

7. URN

Sisältöä päivitetty viimeksi maaliskuussa 2023.

7.1 URN-järjestelmän tausta ja tukiorganisaatio

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

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

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

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

URN-syntaksin uudistaminen oli vaativa prosessi. URI-syntaksin lisäksi oli varmistettava URN-tunnuksen yhteensopivuus myös muiden relevanttien Internet-standardien kuten UNICODEN kanssa. Työ jäi harmittavasti vähän kesken: R-komponentin eli URN-tunnuksen lisättävän komennon (ks. luku 6.3) syntaksia ja semantiikkaa ei määritelty, vaan se päätettiin siirtää siksi kunnes URN-resolveiden kehittäjät alkavat soveltaa tätä ominaisuutta. Kansalliskirjasto on tätä kirjoitettaessa jo lisännyt resolveriinsa ensimmäisen R-komponentin avulla toteutetun palvelun. Osana resolverin kehittämistä URN-nimialueiden rekisteröintianomuksia tarkistava ryhmä käynnisti helmikuussa 2023 keskustelun siitä, miten R-komponentilla toteutetut palvelut ja niiden parametrit rekisteröidään, ja millainen R-komponentin sisäisen rakenteen tulisi olla.

7.2 URN:n soveltamisalat ja keskeiset käyttäjät

URN on globaali tunnistejärjestelmä, jonka avulla voi identifoida mitä tahansa. URN-tunnistejärjestelmän jakautuminen tunnistejärjestelmäkohtaisiin nimialueisiin tekee sen laajentamisen helpoksi: mikä tahansa olemassa oleva tunnistejärjestelmä voidaan ottaa mukaan, ja tarvittaessa on mahdollista luoda uusia. URN on sateenvarjo, joka voi sisällyttää itseensä sekä perinteiset tunnisteet että muut PID-tunnistejärjestelmät. Järjestelmän joustavuuden takaa se, että kullakin nimialueella sovelletaan sen rekisteröineen tunnisteiden tai PID-järjestelmän taustalla olevaa standardia ja toimintaperiaatteita, URN: ISBN-tunnus on edelleen ISBN, mutta toiminnallinen ISBN. URN:DOI on DOI, ja resoluutuu kuten DOI.

Missään muussa PID-tunnistejärjestelmässä ei ole nimialueita. Jokaiselle DOI-tunnukselle pätevät samat säännöt, kun taas URN-järjestelmässä tunnisteiden rakenne ja soveltamisala määritellään kullakin nimialueella erikseen.

URN tunnettiin alun perin kirjastojen ja erityisesti kansalliskirjastojen tunnisteena. Tämän rajoittuneen käsityksen aiheutti URN:NBN-nimialue, jota sovelletaan 13 Euroopan maassa, kuten Suomessa, Ruotsissa, Norjassa, Saksassa ja Italiassa. URN:NBN (national bibliography number) on kansalliskirjastojen hallinnoima URN-nimialue, joka mahdollistaa niiden käyttämien paikallisten tunnistejärjestelmien käytön URN-tunnuksina.

Muita kirjastojen kannalta tärkeitä URN-nimialueita ovat myös esimerkiksi ISBN- ja ISSN-tunnistejärjestelmien URN:ISBN ja URN:ISSN. Niiden nimialuerekisteröinneissä korostetaan sitä, että URN:ISBN- ja URN:ISSN-tunnuksia tulee soveltaa ISBN- ja ISSN-tunnuksien tapaan. Mutta URN ei ole vain kirjastojen tunnistejärjestelmä; useimmilla nimialueilla ei ole mitään tekemistä kirjastojen kanssa. Eikä URN-tunnuksella ole monopoliasemaa edes kansalliskirjastoissa. Useimmat kansalliskirjastot käyttävät muita tässä esityksessä kuvattuja PID-tunnistejärjestelmiä, ja monet luottavat Cool URI – tunnuksiinkin joidenkin aineistojen identifiointiin. Valinta ei välttämättä ole oma, vaan sovellustoimittajan.

