestelmien sopeuttaminen uuteen ISBN:ään oli iso työ joka vaati runsaasti aikaa. Tätä kirjoitettaessa läheskään kaikki ohjelmat eivät toimi tyydyttävästi. Kokonaisuutena ottaen uuden ISBN-tunnuksen käyttöönotto on kuitenkin sujunut hyvin, lukuun ottamatta bibliografisten tietojen luovutusta. Suomessa edellytykset siihen syntyvät vuoden 2007 mittaen, ja käytännön toimintaan päästäneen 2008. Kansainvälisessä katsannossa Suomi lienee ONIXin soveltamisessa yksi edelläkävijöistä.

3. BICI

Kirjan lukujen yms. osien identifiointiin kehitettiin USA:ssa tunnus nimeltä Book Item and Component Identifier eli BICI, jonka luonnoksen NISO julkisti standardiluonnoksena elokuussa 2000. Teksti on edelleen poimittavissa osoitteesta <http://www.niso.org/pdfs/BICI-DS.pdf>, vaikka se oli testikäytössä vain 31.12.2002 saakka. Testikäyttäjiä oli vähän, eikä standardiluonnoksen tien jälkeenkään ole ollut ruusuinen. BICIä pidettiin vanhanaikaisena ja käytännössä kustantajien DOI-aloite syrjäytti sen.

BICI toimii artikkelien ja lehden numeroiden identifiointiin tarkoitettuna, ISSN:ään perustuvan SICI-tunnuksen tavoin; sen avulla voidaan identifioida kirjan yksittäisiä lukuja, kuvia tai vaikkapa sisällysluettelo. Verkkokäytössä tämä on erittäin kätevää, kirjan kaikki loogiset osat voidaan identifioida ja löytää - ja haluttaessa myös myydä - erikseen.

BICI:n syntaksi on lähes sama kuin SICI:n, sillä erotuksella että BICI:n alussa on teoksen ISBN ISSN:n asemesta. Innostus BICI:n käyttöön oli vähäistä, muun muassa koska BICI:n soveltaminen olisi edellyttänyt kustantajien ja kirjakauppojen tietojärjestelmien mittavaa remontointia. Mutta vielä hankalampi vaatimus oli julkaisujen rakenteistaminen siten, että BICI-tunnus olisi voitu generoida automaattisesti.

Kaikki edellä kuvattu olisi voitu tehdä. Mutta BICI oli teknisesti ongelmallinen koska se ei ollut yhteensopiva EAN-järjestelmän kanssa, eikä tunnuksella ollut takanaan kirja-alan poliittista tukea. Tarjolla oli myös oleellisesti helpompi ratkaisu: ISBN-tunnuksen soveltamisalan laajentaminen kirjojen osiin. Uusi ISBN-tunnus on käytännössä kuolinisku jo ennestään henkitoreissaan olleelle BICI:lle. Jos ISBN ei selviä uudesta tehtävästään (eli sen kapasiteetti loppuu kesken), kirjojen osien identifiointiin ryhdyttäneen ennen pitkää kehittämään vaihtoehtoista ratkaisua, mutta tähän menee vielä vuosia.

4. ISMN

Nuottijulkaisujen tunnistamiseen tarkoitettu ISMN (International Standard Music Number) -tunnus otettiin käyttöön 1990-luvun puolivälissä. Se vastaa kirjojen ISBN-tunnusta, ja on tarkoitettu apuvälineeksi nuottien kustantamiseen ja myyntiin sekä musiikkikirjastoimintaan ja muuhun nuottikirjallisuuden tiedonhakuun. Vuoden 2007 keväällä ISMN-järjestelmään kuului jo 52 maata, joista 20 on käynnistänyt toiminnan viimeisen viiden vuoden aikana. Jäsenmaiden määrä kasvaa edelleen nopeasti.

ISMN muistuttaa rakenteeltaan ISBN-tunnusta. Merkittävin ero on, että ISMN:ssä maantunnus on korvattu M-kirjaimella. Muuten tunnuksessa on samat elementit kuin ISBN:ssä eli kustantajantunnus, julkaisuntunnus ja tarkistusmerkki. Esimerkki:

ISMN M-321-76543-1

Kustantajatunnusten jakelu on koordinoitu kansainvälisesti. Kaikkiaan ISMN-järjestelmään on rekisteröity noin 15.000 musiikin kustantajaa, joista suomalaisia noin 120. ISMN-tunnuksia on jaettu Suomessa sekä painetuille että elektronisille nuottijulkaisuille.

ISBN-standardin uusimisen jälkeen myös ISMN-standardi päätettiin modernisoida samalla tavalla, vaikka välitöntä teknistä perustetta tälle toimenpiteelle ei ollutkaan. Julkaisuala katsoi ISMN-tunnuksen muuntamisen 13-merkkiseksi olevan tarpeen, koska molempia tunnuksia käyttävät osittain samat toimijat. ISMN-standardin uudistaminen on parhaillaan käynnissä ja sen odotetaan valmistuvan vuoden 2007 lopulla. 13-merkkiseen ISMN-tunnuksen käyttöön siirrytään luultavasti vuoden 2008 alkupuolella. ISMN-tunnus muuntuu 13-merkkiseksi helpommin kuin ISBN. Tunnuksen eteen lisätään etuliite 979, M-kirjain muutetaan numeroksi 0 ja tarkistusmerkki pysyy samana. Esimerkki:

ISMN 979-0-321-76543-1

Kuten ISBN-tunnuksen muutos, edellyttää ISMN:n muuttaminen 13-merkkiseksi julkaisualan järjestelmien mukauttamista 13-merkkisten tunnusten käyttöön.

5. ISSN

International Standard Serial Number -tunnusta käytetään jatkuvien julkaisujen, siis (painettujen ja elektronisten) sanoma- ja aikakauslehtien, vuosikirjojen, sarjojen yms. identifiointiin. ISSN:n rakenne on määritelty standardissa ISO 3297. Tunnus koostuu kirjaimista ISSN, tyhjämerkistä ja kahdesta neljän numeron ryhmästä, joita erottaa tavuviiva. Kahdeksas merkki on tarkistusmerkki, joka voi olla myös X. Esimerkki:

ISSN 0730-9295

Julkaisuissa numerotunnuksen eteen pitää aina merkitä kirjaimet ISSN luettavuuden parantamiseksi.

Kansainvälinen ISSN-keskus (<http://www.issn.org>) ylläpitää globaalia ISSN-tietokantaa ja koordinoi ISSN:n käyttöohjeiden kehittämistä. Se jakaa vapaita ISSN-tunnuksia muutaman tuhannen tunnuksen "blokkeina" kansallisille ISSN-keskuksille, jotka puolestaan huolehtivat ISSN-tunnusten jakelusta kustantajille. Suomessa ISSN-tunnuksen saa Suomen ISSN-keskuksesta (<http://www.kansalliskirjasto.fi/julkaisuuala/issn.html>), joka toimii Kansalliskirjastossa.

ISSN-tunnus on ISBN:ään verrattuna sikäli "tyhmä", että siitä ei mitenkään ilmene kausijulkaisun julkaisumaa tai kustantaja. ISSN-tunnuksia on jaettu vasta alle pari miljoonaa vielä ilmestyvien ja jo päättäneiden kausijulkaisujen identifiointiin. ISSN skaalautuu siis hyvin, sillä yli kahdeksan miljoonaa tunnusta on käytettävissä. Uusia tunnuksia jaetaan kausijulkaisuille noin 50.000 vuodessa; osa tästä käytetään vanhojen digitoitavien kausijulkaisujen identifiointiin. Kansainvälisen ISSN-keskuksen ylläpitämän ISSN-tietokannan viitemäärä ohitti miljoonan rajan vuodenvaihteessa 2002.

ISSN-tunnuksia on annettu verkkolehdlle jo kymmeniä tuhansia. Suomessa identifioituista lehdistä valtaosa näistä on verkkolehtiä, mutta joukossa on myös sarjajulkaisuja sekä jatkuvasti päivitettäviä kumulatiivisia verkkopalveluita. Oman ryhmänsä muodostavat Kansalliskirjaston Historialliseen sanomalehtikirjastoon digitoimat vanhimmat suomalaiset sanomalehdet.

Kansainvälisen ISSN-keskuksen ohjeiden mukaan saman lehden painetulle ja verkkoversiolle tulee antaa eri ISSN-tunnus. Verkkoversio ei kuitenkaan saa ISSN-tunnusta lainkaan, jos sen sisältö on hyvin niukka ja lähinnä esitteellinen (sisältönä esimerkiksi vain lehden yhteystiedot ja artikkelinäytteitä tai lyhennelmiä).

Verkkolehden eri tiedostomuodoissa julkaistuilla versioilla (esimerkiksi sama lehti HTML- ja PDF-muodossa) on sama ISSN-tunnus. Painetusta lehdestä digitoimalla valmistetussa, eli alkuperäisen julkaisun kanssa täysin identtisessä verkkoversiossa saa käyttää painetun julkaisun ISSN-tunnusta.

Vaikka verkkolehden eri versiot saavat saman ISSN-tunnuksen, ISSN-yhteisössä on koettu ongelmaksi saman lehden eri varianttien löytäminen. Uusi ISSN-standardi ratkaisee tämän pulman, mutta aiheuttaa päänvaivaa järjestelmien kehittäjille.

Vuonna 2008 käyttöön otettavan uuden ISSN-standardin suurin uutuus on Linking ISSN eli ISSN-L. Sen tarkoituksena on sitoa yhteen lehden eri manifestaatiot. Toisin sanoen lehden painetulla ja verkkoversiolla on sama ISSN-L, mutta eri ISSN. Uuteen ISBN-standardiin verrattuna uuden ISSN:n muutokset ovat vähäisiä; tunnuksen rakenne ei muutu, ja esimerkiksi metadattaa ja kansainvälisen keskuksen roolia koskevat määräykset ovat ennallaan. ISSN-L on kuitenkin suurempi muutos kuin miltä se äkikseltään näyttää.

Monimutkaiseksi ISSN-L –konseptin tekee se, että ISSN-L –tunnisteksi poimitaan lehden ensimmäisen manifestaation ISSN. Jokaisessa kausijulkaisutietueessa on siis jatkossa kaksi kenttää ISSN-tunnuksille; toiseen tallennetaan ISSN ja toiseen ISSN-L. Lehden ensimmäisellä julkaistulla versiolla sama ISSN tallennetaan kahteen paikkaan, mutta muilla lehden manifestaatioilla ISSN- ja ISSN-L –tunnukset eroavat toisistaan.

Järjestelmäkehittäjille ISSN & ISSN-L –parivaljakko on mielenkiintoinen haaste. MARC-formaatissa tunnukset on tallennettava eri kenttiin, ja kirjastojärjestelmässä ne on määriteltävä haettaviksi eri indekseistä ja erotettava toisistaan käyttöliittymässä jotta asiakas voisi hakea makunsa mukaan lehden tiettyä manifestaatiota tai niitä kaikkia.