URN-standardit kehittänyt IETF ei valvo eikä aktiivisesti edistä URN-tunnistejärjestelmän käyttöä edes ylätasolla, saati nimialuekohtaisesti. Ja jos jokin taho URN-markkinointia tekisi, ensimmäinen ongelma olisi sen kohdentaminen. URN-nimialueita ovat rekisteröineet hyvin erilaiset toimijat erilaisiin tarpeisiin. Muiden PID-sovellusten käyttäjillä on yhteinen resolverisovellus ja keskitettyjä palveluita, mutta URN-järjestelmässä yhteistyö ja keskitetty hallinnointi voi toteutua vain kunkin nimialueen sisällä.

Koska URN on Internet-standardi, URN-nimiavaruuksien rekisteröinnin vastuunorganisaatio on IANA (Internet Assigned Names Authority). Se vastaa myös rekisteröityjen nimialueiden luettelosta. Käytännön työtä eli rekisteröintianomusten tarkistamisesta vastaa viisi asiantuntijaa, joista yksi on Juha Hakala Kansalliskirjastosta. Jatkossa IANA voisi vasta myös esimerkiksi URN-tunnistejärjestelmän palveluiden rekisteristä, mutta tästä ei ole vielä päätöstä.

URN-nimialueet ovat keskenään perin erilaisia, mikä tekee rekisteröintianomusten käsittelystä haasteellista: sovellusympäristö on yleensä kaikille arvioijille tuntematon. Monet nimialueet eivät tarjoa resoluutiopalveluja, eikä URN-tunnisteiden määrittelevä standardi RFC 8141 tätä edellytäkään. Näillä nimialueilla URN vastaa perinteistä ei-toiminnallista tunnistejärjestelmää. Esimerkkejä ei-toiminnallisista URN-nimialueista ovat URN:MRN, URN:GDST ja URN:UVC. International Association of Marine Aids to Navigation and Lighthouse Authorities'n (IALA) rekisteröimä nimialue URN:MRN mahdollistaa erilaisten merenkulkuun liittyvien fyysisten tai elektronisten objektien identifiointiin. Global Dialogue on Seafood Traceability (GDST) on käynnistänyt kalastukseen liittyvän URN:GDST-nimialueen rekisteröinnin. Tavoitteena on minimoida laitton kalastus mahdollistamalla kuluttajille myyntiin tulevien kalaerien seuranta koko tuotantoketjun ajan.

URN:UVC on EU:n koronarokotustodistuksen tunnisteiden nimialue, joka on jo käytössä vaikka sen rekisteröintiprosessi on kesken. EU ei ole ainoa varaslähdön ottanut toimija, rekisteröimättömiä URN-tunnuksia käytetään myös ainakin LinkedIn-palvelussa ja DDI-metadatatandardissa.

Uusia rekisteröintipyyntöjä tulee muutamia vuosittain. Mielenkiintoinen uusi aluevaltaus oli maaliskuussa 2023 jätetty Itävallan liittokanslerinviraston hakemus GVAT-nimialueesta. Sen tunnuksia on tarkoitus käyttää Itävallan julkishallinnon tuottamien asiakirjojen identifiointiin. Nähtäväksi jää, seuraavatko muut maat Itävallan esimerkkiä.

Kulttuurialalla URN-tunnistetta soveltaa kirjastojen lisäksi elokuvateollisuus. Sen "oma" nimialue URN:EIDR on toiminnallinen ja hyvin hallintoitu. URN-järjestelmässä tunnisteiden soveltamisen valvonta on mahdollista vain nimialuekohtaisesti. Jotkut nimialueet, kuten URN:NBN, voidaan jakaa edelleen pienempiin osiin, joita hallinnoidaan itsenäisesti. Esimerkiksi Kansalliskirjaston vastuulla on vain Suomen osuus NBN-järjestelmästä eli URN:NBN:FI-tunnukset. On myös URN-nimialueita, joilla ei voi olla keskitettyä valvontaa. Esimerkiksi URN:UUID^[1]-tunnuksia voi generoida kuka tahansa, koska tahansa ja mille tahansa elektroniselle objektille.

7.3 URN-tunnuksen rakenne

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

Label "URN:", URN-tunnuksen nimiö. Jokainen URN-tunnus alkaa näin; tarkoituksena on helpottaa URN-tunnuksen löytämistä tekstin keskeltä, ARK-tunnuksen ARK:-nimiön tapaan.

Namespace Identifier (NID), nimialueen tunnus. Esimerkiksi "ISBN".

Namespace Specific String (NSS), resurssin tunnus. Esimerkiksi ISBN-numero.

ISBN-tunnuksesta 978-951-51-7661-5 saadaan URN-tunnus URN:ISBN:978-951-51-7661-5, joka voidaan esittää hyperlinkkinä <https://urn.fi/URN:ISBN:978-951-51-7661-5>. Koska tunnuksen alkuun lisättävä merkkijono on kunkin nimialueen sisällä aina sama, URN-tunnukset voidaan luoda ohjelmallisesti olemassa olevista perinteisistä tunnuksista. Jos URN ei ole toiminnallinen, sen tarjoama lisäarvo on tunnuksen globaali ainutkertaisuus ja sovellettavuus tiedonhaussa.

URN-tunnuksen perään voidaan lisätä vapaavalintaisia komponentteja, jotka eivät ole osa itse tunnusta eivätkä siis identifioi mitään. Ne ovat URI-syntaksin Fragment-ominaisuutta vastaava F-komponentti, URI Query -ominaisuutta vastaava Q-komponentti sekä R-komponentti, josta lisää tuonnempana tässä luvussa.

Nimialueiden tunnuksot ovat ainutkertaisia ja ne identifioivat pysyvästi URN-tunnuksessa käytetyn tunnistejärjestelmän. Kullakin nimialueella noudatetaan kyseisen tunnistejärjestelmän pelisääntöjä, jotka koskevat muun muassa tunnusten rakennetta ja identifioinnin kohteita. Esimerkiksi URN:ISBN-nimialueella toimitaan ISBN-standardin ja käyttöoppaan mukaisesti. URN siis tukee olemassa olevia tunnistejärjestelmiä eikä kävele niiden yli, kuten URI. Maaliskuuhun 2023 mennessä nimialueita on rekisteröity noin 80^[2]. Vain murto-osa näistä nimialueista kuuluu perinteisille bibliografisille tunnisteille tai ylipäättään julkaisujen tunnisteille. Jos on tarpeen identifoida esimerkiksi mittalaitteessa oleva anturi, sen voi tehdä käyttäen IETF:n kehittämää Internet of Things -laitetunnistetta, jonka nimialue on URN:DEV (<https://www.rfc-editor.org/rfc/rfc9039.html>).

URN-tunnuksen NSS-osioon tallennetaan varsinainen tunnus. Niiden tarvitsee olla ainutkertaisia vain oman nimialueensa sisällä. Tunnus 978-951-51-7661-5 voi siis toistua esimerkiksi DEV-nimialueella. Jos URN-tunnuksen saavalla dokumentilla on jo olemassa standarditunnus, sitä on käytettävä URN-tunnuksessa. Jos kirjalla on ISBN, URN-tunnuksen NSS-osion on oltava tuo ISBN-tunnus sellaisenaan, siihen ei saa lisätä mitään eikä siitä saa ottaa pois mitään.

URN-tunnuksina paikalliset tunnuksot muuttuvat globaaleiksi ja ainutkertaisiksi. Esimerkiksi kansalliskirjastojen sisäisten tunnistejärjestelmien vanhoista NBN-tunnuksista saadaan URN:NBN-tunnuksia ohjelmallisesti, lisäämällä paikalliseen tunnukseen URN:NBN:-merkkijono ja maakoodi. URN:NBN-tunnuksia voidaan luoda myös ohjelmallisesti.

Esimerkiksi URN:NBN-tunnuksessa

<http://urn.fi/URN:NBN:fi-fd2010-00003245>

NSS on Kansalliskirjaston koneellisesti luotu digitoidun aineiston tunnus fd2010-00003245, johon on lisätty Suomen maakoodi fi.

NSS-osiossa sallitut merkit määräytyvät sovelletun tunnistestandardin mukaan nimialuekohtaisesti, mutta RFC 8141:n asettamissa rajoissa. Esimerkiksi ISBN-nimialueella käytetään ISBN-standardissa määriteltyjä merkkejä, jotka ovat kaikki RFC 8141:n mukaan sallittuja. DOI-nimialueella voi käyttää Unicode-merkkivalikoimaa, mutta URN-standardin kieltämät merkit pitää koodata ns. percent encoding -menettelyn avulla. Esimerkiksi välilyönnistä tulee koodattuna %20.