Kirjastojärjestelmätoimittajat kykenevät epäilemättä toteuttamaan nämä muutokset. Mahdollisesti pystymme myös selvittämään asiakkaillemme, mistä on kyse. Ongelmia tulee silloin, kun bibliografinen tieto haetaan sellaisen järjestelmän kautta joka ei erota eri ISSN-tunnuksia toisistaan. Oletetaan että kausijulkaisun tiedot ovat haettavissa Googlasta tai Googlen kautta. Tällöin käytettävä indeksointi ja/tai hakuprotokolla ovat niin yksinkertaisia, että tieto saman ISSN-tunnuksen kaksoisroolista katoaa välttämättä. Riippuen siitä miten järjestelmät on koodattu toimimaan, asiakas saa haun tulokseksi joko lehden kaikki tai yhden manifestaation, riippumatta siitä mitä hän halusi.

ISSN-L –ongelma olisi voitu välttää kieltämällä ISSN-tunnuksen uusiokäyttö ISSN-L –tunnuksena. Suuret (tieteelliset) kustantajat olivat kuitenkin haluttomia tähän, mikä ei ole ihme: ne ovat vastustaneet jo uusien ISSN-tunnusten antamista eri manifestaatioillekin. Tulevaisuus näyttää, miten hyvin uusi ISSN-standardi soveltuu monimutkaistuvaan ohjelmistoympäristöömme.

Internet-käyttö

ISSN-tunnusten tallennus verkkolehtiin on suunniteltava huolella. ISBN identifioi aina periaatteessa vain yhden dokumentin, joka tosin saattaa verkossa jakautua muutamiksi tiedostoiksi. Sitä vastoin esimerkiksi haku ISSN-tunnuksella 0040-781X voisi pahimmassa tapauksessa, eli jos Time-lehden jokaisessa artikkelissa olisi lehden ISSN-tunnus, tuottaa tuloksena tuhansia sivuja lehden verkossa julkaistuja artikkeleita aikojen alusta nykypäivään asti. Yksittäisten artikkeleiden asemesta ISSN-haun pitäisi tuottaa tulokseksi vain elektronisen lehden kotisivu; toisin sanoen ISSN pitää tallentaa vain verkkolehden kotisivulle, ei artikkeleihin. Tämä onkin yleinen käytäntö muun muassa Suomessa.

Oikein sovellettuna ISSN on erittäin sopiva Internet-käyttöön. ISSN-tunnuksella on helppo hakea verkkolehtien kotisivuja Googlasta ja muista vastaavista verkon hakukoneista. Sen ainoa merkittävä puute on soveltumattomuus kausijulkaisujen numeroiden ja artikkeleiden identifiointiin. Tähän tarkoitukseen on kehitetty Serial Item and Contribution Identifier eli SICI-tunnus, josta lisää seuraavassa luvussa.

ISSN:n soveltamisessa on kuitenkin yksi mielenkiintoinen rajausongelma: milloin jokin resurssi on tulkittavissa verkkolehdeksi ja milloin on kyseessä jokin muu verkkojulkaisu? Toistaiseksi useimmat lehdet ilmestyvät verkossakin perinteiseen tapaan numeroina, mutta verkkolehdet ovat jo alkaneet tiedonvälityksen nopeuttamiseksi julkaista artikkeleita sitä mukaa kun tekijät saavat ne valmiiksi. Artikkelit on usein myös ryhmitelty esimerkiksi aihepiirin mukaan eikä numeroittain ja vuosikerroittain. Rajasta verkkolehden ja WWW-palvelun välillä on usein kuin veteen piirretty viiva.

ISSN:n soveltamisohjeissa tähän muutokseen on jo varauduttu siten, että ISSN:n voi antaa ns. jatkuville julkaisuille. Niiden määritelmä on mukana kausijulkaisujen uusituissa luettelointisäännöissä (ISBD(CR)) sekä ISSN-keskuksen omissa säännöissä. Jatkossa kansalliset ISSN-keskukset voivat jakaa ISSN-tunnuksia nykyistä laajemmin myös verkon resursseille jotka eivät ole perinteisiä kausijulkaisuja. Tunnuksen soveltamisalueen raja määrittynee käytännössä sen mukaan, miten paljon kansallisilla ISSN-keskuksilla on henkilökuntaa kustantajien neuvontaan ja verkon resurssien luettelointiin. Suomen ISSN-keskus on jo useamman vuoden ajan antanut ISSN-tunnuksia kaikenmuotoisille ”verkkolehdille”, siis myös artikkelitietokannoille ja sivustoille, joissa artikkelit on hajautettu sivuston eri osioihin. Valintakriteerit muotoutuvat sisällön, ei muodon mukaan.

Kansainvälinen ISSN-keskus modernisoi atk-järjestelmänsä vuosina 2001-2004. Tavoitteena oli ottaa käyttöön kaupallinen kirjastojärjestelmä niin vähin toiminnallisina muutoksina kuin mahdollista. Tässä onnistuttiin hyvin; vanhan kotitekoisen järjestelmän korvaava Virtua-järjestelmä palvelee ISSN-keskuksen tarpeita hyvin. Globaali ISSN-tietokanta on nyt paremmin kirjastojen ja muiden tiedontarvitsijoiden ulottuvilla tiedonhaku- ja kopioluettelointia varten. Tekninen tavoitettavuus – tässä tapauksessa Z39.50-standardin välityksellä – ei kuitenkaan yksinään riitä.

ISSN-tietokannan ylläpidon on muututtava nykyistä ajantasaisemmaksi, koska harvakseltaan – esimerkiksi neljä kertaa vuodessa – tehtävät eräpäivitykset eivät riitä kopioluetteloinnin tarpeisiin. Kahden kirjastojärjestelmän välillä tiedonsiirto voidaan periaatteessa automatisoida niin, että tietueet voidaan siirtää vaikkapa kerran päivässä. Käytännössä tämä ei ole mahdollista. Kansallisten ISSN-keskusten on sovittava kansainvälisen keskuksen kanssa kompromissiratkaisu, jonka avulla paikallinen työmäärä säilyy kohtuullisena mutta globaali ISSN-tietokanta palvelee käyttäjiään nykyistä paremmin.

6. SICI

Serial Item and Contribution Identifier (SICI; ANSI/NISO Z39.56-1996; <http://www.niso.org/standards/resources/Z39-56.pdf>) on tunnus, jonka avulla voidaan identifioida kausijulkaisujen numerot (serial item) ja artikkelit (serial contribution). SICI hyväksyttiin amerikkalaiseksi standardiksi jo vuonna 1991, ja sen toinen, uusittu versio ilmestyi 1996. SICIn modernisointia selvitettiin viimeksi vuonna 2002. Tuolloin todettiin, ettei uuden version laadintaan ole tarvetta.

Ted Koppel kuvaa SICIn rakennetta ja tunnuksen suunnittelun tavoitteita optimistisesti nimetyssä artikkelissaan ”The SICI emerges, cicada-like, after eight years of dormancy” (http://www.doi.org/topics/July_04_ISQ_online_version.pdf). Vuonna 2004 ilmestyneen artikkelin keskeinen sanoma on, että SICI-tunnuksesta on vähittäin tulossa suosittu, ja että se välttäisi BICIn kohtalon.

Ei ole varmaa, onko Koppel oikeassa. Hän oli yksi SICIn kehittäjistä ja on luonnollista, että hän uskoo tunnuksen käytön olevan yleistymässä. Päinvastaiseen suuntaan viittaa se, että SICI:lle on kehitetty kilpailijoita. Publisher Item Identifier (PII) on tieteellisten lehtien kustantajien kehittämä, SICIä yksinkertaisempi tunniste, jonka suosio on jäänyt vähäiseksi. PII ei sen vuoksi ole vakava kilpailija, mutta DOI on.

DOI-tunnuksen suhde SICIn on mielenkiintoinen. SICI-tunnukset voidaan esittää DOI-tunnisteina, mikä mahdollistaa SICI-tunnukseen pohjautuvan linkityspalvelun. Mutta toisaalta kustantajat voivat DOI-järjestelmässä identifioida artikkelinsa periaatteessa millä tahansa kotitekoisella tunnisteella, ja välttyä tällä tavoin SICI-tunnusten rakentamisen vaivalta. Tämä ei tietenkään ole vain SICI:n ongelma: DOI voi tunkeutua samalla tavoin kaikkien muidenkin perinteisten tunnusten reviirille.

SICI:n kannalta huono uutinen on, että sen ISO-standardointia ei ole vielä edes aloitettu. Tämä on vakava puute: SICI:n tulisi olla kansainvälinen, ei vain kansallinen standardi. NISO:n (National Information Standards Organization) ja muiden ISO:n yhteistyöorganisaatioiden kehittämiä standardeja voidaan hyväksyä ISO:ssa sellaisenaan niin sanotun Fast track –menettelyn avulla, joten standardointiprosessi sinänsä olisi yksinkertainen.

SICI koostuu ISSN:stä, kausijulkaisun numeroa koskevista tiedoista (Item segment), artikkelia koskevista tiedoista (Contribution segment) sekä tarkistetiedoista (Control segment). Mark Needlemanin artikkeli "Computing Resources for an Online Catalog - 10 years later", joka julkaistiin Information Technology and Libraries -lehden vuosikerran 11 numerossa 2 sivulta 168 alkaen kesäkuussa 1992 saa seuraavan SICI-koodin:

0730-9295(199206)11:2<168:CRFAOC>2.0.TX;2-#

Esimerkistä ilmenee havainnollisesti se, että SICI on monimutkainen tunnus, jonka luonti käsin on työlästä. Mutta periaatteessa jokainen artikkeli sisältää SICI:n edellyttämät tiedot. SICI-tunnuksen rakentamiseen soveltuvia ohjelmia on kehitetty; sen jälkeen kun pohjatiedot on syötetty tallennusalustalle, ohjelma generoi SICI:n. Vielä tehokkaampi ratkaisu on SICI-tunnuksen automaattinen generointi (riittävän) rakenteisesta elektronisesta artikkelista.

SICI-tunnuksen tulkitseminen on itse asiassa varsin helppoa. Esimerkkinä annettu SICI sisältää seuraavat tiedot:

0730-9295	Julkaisun ISSN
(1992069)	Julkaisuaikatiehto
11:2	Numerointitieto (volyymi ja numero)
168	Sijaintitieto, tavallisesti sivun numero. Alun perin elektronisille lehdille tätä tietoa ei voi antaa jos sivunumeroita ei ole. Jos elektroninen versio on digitoitu painetusta, sivunumeroita voidaan tietenkin käyttää.
CRFAOC	Nimekekoodi; muodostetaan nimekkeen kuuden ensimmäisen sanan (mukaan lukien myös artikkelit) ensimmäisistä kirjaimista
2	Code structure identifier. Arvot ovat 1 (lehden numeron tiedot) ja 2 (artikkelin tiedot)
0	Derivative part identifier. DPI:n avulla voidaan kertoa onko kyse artikkelista (0), sisällysluettelosta (1), hakemistosta (2) vai abstraktista (3).
TX	Medium/Format Identifier; kahden kirjaimen mittainen koodi jolla ilmaistaan dokumentin formaatti. TX viittaa perinteiseen artikkeliin. Verkossa ilmestyneen artikkelin tunnus on toistaiseksi aina CO (Online (remote)).
2	SICI-standardin versio; toistaiseksi aina 2
#	Kontrollinumero

Koska tunnukset perustuvat artikkeleiden tiedoille, kansallisia SICI-keskuksia ei tarvita, mutta ainakin periaatteessa kansallisten ISSN-keskusten tulisi kertoa kustantajille miten SICI-tunnuksia rakennetaan ja opastaa heitä SICI:n käytössä. Toistaiseksi tämäntyyppistä tiedotusta ei juuri ole tehty, mutta ISSN-keskusten kannattaisi valmistella julkaisijoille suunnattua SICI-ohjeistusta. Rakenteensa ansiosta SICI ei tarvitse keskitettyä tietokantaa toisin kuin ISSN tai vaikkapa PII joka ISSN:n tavoin on ”tyhmä” tunniste.