Kun URN-tunnus esitetään hyperlinkkinä, merkistöä koskevat myös URI-standardin (RFC 3986) rajoitukset. URI-tunnuksessa ei esimerkiksi saa käyttää merkkejä "?" ja "#" sellaisenaan, koska niillä on erityisrooli. Merkkiä "?" seuraa aina URI query ja merkkiä "#" URI fragment. Nämä ja muut erityisroolin omaavat merkit hyperlinkeissä koodattuna^[3]. Tällöin esimerkiksi kysymysmerkistä tulisi URN-tunnuksen NSS-osan sisällä %3F ja #-merkistä %23. Mutta jos # on normaalissa roolissaan F-komponentin erotinmerkinä, sitä ei saa koodata.

Toisin kuin URI-syntaksin fragmentilla ja queryllä, URN-järjestelmän F- ja Q-komponenteilla ei ole roolia identifioinnissa (mistä syystä niille on annettu RFC 8141:ssä uudet nimet). Käytännössä tämä merkitsee sitä, että kirjan lukuun viittaavan F-komponentin lisääminen URN:ISBN-tunnukseen ei muuta tunnuksen NSS-osaa. Siksi URN johon on lisätty F-komponentti, kuten

<http://urn.fi/URN:NBN:fi-fd2010-00003245#Luku2>

identifioi edelleen saman julkaisun kuin yllä olevan esimerkin URN, mutta F-komponentin avulla selain siirtää käyttäjän luvun kaksi alkuun sen jälkeen kun selain on noutanut julkaisun. F-komponentti osoittaa siis vain tietyn paikan identifioitua julkaisun sisällä. F-komponentti toimii tarkoitetulla tavalla vain silloin, kun linkki ohjautuu suoraan identifioituun julkaisuun eikä esimerkiksi laskeutumissivulle, kuten esimerkissä.

URN-määrittelyssä F-komponentilla ei saa olla roolia identifioinnissa, koska se muuttaisi perinteisten tunnistejärjestelmien toimintaa. Kun URI koostuu ISBN-tunnuksesta jonka perään on lisätty kirjan lukuun kaksi viittaava fragmentti, ISBN muuttuu RFC 3986:n mukaan kirjan tunnuksesta luvun kaksi tunnukseksi. Mutta jos URN:ISBN-tunnukseen lisätään F-komponentti, identifioitu kohde on edelleen koko kirja, koska F-komponentti ei ole osa URN-tunnuksen NSS-osiota eikä siis identifioi mitään.

Q-komponentin avulla voidaan lähettää komentoja suoraan haluttua palvelua tarjoavaan kohdejärjestelmään kuten vaikkapa Melinda-tietokantaan. Tunnus on tällöin muotoa

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

Q-komponentti voi olla esimerkiksi tunnukseen perustuva SRU-haku, jolla noudetaan identifiointia julkaisua kuvaava MARC-tietue. Voidakseen toimia palvelun tuottamisen välikätenä URN-resolverin pitää tietää, minkä tietokannan SRU-palvelimelle haun voi lähettää. Jos URN:ISBN-tunnukseen sisältyy suomalainen ISBN-tunnus, Melindan SRU-palvelin on sopiva kohde. Jos ISBN ei ole suomalainen, haun voi kohdistaa johonkin kansainväliseen yhteisluetteloon.

Perinteisten tunnistestandardien kannalta RFC 3986:n oudoin piirre on, että URI-tunnuksen ja URI queryn yhdistelmä identifioi haun tuottaman objektin. Query-haussa käytetty ISBN voi siis yllättäen muuttua kohdetietokannasta löydetyn, identifiointia kirjaa kuvaavan MARC-tietueen tunnukseksi. Tätä perinteisten tunnistejärjestelmien toimialan hallitsematonta laajennusta oli tietenkin mahdoton hyväksyä URN-tunnistestandardissa, ja siksi URI query korvattiin URN-tunnisteen Q-komponentilla, joka ei identifioi mitään. URI query ja Q-komponentti eroavat toisistaan myös siinä, että query alkaa merkillä "?", Q-komponentti merkkijohdella "?=". URN-syntaksi on yhteismitallinen RFC 3986:n kanssa, mutta siihen oli pakko lisätä myös ominaisuus, jota URI-määrittämissä ei ole.

URI-syntaksi ei ota huomioon sitä, että asiakasohjelman ja kohdejärjestelmän välissä voi olla PID-tunnisteen resolverisovellus. Koska URI-syntaksiin ei sisälly mahdollisuutta lähettää komentoja resolverille, URN-tunnukseen lisättiin R-komponentti. Sen erottaa tunnuksen NSS-osiosta merkkijohdella "?+". R-komponentin avulla haku voidaan kohdistaa resolverille. Siltä voidaan pyytää periaatteessa mitä vain, kuten lista identifioidun julkaisun kopioiden URL-osoitteista. Se, pystyykö resolveri vastaamaan itse vai pitääkö sen lähettää pyyntö edelleen johonkin muuhun tietojärjestelmään, riippuu muun muassa resolverille tallennetuista metatiedoista.

Esimerkkinä käytetty URN näyttää R-komponentin kera tältä:

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

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

PID-järjestelmiin ei olisi tarvinnut rakentaa näitä erillisratkaisuja, jos URI-syntaksi olisi ottanut resolverisovellusten tarpeet huomioon. RFC 3986 ohittaa ne todennäköisesti -tarkoituksella.-

7.4 URN-tunnistejärjestelmän palvelut ja niiden toteuttaminen

PID-järjestelmien vastustajat pitävät tunnuksen resoluutiota ylimääräisenä askeleena käyttäjän ja identifioidun resurssin välissä. Mutta resolverit voidaan nähdä myös palvelutarjonnan välineenä ja keinona irrottaa verkkoresurssien pysyvä tunnus ja sijaintipaikka toisistaan, samalla tavalla kuin kirjastot ovat perinteisesti erottaneet julkaisujen tunnukset ja paikanmerkit (signumit). Kirjastossa on toki mahdollista pitää aineisto pitkään samassa paikassa, ja keskiaikaisissa luostarikirjastoissa kirjat olivat tätä varten kettingeillä hyllyissä kiinni. Webissä kettingin vastine on URI, koska se edellyttää resurssin osoitteen pitämistä toiminnallisena.

Jos kirjat olivat hyllyissä kiinni, kirjastolla oli vain rajoitetut mahdollisuudet palvelujen tarjoamiseen. Sama ongelma on URI-linkityksessä, koska se voi tarjota palveluita vain HTTP-protokollan asettamissa rajoissa. PID-tunnuksilla rajat asettaa resolveiden tarjoama toiminnallisuus.

Q- ja R-komponenttien avulla URN-tunnistejärjestelmään voidaan lisätä palveluita periaatteessa rajattomasti. Resolveri voi itse tarjota palveluita tallentamiensa metatietojen avulla, ja se voi linkittää tunnuksen paitsi identifiointiin resurssiin itseensä, myös mihin tahansa kohteeseen joka jollakin tavoin liittyy kyseiseen resurssiin. Tämä luo mielenkiintoisia mahdollisuuksia linkitetyn datan kehittäjille. URL-osoitteiden korvaaminen tietueissa URN-tunnuksilla tai muilla PID-tunnuksilla helpottaa linkkien ylläpitoa, koska resolvereilla linkkien ylläpito on keskitettyä ja se voidaan myös automatisoida.

Toistaiseksi resolvereiden palveluvalikoima on niukka, ja URN-resolvereiden tilanne on heikompi kuin muiden. Toisin kuin ARK- ja Handle-järjestelmillä, ei ole olemassa yhteistä URN-resolveria, vaan pelkästään URN:NBN-resolvereita on useita, eivätkä ne ole keskenään yhteismitallisia. Uusien palvelujen kehittämistä haittaa myös se, ettei R-komponentin syntaksia ja semantiikkaa ole vielä määritelty. Tämä kehitystyö pitää käynnistää, jotta standardisoinnin puutteesta ei tule pullonkaulaa.

Uusien palveluiden tukeminen voi olla yksinkertaista, jos tarvittavat metatiedot ovat resolverilla. Jos käyttäjä haluaa tietää esimerkiksi jonkin julkaisun kaikkien kopioiden verkko-osoitteet, asiakasliittymä voi lähettää R-komponentin avulla resolverille vastaavan pyynnön. Resolveri palauttaa asiakkaalle halutut URL-osoitteet esimerkiksi HTML-sivuna. Ja jos resolverit kykenevät vaihtamaan tietoja standardirajapinnan kautta, osoitteita ja muita metatietoja voidaan myös vaihtaa resolvereiden kesken.