Nykyisessä SICI-standardissa on se osin ISSN:n soveltamistavasta aiheutuva puute, ettei artikkelin erilaisia elektronisia versioita voida erottaa toisistaan. Artikkelin HTML-, PDF- ja muut versiot saavat saman SICI-tunnuksen, mikä aiheuttaa ongelmia jos SICIä halutaan käyttää digitaalisessa

arkistossa artikkeleiden identifiointitunnisteena. SICI:n mahdollisessa uudessa versiossa Medium/Format Identifier-koodin käyttöä on laajennettava niin, että artikkeleiden erilaisille elektronisille versioille voidaan antaa eri SICI. Tämä edellyttää vähintään dokumenttiformaattien erottelumahdollisuutta; suotavaa olisi jos lisäksi olisi mahdollista määritellä joitakin digitaalikuvienv ominaisuuksia. Sama artikkeli voidaan skannata esimerkiksi TIFF-muotoon tarkkuuksilla 75 ja 300 dpi; nämä näyttö- ja painokelpoiset versiot pitäisi voida erottaa toisistaan.

Internet-käyttö

SICI-tunnusta on käytetty kaupallisten verkkolehtien artikkeleiden identifiointiin DOI-tunnusjärjestelmän osana. SICI skaalautuu erittäin hyvin kattamaan kaiken verkossa ilmestyvän artikkeliaineiston, ainakin niin kauan kun ISSN-tunnuksessa on kasvunvaraa. Jakelun organisoinnin kannalta on tärkeää, että SICI-tunnuksia ei tarvitse pyytää kansalliselta ISSN- tai SICI-keskukselta, vaan kustantaja tai periaatteessa kuka tahansa voi generoida tarvitsemansa SICI:t itse.

SICI:lle ei ole määritelty omaa nimialuetta URN-järjestelmään, mistä huolimatta SICIä on käytetty myös URN-järjestelmässä. Laajan merkkivalikoimansa vuoksi SICI vaatii editointia sekä URN- että DOI-järjestelmässä.

Lisätietoa SICIstä ja sen verkkokäytöstä saa muun muassa Wikipediasta, jonka SICI-artikkeli (<http://en.wikipedia.org/wiki/SICI>) on varsin laadukas. Se sisältää muun muassa esimerkin siitä, miltä URN-tunnisteeksi muokattu SICI näyttäisi:

SICI

1046-8188(199501)13:1<69:FTTHBI>2.0.TX;2-4

URN

URN:SICI:1046-8188(199501)13:1%3C69:FTTHBI%3E2.0.TX;2-4

7. NBN

Kansallisbibliografian ID-tunnuksia eli NBN-tunnuksia käytetään sellaisen (perinteisesti kansallisbibliografiaan luetteloidun) aineiston identifiointiin, jolle ei voida antaa muuta tunnusta. NBN:n syntaksista ja käytöstä päättää jokainen kansalliskirjasto itse; ne myös huolehtivat siitä että tunnukset ovat pysyviä ja ainutkertaisia.

Tunnuksen pysyvyys ei ole vain tekninen ominaisuus; itse asiassa on tärkeämpää, että tunnuksia jakava organisaatio on pitkäikäinen ja että sillä on pysyvä ja mielellään lakisääteinen vastuu tallentaa identifioidut julkaisut. Tästä syystä esimerkiksi yritykset tai yliopistojen atk-keskukset eivät välttämättä sovellu julkaisujen tunnisteiden jakelun vastuuorganisaatioiksi.

NBN-tunnus rakentui Suomessa vanhastaan f-kirjaimesta, vuosiluvusta (kahden tai neljän numeron tarkkuudella) ja juoksevasta numerosta. F-kirjainta voidaan täydentää muilla kirjaimilla; "fe" kertoo, että kyseessä on elektronisen julkaisun NBN. Suomen ja muiden maiden NBN-numeroilla ei välttämättä ole mitään muuta yhteistä kuin tunnusten uniikkisuus kansallisella (ja URN-tunnisteena esitettynä kansainvälisellä) tasolla.

Eri kansalliskirjastot voivat antaa julkaisuille identtisiä NBN-tunnuksia, koska mitään yhteisiä tunnusten jakeluperiaatteita ei ole. Internet-käyttöä varten tunnuksiin on lisättävä ainutkertaisuuden

varmistava koodi eli esimerkiksi maantunnus. Tämä on otettu huomioon Internet-standardissa, joka määrittelee NBN-tunnusten käytön URN-järjestelmässä (<http://www.ietf.org/rfc/rfc3188.txt>).

Internet-käyttö

NBN-tunnusten jakelu samoin kuin tunnisteiden jakelu yleensä on perinteisesti ollut manuaalista ja kansalliskirjaston henkilökunnan vastuulla. Manuaalinen jakelu ei ole toimiva vaihtoehto silloin, kun tunnuksia tarvitaan paljon.

Useilla kansalliskirjastoilla digitaalisiin arkistoihin tallennettavien elektronisten resurssien identifiointi perustuu URN-tunnuksena käytettävään NBN:ään, joka luodaan automaattisesti.

Suomen verkkoarkisto rakennetaan ”haravointiohjelman” avulla. Arkisto sisältää jo kymmeniä miljoonia dokumentteja, joita ei voida käsitellä manuaalisesti. Sen vuoksi verkkoarkistossa NBN-tunnuksena käytetään tallennettavista tiedostoista laskettuja MD5-tarkistussummia, joiden eteen lisätään etuliite ”fea”. Laskennan hoitaa sovellus, jolla aineisto haravoidaan arkistoon. MD5 soveltuu ID-tunnukseksi hyvin, koska epäidenttiset julkaisut eivät periaatteessa milloinkaan saa samaa tarkistussummaa. Verkkoarkistossa tunnistetta voidaan käyttää paitsi dokumentin uniikkina hakuavaimena myös tuplakontrollin välineenä: jos kahdella tiedostolla on sama MD5, on käytännössä varmaa että tiedostot ovat identtisiä.

Saksassa rakennettiin 2000-luvun alussa Carmen-projektissa järjestelmä, jossa yli 70 yliopiston julkaisemat elektroniset väitöskirjat tallennetaan kansalliskirjaston digitaaliseen arkistoon. Tallennuksen yhteydessä väittelijät tai heidän yhdyshenkilönsä ao. kirjastossa tallentavat julkaisulle metatiedot, mukaan lukien NBN-tunnuksen. URN-linkityspalvelu ja siihen liittyvää dokumentaatiota on saatavilla osoitteesta <http://www.persistent-identifier.de/>.

Norjan ja Hollannin kansalliskirjastot käyttävät NBN-pohjaista URN-tunnusta kaikkien arkistointiensa elektronisten resurssien identifiointiin. Jaettuja tunnuksia on pelkästään Hollannissa noin 10 miljoonaa. Norjassa dokumenttien määrä nousee satoihin miljooniin sitä mukaa kun norjalaisen julkaistun kulttuuriperinnön digitointi etenee. Muutamassa vuodessa NBN:stä tulee maailman yleisin (elektronisten) julkaisujen tunniste.

Ruotsissa NBN-pohjaisten URN-tunnusten jakelu käynnistettiin osana Uppsalan yliopiston koordinoimaa Diva-projektia, jota hankkeen vetäjä Eva Muller on kuvannut DLIB-lehdessä julkaistussa artikkelissa (<http://www.dlib.org/dlib/november03/muller/11muller.html>). Uppsalassa kehitetty URN-linkityspalveluohjelmisto valittiin sittemmin kansalliskirjaston ylläpitämän Suomen kansallisen linkityspalvelun pohjaksi, koska arvioimme Diva-sovelluksen saatavilla olevista avoimen lähdekoodin linkityspalveluohjelmistoista parhaaksi.

NBN-tunnusten käyttö ei Suomessa ole vielä yhtä laajaa kuin esimerkiksi Saksassa, mutta sovelluskohteita on verkkoarkiston lisäksi muitakin. Kansalliskirjasto on jakanut omasta NBN-alueestaan useille virastoille omia NBN-blokkeja. Vähäisempiin tunnistetarpeisiin alettiin varautua jo vuosia sitten. Kansalliskirjasto rakensi 2000-luvun alussa yhdessä Lundin yliopiston kirjaston sittemmin edesmenneen Netlab-yksikön kanssa URN-generaattorin, joka rakentaa kansallisbibliografian ID-numeroon perustuvia URN-tunnuksia.

URN-generaattorin ansiosta periaatteessa kuka tahansa voi Suomessa identifioida oman (verkko)julkaisunsa. Tästä syystä on tarpeen opastaa ”suurelle yleisölle” julkaisujen identifiointin periaatteita sekä antaa perustiedot siitä, miten tunnuksia voidaan upottaa julkaisuihin.

Kansantajuista opastusta (verkko)julkaisujen identifioinnista ei juuri ole tarjolla (tämäkin teksti on tarkoitettu kirjastonhoitajille ja muille ammattilaisille); URN-generaattorin käyttöohje ja mahdolliset palvelukohtaiset opasteet antavat kuitenkin suuntaa-antavaa opastusta ja ohjaavat tunnisteiden tallentamista julkaisuihin.

Käytännön esimerkki URN-generaattoria käyttävästä sovelluksesta on Helsingin yliopiston E-thesis-palvelu (<http://ethesis.helsinki.fi/>). Sen sisältämällä väitöskirjoilla on ISBN (URN-tunnukseksi tallennettuna), mutta graduille annetaan NBN. Esimerkiksi Jaakko Jäätmaan pro gradu ”Taustamusiikki markkinoinnin työvälineenä suomalaisissa tavarataloissa” omaa NBN-tunnuksen fe20071482, ja löytyy verkosta pysyvästi osoitteesta <http://urn.fi/URN:NBN:fi-fe20071482>, olipa gradun URL mikä tahansa. URN-tunnuksen edessä oleva kansallisen linkityspalvelun osoite (<http://urn.fi/>) pitää tallentaa niin kauan kuin verkossa ei ole linkityspalveluiden paikallistamismahdollisuutta eivätkä WWW-selaimet osaa käsitellä URN-tunnuksia oikein.

NBN-tunnuksen yksinkertaisen rakenteen vuoksi sen jakelu joko sellaisenaan tai URN-tunnuksena on helppo hajauttaa ja automatisoida. Tunnuksia voidaan jakaa periaatteessa rajattomasti; mitään teknisiä tai taloudellisia rajoitteita ei ole. Esimerkiksi verkkoarkiston kymmenille miljoonille dokumenteille saadaan tunnukset ilman vähintäkään ihmistyötä käyttäen edellä kuvattua tekniikkaa. Pro gradu –toille tekijöiden itsensä on tehtävä kuvailu ja haettava samalla tunniste URN-generaattorilta.