Tietojen hakuun muista järjestelmistä voidaan käyttää Q-komponenttia. Jos käyttäjä haluaa kotimaisen kirjan metatiedot, resolverille lähetettävään URN: ISBN-tunnukseen lisätään Q-komponenttina SRU-standardiin perustuva hakulause, jossa hakuheutona on kirjan ISBN. Resolveri muuttaa Q-komponentin URI queryksi ja uudelleenohjaa haun Melindan SRU-palvelimelle, joka toimittaa kirjan metatiedot asiakkaalle.

R-komponentin ja Q-komponentin perusero on se, että resolveri vastaa pyyntöön oletusarvoisesti itse. Se voi kuitenkin välittää pyynnön eteenpäinkin. Jos resolverilta pyydetään identifioidun julkaisun metatiedot eikä sillä ole niitä, se voi muuttaa R-komponentin sopivan kohdejärjestelmän ymmärtämään muotoon, kuten SRU-hauksi. Jos kohdejärjestelmän tukema rajapinta muuttuu SRU:sta joksikin muuksi, R-komponenttia käytettäessä vain resolverilla tehtävää muunnosta R-komponentista hakulauseeksi pitää päivittää. Mutta jos käytetään Q-komponenttia, jonka resolveri aina toimittaa sellaisenaan suoraan kohdejärjestelmään, hakulauseeseen päivitys pitää tehdä jokaisessa resolveria käyttävässä asiakasohjelmassa erikseen.

On vaikeaa ennakoita, mitä palveluita ja missä järjestyksessä URN-resoluutiopalvelimiin tullaan lisäämään. Sama ongelma koskee myös muita PID-järjestelmiä; yksikään niistä ei ole julkistanut suunnitelmiaan. Palvelutarjonnan kehittämiseen liittyy myös laajempia kysymyksiä resolvereille tallennettavista ydinmetatiedoista sekä resolvereiden välisen tiedonsiirron rajapinnoista. Näitä haasteita pitäisi mahdollisuuksien mukaan ratkoa eri PID-järjestelmien yhteistyönä.

7.5 URN julkaisujen tunnisteena

URN perustuu olemassa oleviin tunnistejärjestelmiin. Jos kirjalla on ISBN, sen URN-tunnuksen on oltava URN:ISBN. URN-tunnukseen otettavaa ISBN:ää ei tarvitse muokata eikä päivittää. Toisin kuin DOI-pohjaisessa 13 merkin mittaisessa ISBN-A -tunnuksessa, myös vanha 10 merkin mittainen ISBN-tunnus kelpaa.

Kirjastoalalla käytettyjen tunnistejärjestelmien soveltuvuutta URN-tunnuksiksi arvioitiin jo yli 20 vuotta sitten (Using Existing Bibliographic Identifiers as Uniform Resource Names, RFC 2288^[iv]). Mitään merkittäviä ongelmia ei havaittu, mutta siirtyminen teoriasta käytäntöön paljasti eroja eri tunnistejärjestelmien välillä.

Semanttinen PID-tunnus kertoo, mistä sen resolveri löytyy. Ei-semanttinen ("tyhmä") PID-tunnus voidaan resoluoida vain yhdessä paikassa, koska tunnus ei anna mitään vinkkiä siitä, mistä päin Internetiä resolveri löytyy. "Tyhmille" URN:ISSN-tunnuksille tämä ei ole ongelma, koska kaikki URN:ISSN-tunnukset resoluoidaan kansainvälisen ISSN-keskuksen ISSN-tietokannassa. ISBN-tunnisteella mitään vastaavaa tietokantaa ei ole, mutta koska tunnus on semanttinen, oikea resolveri on yleensä mahdollista löytää. Esimerkiksi 978-951- ja 978-952 -alkuiset URN:ISBN-tunnukset voidaan ohjata Suomeen Kansalliskirjaston resolverille. 978-3 -alkuiset ovat hankalampi tapaus, koska ne pitää ohjata joko Itävallan tai Saksan kansalliskirjaston resolverille. Jälkimmäinen järjestelmä kattaa myös Sveitsissä annetut 978-3 -alkuiset tunnukset.