Verkojulkaisujen identifioinnissa NBN:llä on tärkeä rooli. Monille verkon julkaisuille ja muille aineistoille (esim. huomattava osa kotisivuista, keskustelulistat) ei ole mahdollista eikä edes järkevää antaa perinteistä tunnistetta. Ne eivät yksinkertaisesti kata verkossa olevaa aineistoa; kotisivulle ei voi antaa sen paremmin ISBN:ää kuin ISSN-tunnustakaan, vaikka jälkimmäisen käyttöä sääntöjen sen voisivatkin sallia. ISSN-tunnuksia on vielä jäljellä useita miljoonia, mutta kotisivuja on jo nyt luultavasti useita satoja miljoonia.

Joissakin tilanteissa kahdella eri dokumentilla/tiedostolla voi hyvin olla sama identifiointitunnus. Esimerkiksi elektronisen lehden kotisivun layout voi ajan mittaan vaihdella moneen otteeseen, mutta kaikilla varianteilla on sääntöjen mukaan sama ISSN - kunhan lehden nimi ei muutu. Lisäksi ISSN on voitu tallentaa myös artikkeleihin tai lehden numeroiden yhteyteen. Digitaalisessa arkistossa lehden kotisivun eri variantit on kuitenkin erotettava toisistaan, ja käytännössä tämä on tehtävä koneellisin menetelmin kuten MD5-tarkistussumman avulla. Verkoarkistossa perinteiset identifiointitunnisteet toimivat ensisijaisesti vain tiedonhaussa; identifiointi ja halutun version haku tehdään tarkistussummaan perustuvalla URN/NBN-tunnuksella.

Verkojulkaisujen luetteloinnissa pitäisi siirtyä URL-linkityksestä URN-tunnusten käyttöön. Jos julkaisun sijainti verkossa muuttuu tai se tulee saataville uusista osoitteista, perinteisessä toimintamallissa URL pitää korjata bibliografisissa tiedoissa. Tämä on ongelma silloin, kun tietue on kopioitu moneen paikkaan. Esimerkiksi elektronisen väitöskirjan tiedot voivat löytyä usean kirjaston näyttöluettelosta sekä monista julkaisuarkistoista. Linkkitiedon päivittäminen näihin kaikkiin on iso työ. Helpommalla päästään, jos viitteissä on vain URN, ja sijaintitiedot löytyvät linkityspalvelussa ylläpidetystä taulukosta, johon viedään URN ja sitä vastaavat URL-tunnukset. Muutos pitää tehdä vain yhteen paikkaan, riippumatta siitä, miten monen kirjaston näyttöluettelossa, julkaisuarkistossa tai muissa hakujärjestelmissä ao. julkaisun tiedot ovat.

8. Teosten tunnisteet

IFLA:n Functional Requirements for Bibliographic Records (FRBR) –mallin (<http://www.ifla.org/VII/s13/frbr/frbr.pdf>) mukaan voimme eritellä:

- Teokset, esimerkiksi Väinö Linnan Tuntemattoman sotilaan, kirjan pohjalta luodun näytelmän ja kirjan kaksi elokuvaversiota.
- Ekspressiot, esimerkiksi Tuntemattoman sotilaan käännökset eri kielille sekä kirjan kuvitettu laitos
- Manifestaatiot, esimerkiksi Tuntemattoman sotilaan suomenkielisen alkuperäisteoksen eri painokset sidottuina tai nidottuina.
- Niteet, esimerkiksi Tuntemattoman sotilaan ensimmäisen painoksen bibliofiilien kirjajhylyihin tallennetut kappaleet.

Keskeinen ongelma koko tämän järjestelmän kannalta on teoksen määrittely. Kirjastoalan valitsema lähtökohta on teoreettisesti varsin kaunis, ja sitä täydentävät FRBR-julkaisun seikkaperäiset ohjeet siitä, milloin kyse on teoksesta, milloin ekspressiosta. Esimerkiksi alkuperäisteokselle uskollinen käännös on ekspressio, mutta luova käännös on teos; siten Saarikosken paikoin varsin omaperäinen tulkinta Joycen Odysseuksesta on taatusti teos, mutta Mannin Doktor Faustus –teoksen kaunis mutta siitä huolimatta uskollinen käännös ekspressio.

ISON kehittämät teosten tunnisteet identifioivat teoksen tai sen ekspression; esimerkiksi Väinö Linnan romaani Tuntematon sotilas ja sen ruotsinkielinen käännös saavat aikanaan kumpainkin oman ISTC (International Standard Text Work Code) -tunnuksen. ISTC:n ja muiden teostietojen avulla on helppo löytää kaikki kyseisen teoksen/ekspression manifestaatiot, sitten aikanaan kun tarvittavat muutokset on tehty kirjastojen järjestelmiin. OCLC:n saamien tutkimustulosten mukaan (lisätietoja osoitteesta <http://www.oclc.org/research/projects/frbr/>) teostason tietueet voidaan luoda olemassa olevista manifestaatioiden tietueista. Vaikeuksia tuottavat esimerkiksi raamatun kaltaiset teokset, joista on olemassa huikea määrä manifestaatioita ja ekspressioita.

Teostason tietojen luonti sinänsä ei ole itsetarkoitus. Kirjastojen ja niiden käyttäjien kannalta mielenkiintoista on, että FRBR-mallin toteutuksella voidaan rikastaa näyttöluetteloita ja kehittää täysin uudenlaisia palveluita. Mutta jos teosten tiedot luodaan, tarvitsemme myös teosten tunnisteet.

Tätä kirjoitettaessa AV-teoksille tarkoitettu International Standard Audiovisual Number (ISAN; <http://www.isan.org/>) ja musiikkiteosten ISWC (International Standard Musical Work Code; <http://www.iswc.org/iswc/en/html/home.html>) ovat valmistuneet ja niitä on sovellettu jo vuosien ajan. ISWC-jakelusta vastaa Suomessa Teosto; omaa ISAN-toimistoa ei ole vielä perustettu.

Kirjastojen kannalta ISAN- ja ISWC-tunnuksia tärkeämpi on ISTC eli tekstimuotoisten teosten tunniste. Sen kehittäminen eteni Committee Draftiin eli ensimmäiseen julkiseen luonnokseen jo vuoden 2001 lopulla, ja kommenttikierros saatiin läpi keväällä 2002. Draft International Standard eli DIS-versio valmistui kommenttien pohjalta toukokuussa 2002, ja teksti lähti kommentoitavaksi heinäkuussa 2002. Tulos oli mainio: yksimielinen hyväksyntä.

Sisällöstä oli tässä vaiheessa päästy yksimielisyyteen, mutta kansainvälisen ISTC-keskuksen vastuuorganisaation nimeämisestä tuli ongelma. Loppusuoralla vastakkain olivat OCLC ja konsortio, jossa olivat mukana säveltäjien ja kirjailijoiden järjestöjen kattojärjestö CISAC sekä Nielsen Book Data ja Bowker. Jälkimmäiset voittivat kisan niukasti, vain huomatakseen että

standardin edellyttämän globaalin ISTC-tietokannan kehittämis- ja ylläpitokustannukset olivat liian korkeat.

Standardointiprosessi pysähtyi tähän havaintoon pitkäksi aikaa. Vasta 2007 keväällä tehdyt uudet laskelmat ja tekijänoikeusjärjestöjen kattojärjestön IFRROn (<http://www.ifrro.org/>) liittyminen konsortioon ovat ilmeisesti ratkaisseet ongelmat, ja standardointiprosessi ja ISTC:n käytännön implementointi voivat edetä.

Edellä mainittujen kolmen tunnusjärjestelmän seuraksi on tarkoitus rakentaa vielä still-kuville tarkoitettu identifikaatiojärjestelmä, mutta toistaiseksi sen työstämistä ei ole vielä edes aloitettu. Ainakin tähän asti teosten tunnusten kehitystyölle on ollut luonteenomaista se, etteivät eri työryhmät ole pyrkinet terminologian, tunnusten syntaksin tai tunnusten käyttösääntöjen harmonisointiin, vaan kaikki ovat soveltaneet oman alansa vanhoja käytäntöjä. ISANia, ISWC:tä ja ISTC:tä vertaillen lopputulos voi näyttää varsin sekavalta, ja sitä se onkin. Tästä ei välttämättä aiheudu ongelmia, koska standardit ovat eri yhteisöjen käytössä. Toiminnallisesti tunnusten erilaisuudesta ei liene mitään haittaa.

Eri alojen erilaisuutta osoittaa hyvin se, ettei AV-aineistolle ole olemassa kattavaa manifestaatioiden identifiointitunnusta. Ongelman poistamiseksi ISANille alettiin 2002 kehittää versiot (manifestaatiot) kattavaa laajennusta, V-ISAN. Sillä voidaan identifioida vaikkapa elokuvasta tehtävät eri manifestaatiot, kuten DVD- ja VHS-tuotteet, sekä niiden erikieliset variantit. Vuoteen 2006 mennessä V-ISAN -tunnuksesta oli saatu valmiiksi yksimielisesti hyväksytty DIS-versio. Lisätietoja tunnuksesta saa osoitteesta <http://www.collectionscanada.ca/iso/tc46sc9/v-isan.htm>.

ISTC

ISTC-tunnukseen tulee DIS-standardiluonnoksen eli käytännössä lopullisen tekstin mukaan 16 heksadesimaalimerkkiä. Tunnus koostuu neljästä osasta: kolmen merkin mittaisesta tunnuksen antajan tunnisteesta, vuosiluvusta (neljä merkkiä), kahdeksan merkin mittaisesta teoksen tunnisteesta sekä tarkistusluvusta. Järjestelmän kapasiteetti on valtava; ISTC-tunnuksien jakelukeskuksia voi olla 4096, ja jokainen niistä voi antaa vuosittain miljardi tunnusta vuoteen 9999 saakka.

ISTC:n kehittämisessä yksi kompastuskivi oli ainakin kirjastoalan edustajille kustantajien FRBR-mallista poikkeava perusideologia ja sen käytännön ilmentymänä kirjastoalan käytänteistä poikkeava terminologia. Esimerkiksi teoksen määrittelyssä kustantajien edustajien perusasenne oli hyvin pragmaattinen: jokainen myytävissä oleva itsenäinen teos on voitava identifioida erikseen. Kootut teokset ovat siten aina omia itsenäisiä teoksiaan, kuten teoksesta pilkotut osakohteet, vaikkapa kirjan erikseen ostettavissa olevat luvut.

Ekspressio merkitsee kustantajille teoksen luontiprosessia, minä vuoksi ISTC-standardista oli poistettava kaikki viittaukset ekspressio-käsitteeseen. Standardia soveltavien kirjastonhoitajien on kuitenkin muistettava että teostunnukset on annettava myös ekspressioille. Samaa "perheeseen" kuuluvien teosten ja ekspressioiden linkittäminen toisiinsa on onneksi nykytekniikalla varsin helppoa; jos tietää Tuntemattoman sotilaan alkuperäisteoksen ISTC:n, sen metatiedoista löytää ISTC-pohjaiset linkit teoksen käännöksiin, erilaisiin kuvitettuihin laitoksiin sekä itsenäisiin rinnakkaisteoksiin. Kuten ISSN-järjestelmässä, jokainen teos jolle annetaan ISTC on myös kuvailtava.

Formaatiksi on valittu teoksia varten kehitteillä oleva ONIX-variantti. Siitä tiedetään varmasti ainakin se, että bibliografiset tiedot voidaan kohtuullisen hyvin konvertoida MARC-muotoon. ONIX-formaatin määrittelemät tuotetiedot ovat vaikeammin konvertoitavissa.

ISTC-järjestelmää tarvitaan, koska teosten ekspressioiden ja manifestaatioiden määrä on nopeasti kasvamassa elektronisen julkaisemisen vuoksi. Kun pahaa aavistamaton käyttäjä etsii vaikkapa Tuntematonta sotilasta Fennicasta, hän saa melkoisen määrän viitteitä – yli 70. Jos pystymme esittämään näyttöluettelossa aluksi vain teokset (ja niiden ekspressiot) ja vasta tämän jälkeen manifestaatiot erikseen kustakin teoksesta / ekspressiosta, on kokonaisuus käyttäjien kannalta helpommin hahmotettavissa. ISTC-linkkien avulla voidaan helpommin tarjota käyttäjille tietoa teosten välisistä linkeistä; esimerkiksi Shakespearen Romeo ja Julia –näytelmän teostason tiedoista voidaan luoda periaatteessa rajattomasti linkejä niihin teoksiin, jotka ovat tavalla tai toisella hakeneet inspiraatiota Shakespearen työstä.

Arvelin tämän tekstin aiemmassa versiossa että kansainvälisellä tasolla ISTC-jakelutoiminta alkaa vuoden 2003 alkupuolella. Tästä aikataulusta ollaan nyt jäljessä viisi vuotta. Parhaassa tapauksessa ISTC-standardi saadaan valmiiksi 2008, ja sen tekninen infrastruktuuri julkistetaan pian tämän jälkeen.

Toistaiseksi ei ole mitään päätöstä siitä, mihin Suomen ISTC-keskus tai keskuksat perustetaan. Mutta on todennäköistä, että yksi keskus sijoittuu kansalliskirjastoon, koska kansalliskirjasto vastaa tekstimuotoisen aineiston manifestaatioiden tunnisteista, eli ISBN:stä, ISSN:stä ja ennen pitkää mahdollisesti myös SICI:stä. Jos ja kun ISTC-keskuksen toiminta käynnistyy, mielenkiintoisin tehtävä on takautuvien ISTC-tunnusten ja niiden metadatan generointi (minkä standardi yksiselitteisesti sallii, ja mikä on kustantajien ja muiden kirjastojen järjestelmien käyttäjien vuoksi tarpeen).

Teostietojen generoinnin ongelmaa on tutkittu intensiivisesti OCLC:ssä; heidän käsityksensä on, että teosten tiedot voidaan rakentaa suuremmista ongelmista olemassa olevista manifestaatioiden tiedoista. Tämä näkemys on sittemmin todennettu rakentamalla teostason tietueet Fennica-tietueista, jotka toimitettiin OCLC:lle WorldCat:iin tallennettavaksi vuonna 2006.

Ekspressioiden tietojen luonti voi olla vaikeata. Samaan yleiskäsitykseen, pessimistisemmin yleispainotuksin, päätyivät Eeva Murtomaa ja Knut Hegna omassa selvityksessään, jossa tutkittiin IFLA:n FRBR-mallin toteutuksen edellyttämää teosten kuvailutietojen luontia.

Tutkimuksen rinnalla ja sen tulosten rohkaisemana on jo kehitetty ainakin yksi kirjastojärjestelmä, joka tukee teosten luettelointia ja hakua. VTLS:n Virtuassa on ollut tämä ominaisuus jo muutamia vuosia. Valitettavasti muut kirjastojärjestelmät eivät ole seuranneet esimerkkiä, lukuun ottamatta WorldCatia. Koska muun muassa Kongressin kirjasto on määritellyt teosten kuvailun ja ISTC-järjestelmän yhdeksi keskeisistä tavoitteistaan, oletin muutamia vuosia sitten että muutkin kirjastojärjestelmien toimittajat tulevat uusimaan järjestelmänsä teoksien kuvailua silmällä pitäen. Nykykäsitykseni on, että järjestelmätoimittajat ovat haluttomia panostamaan paljon perinteisen kirjastojärjestelmän modernisointiin. FRBR-toteutukset tullevat rajoittumaan käyttöliittymään, missä manifestaatioiden tiedoista voidaan lennossa rakentaa teosten tiedot. Koska niitä ei kuitenkaan tallenneta pysyvästi minnekään, ISTC-järjestelmän pohjaksi tämäntyyppinen ohjelmisto ei kuitenkaan kelpaa.

9. URN

Uniform Resource Name -tunnusten avulla voidaan identifioida pysyvästi Internet-verkosta löytyvät dokumentit. URN-järjestelmän määritteli Internet Engineering Task Force:n (<http://www.ietf.org/>) URN Working Group (<http://www.ietf.org/html.charters/urn-charter.html>). Ryhmän piti saada työnsä valmiiksi jo vuonna 2000 tai viimeistään vuoden 2001 alussa, mutta prosessi venyi kesäkuuhun 2002 asti, koska kaikkiin URN-järjestelmän teknisiä perusteita koskeviin Internet-standardeihin tarvittiin Internet Engineering Steering Boardin hyväksyntä.

URN-pohjaisia tuotantosovelluksia on ollut käytössä jo 2000-luvun alkupuolelta lähtien, koska tunnuksen syntaksi on ollut tiedossa jo vuodesta 1997. Toimiva sovellus voidaan rakentaa käyttämättä globaalia URN-linkityspalvelua – onneksi, sillä mitään koko verkon kattavaa palvelua ei ole toistaiseksi rakennettu.

URN Working Groupin standardointiurakka oli melkoinen, koska työryhmä on määritellyt paitsi URN:n syntaksin, myös menetelmät joilla URN-pohjainen linkitys nykyisessä Internetissä tapahtuu, sekä palvelut joita URN-järjestelmä tarjoaa. Standardoinnin tavoitteena oli, että URN-tunnuksen on oltava ”immuuni” Internetin teknisen infrastruktuurin muutoksille. Tässä onnistuttiin: URN-tunnukset sinänsä eivät tee mitään oletuksia verkon toiminnasta tai linkityspalvelujen sijainnista. Tässä suhteessa periaatteellinen ero DOI-tunnukseen on merkittävä, koska DOI-tunnuksen rakenne määrittyy sen taustalla olevasta Handle-palvelusta (<http://www.handle.net/>).

URN-tunnuksen rakenne määritellään Internet-standardissa RFC2141 (<http://www.ietf.org/rfc/rfc2141.txt>). Sen mukaan URN koostuu kolmesta osasta:

- Kiinteä osa "URN:". Jokainen URN-tunnus alkaa näin; tarkoituksena on helpottaa URN-tunnuksen löytämistä rakenteettomista dokumenteista. Täysin luotettavaksi URN-tunnusten tunnistusta ei tällä keinolla tietenkään saada, koska jossakin dokumentissa voi olla muiden julkaisujen URN-tunnuksia ja ASCII-tekstissä ei ole mitään keinoa osoittaa mikä URN kuuluu juuri asianomaiselle julkaisulle.
- Nimiavaruus (NID eli Namespace Identifier). Koodi joka identifioi URN-tunnuksena käytetyn tunnusjärjestelmän.
- Identifioiva osa (NSS eli Namespace Specific String). Julkaisun varsinainen tunnus kuten ISBN sijoitetaan tähän osaan.

Internet-standardi RFC2141 määrittelee myös sen, mitä merkkejä URN-tunnus saa sisältää. Lähtökohtana on URL-sijaintitietojen rakenne ja merkistö. Valitettavasti standardi on tältä osin vanhentunut. Esimerkiksi skandinaavisia merkkejä ei saa käyttää, vaan ne on koodattava vastaaviksi heksadesimaaliluvuiksi. RFC2141 tulisikin pikaisesti päivittää viimeisimmän URI-tunnisteiden rakenteen määrittelemän Internet-standardin, RFC 3986:n mukaiseksi. (<http://tools.ietf.org/html/rfc3986>).

RFC2141:n vanhentuneisuus ei onneksi ole useimmille tunnisteille ongelma. Esimerkiksi ISBN ja ISSN ovat selviä tapauksia, kuten Suomessa käytetty kansallisbibliografian ID-numerokin. SICI vaatii merkkimuunnoksia vastaisuudessaakin, kaikenlaiset sulkeet kun ovat kiellettyjen merkkien listalla sekä RFC3986:ssa että niissä Internet-standardeissa, jotka se korvasi.

Internet-standardissa RFC2611 (<http://www.ietf.org/rfc/rfc2611.txt>) määritellään prosessi, jolla nimiavaruuksien rekisteröinti tapahtuu. Vastuuorganisaatio on IANA (Internet Assigned Names Authority). Käyttäen RFC2611:ssä määriteltyä menetelmää, on kansallisbibliografian ID-numerolle

varattu nimiavaruus NBN. Varaus on virallistettu Internet-standardissa RFC 3188 (<http://www.ietf.org/rfc/rfc3188>). Sen mukaan NBN-nimiavaruus on hajautettu kansalliselle tasolle varaamalla ISO:n maakoodit tähän tarkoitukseen (esimerkiksi urn:nbn:fi:). Lisäksi jokainen kansalliskirjasto voi jakaa oman "siivunsa" edelleen pienempiin osiin sillä edellytyksellä, että pitää niistä itse kirjaa. Kansalliskirjasto on antanut omia osa-alueita muun muassa useille virastoille, jotka julkaisevat aineistoa verkossa.

Tapa jolla URN-järjestelmä käyttää hyväkseen Internetin nimipalvelua eli Domain Name Service -järjestelmää, on määritelty joukossa Internet-standardeja. Näiden standardien käytännön merkitys on jäänyt vähäiseksi, koska niitä ei ole implementoitu nimipalvelusovelluksissa. Sen vuoksi URN-tunnukset on toistaiseksi tallennettava siten, että ne toimivat hyperlinkkeinä. Vastaavalla tavalla menetellään DOI-tunnusten kanssa.

Esimerkiksi osoitteessa <http://www.lib.helsinki.fi/tietolinja/0206/kaukopalvelu.html> olevan Tietolinjan artikkelin URN-tunnus on URN:NBN:fi-fe20061586. URN-tunnuksesta muodostettu hyperlinkki on silloin <http://urn.fi/URN:NBN:fi-fe20061586>, jossa alun <http://urn.fi/> on Kansalliskirjaston ylläpitämän kansallisen URN-linkityspalvelun osoite.

Kirjastoalalla käytettyjen identifikaatiojärjestelmien soveltuvuutta URN-tunnuksiksi on käsitelty alustavasti Internet-standardissa RFC2288 (<http://www.ietf.org/rfc/rfc2288.txt>). Mitään erityisiä ongelmia ei tässä selvityksessä havaittu; dokumentin laatijoiden motiivi onkin ollut ensisijaisesti Internet-yhteisön informoiminen bibliografisten tunnistejärjestelmien luonteesta ja käyttömahdollisuuksista ja toisaalta URN-järjestelmän tekeminen tunnetuksi kirjastoissa. Tarkempi analyysi - jota on tarvittu esimerkiksi ISBN- ja ISSN-nimialueiden rekisteröinnin yhteydessä - on paljastanut mielenkiintoisia samankaltaisuuksia ja eroja eri järjestelmien välillä.