Edellä esitetty rajaus koskee vain tulevaisuutta, jolloin URN-tunnukset voidaan esittää hyperlinkkein ilman resolverin osoitetta. Vain tällöin tunnukset ovat täysin teknologiarippumattomia, mikä takaa niiden toiminnallisuuden hyvin pitkälle tulevaisuuteen. Tämän tavoitteen vuoksi tulisi pidättäytyä ratkaisuihin, jotka eivät salli resolverin osoitteen poistamista. Esimerkiksi URN:ISSN-tunnukset voitaisiin käsitellä Kansalliskirjaston resolverilla, jos tunnukset tallennettaisiin muodossa <http://urn.fi/urn:issn:<issn>>. Mutta kun resolverin osoite poistetaan, ja URN:ISSN-resolverin osoite noudetaan Internetin nimipalvelusta, resolveriksi tulee kansainvälisen ISSN-keskuksen ylläpitämä järjestelmä. Ei ole mitään toimivaa keinoa rajata suomalaisten ISSN-tunnusten resoluointi Kansalliskirjastolle.

Kun resolverin osoite on poistettu URN-järjestelmä ei pysty käsittelemään ei-semanttisia tunnisteita, joilla on monta resoluutiopalvelua. Jos esimerkiksi ISBN-tunnuksista tehtäisiin ISSN:n kaltainen "tyhmä" numerosarja, uusien URN:ISBN-tunnusten toiminnallisuutta ei enää voisi taata. Tästä syystä URN-tunnisteen käyttäjien on tarpeen seurata tarkoin esimerkiksi ISBN-standardin kehitystä. Muille PID-järjestelmille rakenteeton ISBN ei ole ongelma, koska näiden tunnuksien etuliite osoittaisi resolverin sijainnin.

URN-nimialueet rekisteröitiin eniten käytetyille julkaisujen tunnisteille pian URN-syntaksin määrittelyn jälkeen: ISSN-rekisteröinti vuonna 2000 ja ISBN sekä NBN vuonna 2001. Kaikki edellä mainitut nimialueet on sittemmin päivitetty 2017 julkaistun uuden URN-syntaksin mukaisiksi. NBN-nimialueen rekisteröintiä (<https://www.iana.org/assignments/urn-formal/nbn>) täydentää RFC-julkaisu Using National Bibliography Numbers as Uniform Resource Names (RFC 8458, <https://tools.ietf.org/html/rfc8458>), joka on NBN-tunnistejärjestelmän virallinen kuvaus. IETF katsoi sen julkaisemisen tarpeelliseksi, koska NBN-tunnisteista ei ollut olemassa aiempaa kuvastusta. Tätä samaa toimintamallia voidaan ainakin periaatteessa soveltaa myös ARK-tunnisteseen.

URN:ISBN- ja URN:NBN-tunnuksia on käytetty nimialueiden rekisteröinnistä lähtien. Jaettujen tunnisteiden määrästä ei ole olemassa tilastoja. ISSN-nimialueen testikäyttö alkoi vuonna 2021, ja tarkoitus on siirtyä tuotantoon 2023. Kansainvälinen ISSN-keskus käyttää resolverina omaa versiotaan Kansalliskirjaston ohjelmistosta. Sen avulla voidaan tarjota ISSN-tietokannan eli portaalin mahdollistamia erityispalveluja kuten linkityksen lehden kotisivulle tai saman lehden eri manifestaatioiden metatietoihin.

7.6 National Bibliography Number (NBN)

Kirjastojen kannalta URN-nimialueista ehkä mielenkiintoisin on kansallisbibliografian ID-tunnus eli National bibliography number (NBN). Sen avulla on mahdollista käyttää kansalliskirjastojen ja niiden kanssa yhteistyötä tekevien organisaatioiden sisäisiä tunnistejärjestelmiä globaalisti ainutkertaisina toiminnallisina tunnisteina. Kansalliskirjastolla näitä yhteistyökumppaneita on muutamia kymmeniä, joista aktiivisia oli kesäkuussa 2021 yhdeksän.