URN-tunnusten generointi olemassa olevista tunnisteista on URN-syntaksin ansiosta erinomaisen helppoa. Muutamia esimerkkejä:

URN:NBN:fi-fe19985628 (kansallisbibliografian ID-numero)

URN:ISBN:9138202107 (ISBN)

URN:ISSN:1560-1560 (ISSN)

Kyse on siis vain kiinteän prefiksin lisäämisestä olemassa olevaan tunnukseseen (silloin kun tunnuksessa ei ole erikoismerkkejä jotka on konvertoitava). URN-tunnuksia ei tarvitse edes rakentaa valmiiksi; riittää että hakujärjestelmä tietää mikä URN-prefiksi kullakin tunnisteella on. Esimerkiksi DOI:n osalta tilanne on monimutkaisempi, koska DOI-prefiksi riippuu siitä, kuka tunnuksen antaa, ja lisäksi tunnukset eivät ole maksuttomia.

URN-tunnus ei mene päällekkäin nykyisten järjestelmien kanssa, koska URN-tunnisteen identifioivaa osaa ei saa "vetää hatusta". Jos identifioitava julkaisu on kirja, on käytettävä joko ISBN- tai NBN-nimiavaruutta, riippuen siitä onko kirjalle määritelty ISBN vai ei. Jos kumpaakaan tunnistetta ei ole, jompikumpi pitää hakea. Tältä osin DOI poikkeaa URN-tunnuksesta täysin: kustantaja voi soveltaa DOI-järjestelmässä sisäisiä tunnisteita, joiden soveltamisalat voivat mennä päällekkäin perinteisten tunnisteiden kanssa.

URN-järjestelmän tarjoamat palvelut on määritelty Internet-standardissa RFC 2483 (<ftp://ftp.funet.fi/pub/doc/rfc/rfc2483.txt>). Käyttäjä voi periaatteessa "tilata" linkityspalvelusta dokumentin URN-tunnuksella sen viitetiedot, sijaintitiedot tai dokumentin itsensä. Riippuu järjestelmästä, käyttäjän oikeuksista ja myös dokumentista, voidaanko haluttu palvelu toimittaa.

Käytännössä esimerkiksi Suomen nykyinen URN-linkityspalvelu kykenee vain toimittamaan dokumentin itsensä, jos palvelun käyttäjällä on tähän oikeus.

URN:n ja DOI:n kaltaisilla tunnistejärjestelmillä ja perinteisillä tunnisteilla on periaatteellinen ero. Samalle julkaisulle ei yleensä ole annettu useita tunnuksia (esimerkiksi ISBN- ja ISSN-tunnusta), mutta yhdellä julkaisulla voi olla sekä DOI- että URN-tunnus (ja mahdollisesti myös PURL- ja ARK-tunnukset), jos julkaisu on tavoitettavissa eri linkityspalveluiden kautta.

Selvennetään asiaa esimerkillä. Suomalaisen tutkijan tieteellinen julkaisu voi olla noudettavissa sekä kotimaisesta julkaisuarkistosta että kustantajan palvelimelta. Edellinen voi soveltaa linkitykseen URN-järjestelmää, jälkimmäinen DOI:ta. Käyttäjän pitäisi ohjautua siihen palveluun, jonka käyttöön hänellä on oikeus. Valinta voidaan tehdä OpenURL-järjestelmän avulla ilman että käyttäjä edes tietää asiasta; lopputulos on se, että hän näkee vain ne palvelut, joita saa käyttää.

Verkkoympäristössä myös tunnusten toisensa poissulkevuus on historiaa. Samalle verkkojulkaisulle voidaan antaa useita tunnisteita riippuen toimintaympäristöstä. Suomessa valtioneuvosto on suosittanut OID-numeroon perustuvien tunnusten käyttöä asiakirjojen identifiointissa. OID:t ovat yhteismitallisia URN-tunnusten kanssa. Sama julkaisu voi olla löydettävissä kansalliskirjaston verkkoarkistosta, jossa julkaisuun pääsee käsiksi MD5-tarkistussummaan perustuvan URN-tunnuksen avulla. Lisäksi, jos asiakirja on monografia, sille voidaan antaa ISBN.

Internet-käyttö

URN-kehitystyön tavoitteena on luoda periaatteessa koko Internetin kattava järjestelmä, jossa käyttäjä voi esimerkiksi WWW-selaimessa käyttää julkaisussa olevaa URN-tunnusta hyperlinkkinä kuten URL-tunnusta. Verkon käyttäjä voisi esimerkiksi kopioida URN:n selaimen Location-ikkunaan ja painaa Return-näppäintä. Tämän jälkeen haluttu dokumentti tai sen tiedot haetaan verkosta käyttäjälle riippumatta siitä, missä ne sijaitsevat.

Käyttäjälle tämä on yksinkertaista, mutta tarvittava tekninen infrastruktuuri on kaikkea muuta kuin yksinkertainen. Onneksi suurin osa tarvittavasta teknologiasta (DNS-järjestelmä) on jo olemassa.

Globaalin URN-linkityspalvelun ylin taso on Resolution Discovery Service (RDS), palvelu jonka avulla verkosta löydetään palvelin, joka kykenee avaamaan halutun URN-tunnuksen. Tarkoitus on pystyttää RDS-palvelut osoitteisiin urn.net ja uri.net. Internet Assigned Names Authority (IANA) tallentaa URN-linkityspalveluiden edellyttämät nimipalvelun osoitetiedot näihin järjestelmiin. Internet-nimipalvelun (Domain Name Service, DNS) luonteen mukaisesti nämä tiedot leviävät periaatteessa kaikkialle verkkoon. Oikean URN-linkityspalvelun paikallistaminen ei siis ole riippuvainen RDS-nimipalvelimesta, vaikka tarvittavat tiedot tallennetaankin Internet-nimipalveluun sen kautta.

Tarvittavan URN-linkityspalvelun/-palvelujen löytämisen menetelmät vaihtelevat paljon identifikaatiojärjestelmästä riippuen. Perusjako on "tyhmien" ja älykkäiden" tunnisteiden välillä.

Älykkäät tunnisteet kuten ISBN sisältävät vinkin siitä, mistä päin maailmaa linkityspalvelin voisi löytyä. Maantunnuksella 951 tai 952 alkavat ISBN:t ovat ainakin periaatteessa avattavissa Suomen kansallisbibliografian kautta, ja vastaavasti maantunnus 3 viittaa Saksaan, Itävaltaan tai Sveitsin saksankieliseen osaan (nimipalveluun voidaan määritellä joukko palvelimia ja niille prioriteettijärjestys). Internet-nimipalvelulle voidaan suhteellisen helposti ohjeistaa ISBN-maantunnuksien käyttö. Yhdysvalloissa ja Englannissa maantunnus ei, kustannustoiminnan

kansainvälistymisen vuoksi, aina pidä paikkaansa. Eurooppalaiset kustantajat ovat ostaneet amerikkalaisia alan yrityksiä, ja tämän vuoksi Yhdysvalloissa ilmestyy kirjoja joiden ISBN alkaa numerolla 3. Mutta kansainvälisen ISBN-keskuksen mukaan kaikki nämä julkaisut luetteloidaan Saksan kansallisbibliografiaan, joten URN-linkityksen kannalta mitään ongelmaa ei pitäisi olla.

Globaalin URN-linkityksen kannalta on haasteellista, että monet tunnisteet ovat ISSN:n tapaan "tyhmiä", eivätkä kerro dokumentin julkaisupaikasta mitään. Tällöin linkityspalvelimen löytäminen voi olla vaikeaa, ellei ole käytettävissä globaalia tietokantaa, jonka kautta URN-tunnukset voidaan muuntaa sijaintipaikoiksi. Esimerkiksi kansainvälinen ISSN-tietokanta toimii ISSN:ään perustuvien URN-tunnusten globaalina linkityspalveluna hyvin, edellyttäen että ISSN-tietokannan viitteissä olevat lehtien kotisivujen URL:t ovat ajan tasalla.

Monilla muilla "tyhmillä" tunnisteilla asiat ovat huonommin kuin ISSN:llä. Artikkeleiden ja lehtien numeroiden SICI-tunnukselle ei ISSN-tietokantaa vastaavaa globaalia kantaa voida käytännössä rakentaa, teknisistä ja tekijänoikeussyistä. Globaaliin artikkelitietokantaan pitäisi koota kaikki Artoa vastaavat tietokannat, joita on sadoittain ellei tuhansittain. Virtuaalinen yhteisluettelo ei tässä tapauksessa ole mahdollinen: on mahdotonta tehdä haku aina sadoista tietokannoista kerrallaan, kun halutaan löytää artikkeli verkosta sen SICI:n perusteella.

SICI:n URN-linkitystä varten hahmoteltiin 90-luvun lopulla EU:n DIEPER-projektissa ratkaisu, jossa SICI-tunnukset olisi linkitetty artikkeleihin ISSN-tietokannan kautta. ISSN-tietokannan tietueisiin olisi voitu lisätä uusi tietoelementti, johon olisi tallennettu niiden tietokantojen URL-osoitteet, joihin sisältyy artikkeleita asianomaisesta lehdestä. Jokaisen SICI:n URN-linkityspyyntö olisi tällöin voitu ohjata ISSN-tietokantaan. SICI-tunnukseen sisältyvän ISSN-tunnuksen avulla olisi voitu hakea lehden tiedot, ja tarkistaa löytyykö niistä yhden tai useamman artikkelitietokannan osoite. Jos olisi löytynyt, artikkeli (tai viite) olisi voitu noutaa käyttäjälle, jos hänellä olisi ollut ainakin yhteen palveluun käyttöoikeus.

Yllä kuvattua oleellisesti kätevämpi ratkaisu on se, että ISSN-tietokantaan ei nykyisen käytännön vastaisesti enää toimiteta elektronisten lehtien URL- eli sijaintitietoja lainkaan, vaan kausijulkaisutietueisiin tallennettaisiin vain URN. Tämän jälkeen kaikki sijainti- ja muiden tietojen ylläpito voitaisiin hoitaa (kansallisella) URN-linkityspalvelimella. Kun URN tallennetaan hyperlinkkinä, ISSN-tietokannan Virtua-järjestelmään ei tarvitse tehdä mitään muutoksia.

Tämä esimerkki kuvaa, miten vaikeaa joidenkin perinteisten tunnisteiden sovittaminen Internet-palveluiksi voi olla, ellei toteutustapaa mietitä tarkasti. Artikkelitietokantojen ja lehtien osoitteiden ylläpito on työlästä keskitettynäkin palveluna, jos painettua aineistoa ryhdytään digitoimaan mittavasti.

Kaikki Internet-dokumentit kattavaa URN-linkityspalvelujen verkkoa ei koskaan kyetä rakentamaan, eikä siihen ole edes tarvetta - valtaosa verkkoaineistosta on tarkoitettu lyhytikäiseksi. URN-linkityspalvelua ei kannata rakentaa eikä julkaisua identifioida URN-tunnuksella tai oikeastaan muutenkaan, ellei aineistoa ole tarkoitus säilyttää pitkään. Siksi linkityspalveluiden rakentamisen mielekkyys on julkaisujen osalta vahvasti kytyksissä siihen, tallentaako kansalliskirjasto oman maansa elektronisia julkaisuja, ja miten kattavasti se sen tekee. Muille aineistoille on muita toimijoita: Kansallisarkisto vastaa Suomessa asiakirjoista, Suomen elokuva-arkisto radio- ja TV-aineistoista sekä Yhteiskuntatieteellinen tietoarkisto tutkimusaineistoista. Mutta kaikki voivat soveltaa samaa teknistä infrastruktuuria palvelujaan kehittäessään: asiakirjat eivät tarvitse julkaisuista poikkeavaa linkitystä.

Suomessa kansallisbibliografian ID-tunnukseen perustuvien URN-tunnusten jakelu käynnistyi toukokuussa 1998, yhtä aikaa Ruotsin kansalliskirjaston kanssa. URN-generointiohjelma löytyy osoitteesta <http://www.lib.helsinki.fi/cgi-bin/urn.pl> ja URN-tunnusten käyttöön liittyvää ohjeistusta osoitteesta <http://www.kansalliskirjasto.fi/julkaisuala/urn/>.

Vuodesta 1998 lähtien tunnuksia on jaettu jo kymmeniä tuhansia. Palvelua ei ole mainostettu kirjastoalan ulkopuolelle, muun muassa koska linkityspalvelu on puuttunut. Palvelun julkistamisen jälkeen keväällä 2007 tunnusten menekki ja mielenkiinto URN-järjestelmää kohtaan on voimakkaasti kasvanut, ja ongelmaksi on noussut uusien ja potentiaalisten käyttäjien tukeminen.

Kansalliskirjastot ovat sitoutuneet URN-tunnusten käyttöön, ja URN-pohjaisia järjestelmiä on jo muun muassa Saksassa ja Norjassa, missä kansalliskirjasto aikoo identifioida kaiken aineiston digitaalisessa arkistossaan NBN:ään perustuvien URN-tunnuksin. Kirjastomaailman ulkopuolella URN-tunnusten käyttöönotto ollut hidasta, koska kustantajat ovat panostaneet DOI-tunnukseen.

URN-linkityksen edellyttämät sinänsä yksinkertaiset lisäpiirteet on tietävästi rakennettu nimipalvelinsovelluksiin jo vuosia sitten. Tältä osin tekniikka on siis kunnossa. Valitettavasti linkityspalvelimien paikallistamiseen tarvittavia Resolver Discovery Service –palvelimia ei ole rakennettu, minkä vuoksi Internetin nimipalvelusta ei löydy erikseen URN-linkityspalveluiden tietoja.

Yhtä tärkeää kuin verkon infrastruktuurin päivitys on URN-tuen saaminen WWW-selaimiin ja esimerkiksi kirjastojen käytössä oleviin järjestelmiin. WWW-selaimille on rakennettu URN-linkitystä varten lisäohjelmia (plug-in), mutta ne on tarkoitettu väliaikaisratkaisuuksi. Teknisesti parempi tilapäisratkaisu on sellainen, jossa käyttäjän ei tarvitse asentaa WWW-selaimensa mitään ylimääräistä. Tämäntyyppinen ratkaisu, jossa URN tallennetaan URL-osoitteen tapaan, on käytössä esimerkiksi Suomessa ja Ruotsissa. Tarvittavien ohjelmien rakentaminen ei tietävästi ollut vaikeaa. Tämä ei ole yllätys, koska URN-järjestelmä on pyritty rakentamaan niin, että linkityspalveluiden rakentaminen ja soveltaminen olisi helppoa.

WWW-selainten rikastaminen URN-tuella ei sekään olisi teknisesti hankalaa; käytännössä esteenä on ollut politiikka. Microsoftia ei ole saatu kiinnostumaan URN-järjestelmän implementoinnista.

Kirjastojen järjestelmistä URN-linkityksen suora tuki puuttuu muun muassa sen vuoksi, että kirjastot eivät ole sitä sovellustoimittajilta vaatineet. Eipä sille ole erityistä tarvettakaan: jos URN-tunnukset tallennetaan hyperlinkkeinä, ei kirjastojärjestelmissä sen paremmin kuin MARC-formaatissakaan tarvita mitään muutoksia. URN-tunnukset toimivat kuten URL-tunnukset, ja ne voidaan tallentaa kenttään 856.

Joskus tulevaisuudessa kun URN on osa Internetin yleistä infrastruktuuria, kirjastojen digitaalisiin arkistoihin ja muihin järjestelmiin voi tulla URN-linkityspyyntöjä suoraan Internetin nimipalvelusta. Vasta tällöin saadaan URN-järjestelmästä irti kaikki edut. Kirjastoissa sen avulla voidaan periaatteessa rakentaa erityispalveluita kuten hajautettu kopioluettelointi. Se voisi toimia niin, että luetteloija antaa cataloguing client –sovellukselle ISSN-tunnukseen tai ISBN-tunnukseen perustuvan URN:n ja painaa RETURN-näppäintä. Nimipalvelusta selviää oikean tietokannan – esimerkiksi Uuden Seelannin kansallisbibliografian – osoite, ja sinne lähetetään URN-haku. Uuden Seelannin kansallisbibliografiassa avataan URN-tunnuksen sisältö ja tehdään ISBN-haku, ja löytynyt tietue toimitetaan MARC-muodossa suomalaiselle luetteloijalle.

Tällaisen järjestelmän rakentaminen on paitsi tekninen myös poliittinen ja organisatorinen haaste, mutta lähivuodet osoittavat kykenevätkö kansalliskirjastot sopimaan tämäntyyppisestä luettelointiyhteistyöstä.

10. DOI

Digital Object Identifier -tunnusjärjestelmän (<http://www.doi.org>) taustavoimat ovat Corporation for National Research Initiatives eli CNRI (<http://www.cnri.reston.va.us/>), joka vastaa järjestelmän teknisestä perustasta, sekä International DOI Foundation (IDF; <http://www.doi.org/>), joka koordinoi toimintaa käytännön tasolla. IDF:ssä vaikutusvalta on suurilla kustantajilla, jotka vastaavat käytännössä myös rakennettavien palvelujen rahoittamisesta. IDF ylläpitää myös DOI Handbook -verkkojulkaisua, joka on perusteellinen johdatus tunnusjärjestelmään (<http://www.doi.org/hb.html>).

Pienenä kuriositeettina mainittakoon, että DOI-käsikirjan DOI-tunnus on doi:10.1000/182. Kirjalle ei ole haettu ISBN-tunnusta.

DOI-projekti alkoi syyskuussa 1996 ja sen oli tarkoitus kestää viisi vuotta. Ensimmäiset toimivat tuotantojärjestelmät esiteltiin Frankfurtin kirjamesseilla lokakuussa 1997. Tätä kirjoitettaessa DOI on laajasti tuotantokäytössä, mutta sen sovellusalue rajautuu suuriin tieteellisiin kustantajiin. Näyttävien esimerkkien DOI-pohjaisista palveluista on Crossref-järjestelmä (<http://www.crossref.org/>), johon oli tallennettu heinäkuussa 2007 yli 27 miljoonaa tieteellistä artikkelia. Crossref on suurin Internetissä toimiva linkityspalvelu, ja ehdottomasti isoin DOI-vetoinen palvelu: kaikkiaan DOI-tunnuksia on jaettu noin 28 miljoonaa.

Teknisesti DOI perustuu CNRI:n kehittämään Handle systemiin (<http://www.handle.net/>), joka on periaatteessa yhteensopiva URN-tunnusten kanssa. Mitä tahansa DOI-tunnusta voi siis käyttää myös URN-tunnuksena, kunhan DOI:lle rekisteröidään oma URN Namespace Identifier. Vastaavasti URN voitaisiin muuttaa DOI-tunnukseksi, mutta tässä ei juuri ole mieltä. Ylipäätään DOI:n edellyttämä tekninen ratkaisu (riippuvuus Handle system'stä) on ongelma, koska linkityspalvelulle riippuvuus mistä tahansa yksittäisestä tekniikasta on riski, vaikka kyseinen tekninen ratkaisu olisi hyväkin. Kehitys on jatkuvasti nopeaa, minkä vuoksi on vaarallista olettaa että Handle system sen paremmin kuin http-protokollakaan säilyisi edes vuosikymmenien, saati vuosisatojen ajan. Kustantajille muutama vuosikin voi olla tarpeeksi; kun aineisto ei enää käy kaupan sillä ei enää ole arvoa. Kirjastojen, ja eritoten kansalliskirjastojen, aikaperspektiivi on kuitenkin aivan toinen.

Kirjastoille DOI-palveluiden yksipuolisuus (ei mahdollisuutta pyytää bibliografisia tietoja, vaan vain julkaisu itse) on toinen ongelma. DOI:n varaan voidaan luoda (kaupallisten) dokumenttien välityspalvelu, mutta ei esimerkiksi kopioluettelointipalveluita. Lisäksi kirjastojen mahdollisuudet lisätä uusia palveluita DOI-järjestelmään ovat hyvin rajalliset, koska IDF kontrolloi järjestelmää eikä kirjastoilla ole ollut juuri sananvaltaa järjestelmän kehitykseen. URN-järjestelmään meidän on mahdollista ja oleellisesti helpompi vaikuttaa, koska kehittäjäyhteisö on avoin ja kirjastot ovat suurimpia URN-käyttäjiä.

DOI-järjestelmää ja sen teknistä perustaa on yritetty standardoida vaihtelevalla menestyksellä. CNRI:n kehittämä Handle system saatiin viiden vuoden yrittämisen jälkeen Internet-standardiksi. Palvelun yleiskuvaus RFC 3650 julkaistiin vuonna 2003 (<http://www.ietf.org/rfc/rfc3650.txt>). RFCt 3651 ja 3652 määrittelevät Handle-palvelut ja protokollan. Kaikkien näiden RFCiden status on

Informational, eli ne eivät varsinaisesti ohjaa Internetin rakentamista, vaan ovat verkon käyttäjille ja rakentajille tiedoksi.

DOI-tunnuksen varsin Handle-vetoinen syntaksi on määritelty amerikkalaisessa standardissa ANSI/NISO Z39.84-2000, joka on muiden NISO-standardien tapaan haettavissa verkosta maksutta (<http://www.niso.org/standards/resources/Z39-84-2005.pdf>).

DOI-järjestelmän standardointi ISO:ssä käynnistyi alkuvuodesta 2006. Tätä kirjoitettaessa ei ole vielä saatu edes Committee draftia julki, mutta yksimielisyys tekstin sisällöstä ei liene enää kovin kaukana. Suurin kompastuskivi on DOI:n kaksinaisuusluonteessa. Tunnistejärjestelmänä se vain täydentäisi ISON nykyisiä tunnisteita tarjoamalla niille linkitysominaisuudet. Mutta DOI on myös tunniste, ja DOI-käsikirjan lisäksi sillä voidaan tunnistaa myös muita kirjoja ja vaikkapa kausijulkaisuja ja niiden artikkeleita – joihin DOI:ta onkin ahkerimmin sovellettu. Muiden tunnisteiden edustajat ISO:n DOI-työryhmässä ovat olleet huolissaan siitä, että DOI voi syrjäyttää ISBN:n, ISSN:n ja periaatteessa kaikki muutkin ISO-tunnisteet. Paljon aikaa on työryhmässä käytetty sen pohtimiseen, miten määritellä ”tunniste-DOI:n” ja muiden tunnisteiden suhde, eikä lopullista tekstimuotoilua ollut vielä heinäkuussa 2007 keksitty.

DOI rakennettiin alun perin niin, että se on kustantajan kannalta mahdollisimman kätevä: mitään ei tarvitsisi jättää tunnistamatta, ja tunnuksen rakenteen saisi päättää itse (jotta vanhat kotitekoiset tunnisteetkin voisi sovittaa DOI-järjestelmän sisään). Samat piirteet jotka tekivät DOI:sta kustantajille mainion järjestelmän, herättävät epäilyksiä muiden tunnisteiden ylläpitäjissä.

Kansallisbibliografian ID-numero palvelee kansalliskirjastoja samalla tapaa kuin DOI:n tunnisteosa kirjastoja. NBN ei kuitenkaan ole herättänyt huolestumista esimerkiksi kansainvälisessä ISSN-keskuksessa, koska NBN-tunnus annetaan vain aineistolla jolle mikään perinteinen tunnus ei sovi.

DOI koostuu kahdesta osasta: julkaisijan ID-tunnuksesta (prefix) sekä teoksen ID-tunnuksesta (suffix), joita erottaa kauttaviiva. Esimerkiksi edellä mainittu DOI-tunnus

10.1000/182

sisältää prefiksin "10.1000" ja suffiksin eli julkaisun tunnisten "182".

Prefiksi jakautuu hakemiston tunnukseen (Directory Code) sekä rekisteröijän tunnukseen (Registrant Code), jotka erotetaan pisteellä. Toistaiseksi ainoa sallittu hakemiston tunnus on "10". Rekisteröijän tunniste identifioi kustantajan tai muun organisaation, joka on antanut teokselle DOI:n. Esimerkissä tämä tunniste on "1000". Rekisteröijän tunnuksia jakaa tätä kirjoitettaessa vain DOI Registration Agency. Jakelun keskityksellä varmistetaan DOI-prefiksien ainutkertaisuus, joka on tunnisten toimivuuden kannalta ehdottoman välttämätöntä.

DOI-prefiksin pituutta ei ole rajattu, joten DOI:ssa on periaatteessa tilaa on mielivaltaiselle määrälle kustantajia. Käytännössä rekisteröijien tunnisteiden jakelusta tuli järjestelmän toiminnan alkutaipaleella pullonkaula, koska se oli käsityötä. International DOI Foundation keskusteli kustantajatietojen vaihdosta ISBN-järjestelmän edustajien kanssa, mutta kaikkia osapuolia tyydyttävään kompromissiin ei tuolloin päästy. Sittenkin suhteet ovat lämmenneet; vuonna 2007 ISBN-keskus ja DOI-jakelukeskus toimivat läheisessä yhteistyössä Saksassa, Italiassa, Englannissa ja Yhdysvalloissa eli monissa niistä maista, joissa tieteellinen kustantaminen on merkittävää liiketoimintaa.

DOI:n kustantajatunnuksen hankkiminen maksaa kustantajan koosta riippuen vaihtelevan summan rahaa, ja myös DOI-tunnukset ovat maksullisia (eivät kuitenkaan kirjastoille). Vastineeksi tästä investoinnista saa mahdollisuuden käyttää DOI-linkityspalveluja. DOI-järjestelmän tukijat ovat usein huomauttaneet, ettei ilmaista linkityspalvelua ole olemassakaan; maksuton palvelu voidaan toki tarjota, mutta joku jossakin maksaa viulut. DOI:n tapauksessa tämä joku on palvelua käyttävä kustantaja tai muu kirja-alan toimija.

DOI-järjestelmän on tulevaisuudessa tarkoitus kattaa paljon muutakin kuin vain linkityspalvelu. Priscilla Caplanin mukaan:

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 (<http://info.lib.uh.edu/pr/v9/n1/capl9n1.html>).

DOI on siis paljon enemmän kuin vain ID-tunnus - sen varaan rakennetaan elektronisten julkaisujen jakelujärjestelmää. Tunnuksesta maksaessaan kustantajat maksavat infrastruktuurista, joka mahdollistaa julkaisujen välittämisen, maksullisesti tai ilmaiseksi. Toistaiseksi tästä infrastruktuurista on saatu rakennetuksi vasta yksi erittäin toimiva palvelu (CrossRef), mutta on syytä uskoa että DOI:n käyttö tulee edelleen laajenemaan. Vauhtia voi hidastaa se, ettei IDF:llä ole omia kehittämisresursseja. Se on riippuvainen kustantajien omista hankkeista.

Teoksen DOI-numeron antaa joko kustantaja tai muu rekisteröijä. DOI-tunnukseksi kelpaa mikä tahansa tunniste, mutta on todennäköistä tai ainakin toivottavaa että se perustuu silloin kuin mahdollista olemassa oleviin järjestelmiin kuten SICI:iin. Prefiksin eli tunnisteen alussa voidaan sulkeissa tai muulla tavoin ilmoittaa mistä järjestelmästä tunnus on peräisin, mutta tämä tieto on vapaaehtoinen. Niinpä vanhaan 10 merkin mittaiseen ISBN-tunnukseen perustuva DOI voi näyttää tältä:

10.1002/(ISBN)0-471-58064-3

tai tältä:

10.1002/0-471-58064-3

Tunnisteen identifioinnin vapaaehtoisuus vaikeuttaa DOI-tunnuksen oikeellisuuden ohjelmallista tarkistamista. Tunnuskoodin puuttuessa kirjastojärjestelmän tai muun DOI:ta tutkivan ohjelman pitäisi kokeilla onko prefiksin sisältönä ISBN, ISSN, SICI, BICI tai muu ylipäättään tarkistettavissa oleva tunnus, ja onko prefiksin alussa mahdollisesti tieto käytetystä tunnistesta jollakin tavoin tallennettuna. DOI:ssa ei ole tarkistusnumeroa, jonka avulla koko DOI-tunnuksen oikeellisuuden voisi tarkistaa. Niinpä DOI-prefiksi voi olla väärä ja koko DOI sen vuoksi virheellinen, vaikka suffiksiin tallennettu ISBN-tunnus olisi kunnossa.

Pahin tekninen DOI-ongelma on kirjastojen kannalta kuitenkin se, että kustantaja voi käyttää DOI-järjestelmän sisällä mitä tahansa omatekoista tunnusta. Siksi DOI:n syntaksin standardoinnin arvo on teknisesti verraten vähäinen; kaaoksen standardoimisesta ei juuri ole iloa. Periaatetasolla on tietenkin syytä olla iloinen siitä, että IDF haluaa julkistaa DOI:n ISO-standardina ja avata tällä tavoin muillekin mahdollisuudet vaikuttaa DOI:n sisältöön.

Internet-käyttö

DOI-tunnusten linkityspalvelu toimii HTTP-protokollan varassa siten, että DOI-tunnukset upotetaan aina URL-tunnukseen, joka on <http://dx.doi.org>. Jos siis DOI-tunnus on 10.1006/rwei.1999.0001, se tallennettiin hyperlinkkinä muodossa <http://dx.doi.org/10.1006/rwei.1999.0001>.

CNRI:n ylläpitämän keskitetyn DOI-linkityspalvelun (<http://dx.doi.org>) kautta voi hakea mitä tahansa 28 miljoonasta DOI-tunnuksesta. Järjestelmän tekninen keskittäminen takaa, että tämä toimii. Toistaiseksi CNRI on kyennyt kasvattamaan palvelun kapasiteettia niin, että se toimii luotettavasti.

DOI-prefiksillä ei siis toistaiseksi ole mitään tekemistä oikean DOI-linkityspalvelimen paikallistamisen kanssa, vaan ”sen ainoan oikean” löytäminen hoidetaan URL-tunnuksella, joka ei ole osa DOI:ta. Tässä suhteessa DOI ja URN poikkeavat toisistaan oleellisesti jo nyt, koska URN-linkityspalveluita on useita. Euroopan kansalliskirjastot ovat neuvotelleet keskitetyn URN-linkityspalveluiden paikallistamisjärjestelmän rakentamisesta, tavoitteena päästä asiasta sopimukseen syksyllä 2007.

Keskitetyn teknisen perustansa ja kaupallisen orientaationsa vuoksi DOI:n varaan ei tulla rakentamaan todella kattavaa linkityspalvelua. DOI rajautuu etupäässä suurten kustantajien kaupallisen aineiston jakeluun, ja tässä sillä voi olla paljon annettavaa. Vapaasti käytettävissä järjestelmissä kuten julkaisuarkistoissa URN on suosituimpi ratkaisu. Jos ja kun DOI-tunnusten ja niitä käyttävien kustantajien määrä ja DOI-järjestelmän hakukäyttö edelleen kasvavat, on järjestelmän tekninen ja hallinnollinen infrastruktuuri rakennettava nykyistä järeämmäksi ja hajautetuksi.

11. Tunnusten tallennuksesta

Verkossa olevan elektronisen julkaisun tunnisteiden tehokas hyödyntäminen edellyttää sitä, että (perinteinen) tunniste tallennetaan dokumentin sisään siten, että Google ja muut hakuohjelmistot indeksoivat datan. Varmin ratkaisu on tallentaa tunniste julkaisun nimiösivulle. Indeksintisovellus ei tällöin voi kuitenkaan päätellä, että kyse on ylipäätään tunnisteesta tai ainakaan asianomaisen julkaisun tunnisteesta. Tätä ongelmaa voidaan lievittää – ja varmistaa linkityspalvelun käyttö – tallentamalla tunnus URN- tai mahdollisesti DOI-tunnisteena.

Jos julkaisu sisältää oman kuvailunsa, tunniste on hyvä tallentaa myös sille varattuun metadatakenttään. Jos hakuohjelmisto indeksoi metadatan, saadaan optimitulos. Valitettavasti verkkojulkaisujen metadatan laatu on ollut niin huono, että monet hakuohjelmistot eivät ole käyttäneet sitä indeksoinnissa hyväksi.

Ulkoista metadattaa tallennettaessa perusvaatimus on, että tallennukseen käytetyssä formaatissa on paikka tunnisteelle. MARC21-formaatissa sekä sen varianteissa kuten MARC21-Fin:ssä voidaan tallentaa periaatteessa mitä tahansa. Kirja-alan ONIX ei ole yhtä joustava, sen avulla ei voida toistaiseksi tallentaa esimerkiksi URN-tunnusta.

Kaikki dokumenttiformaatit eivät tarjoa yhtä hyviä mahdollisuuksia sisäisen metadatan tallentamiseen. Adobe Acrobat ja tulevan MS Officeen XML-pohjainen formaatti tukevat Dublin

Core –formaattia ja tunnisteiden tallentaminen esimerkiksi PDF- ja Word-dokumentteihin on sen vuoksi periaatteessa helppoa. Ilman apuvälineitä kunnon jälkeä ei synny; metadatan luontivaiheessa pitää olla käytettävissä hyvillä opastetiedoilla varustettu tallennusalusta, josta on linkki tunnisteiden luontisovellukseen. Hyvä esimerkki on Helsingin yliopiston E-thesis –järjestelmän väitöstietolomake (<http://ethesis.helsinki.fi/forms/diss/>).

Monet tekstiformaatit ovat indeksointiohjelmille outoja, eikä niitä voida indeksoida lainkaan. Tällöin ainoa tallennustapa josta on edes jotakin hyötyä on tunnuksen tallentaminen nimiösivulle. Kuvaformaattien osalta tilanne voi olla vielä pahempi: identifiointitunnusta ei voi tallentaa kuvankäsittelyohjelmalla kuvan päälle, mutta ei tunnusta ei voi välttämättä tallentaa myöskään kuvan tekstimuotoiseen nimiöön.