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

NBN-tunnuksia käyttävät kaikki kansalliskirjastot, mutta ne päättävät itse, alkavatko soveltaa myös URN:NBN-tunnuksia. Niitä hyödyntävät kirjastot vastavaavat tunnuksien pysyvyydestä ja paikallisesta ainutkertaisuudesta. URN:NBN-tunnukset ovat oletusarvoisesti toiminnallisia, mutta pakkoa siihen ei ole.

Kansalliskirjastossa NBN-tunnuksen osia olivat aiemmin f-kirjain, vuosiluku kahden tai neljän numeron tarkkuudella ja juokseva numero. Tunnuksen soveltamisalueen laajentamisen myötä f-kirjaimen jälkeen on alettu lisätä muita kirjaimia. Esimerkiksi merkintä "fe" kertoo, että Kansalliskirjaston NBN-tunnus kuuluu syntyjään elektroniselle julkaisulle. Vastaavasti "fd" tunnuksessa fd2010-00003245 osoittaa, että kyseinen julkaisu on digitoitu.

Kansalliskirjastojen antamien NBN-tunnusten ainutkertaisuus on taattu vain kansallisella tasolla. Koska mitään kansainvälistä NBN-tunnusten muodostamismenetelmää tai keskitettyä valvontaa ei ole, eri kirjastot voivat antaa aineistoilleen samoja tunnisteita. Esimerkiksi Saksan kansalliskirjasto voisi antaa tunnuksen fd2010-00003245 jollekin kokoelmassaan olevalle julkaisulle. URN:NBN-tunnuksiin lisätään sen vuoksi ainutkertaisuuden varmistava ISO-standardin mukainen maakoodi. Esimerkiksi Kansalliskirjaston hallinnoima Suomen NBN-aliavaruus on URN:NBN:FI, ja Saksan kansalliskirjasto hallinnoi NBN-aliavaruutta URN:NBN:DE. Kansalliskirjasto voi jakaa oman aliavaruutensa pienempiin osiin, ja delegoida tällä tavoin tunnistejakelua muille organisaatioille. Saksassa näitä kansallisia aliavaruuksia on yli 450 (<https://nbn-resolving.org/namespaces>).

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

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

Perinteisten tunnisteiden soveltamisaloja ei voi laajentaa niin, että ne kattaisivat verkkoaineistot. Käytettävissä olevat tunnusavaraukset ovat liian pieniä (esimerkiksi ISSN-tunnuksia voidaan antaa korkeintaan 10 miljoonaa), eivätkä henkilöresurssit riitä tunnuksien manuaaliseen jakeluun eivätkä identifiointien aineistojen kuvailuun. Siksi URN:NBN-tunniste ja muut automaattisen tunnusjakelun mahdollistavat PID-järjestelmät ovat välttämättömiä, jotta esimerkiksi kaikki pitkäaikaissäilytettävät aineistot saavat pysyvän tunnusteen.

7.7 URN:n tulevaisuuden näkymät

URN-järjestelmällä on muita PID-järjestelmiä paremmat mahdollisuudet tulla kiinteäksi osaksi Internetin perusrakennetta, koska se on Internet-standardi. Tunnisteen suosiota edistää myös maksuttomuus, joka on ehto sille että esimerkiksi esineiden Internetin vaatimat miljardit tunnukset on mahdollista luoda kesäkuussa 2021 julkaistussa Internet-standardissa RFC 9039 kuvatulla tavalla^[vii]. Maksuttomuus ei koske nimialueen rekisteröintyä tunnistejärjestelmää. DOI-tunnus ei ole ilmainen, mutta sen jälkeen kun DOI on ostettu, sen muuttaminen URN:DOI-tunnukseksi ei enää saa maksaa mitään.

IETF:n kannalta URN on ongelmaton tunniste myös siksi, että se ei Handlen tavoin tarjoa vaihtoehtoja ratkaisua Internetin nimipalvelulle ja Webille. Mutta kun nimipalvelu aikanaan korvautuu jollakin muulla teknologialla tai kun HTTP-protokolla joskus kaukaisessa tulevaisuudessa ei ehkä enää toimi, URN sopeutuu myös niihin, koska tunniste on teknologiariippumaton. Muutoin sen elinkaari ei olisikaan periaatteessa rajaton. On tietenkin mahdollista että uudet tekniikat tehdään tästedes aina alaspäin yhteensopiviksi, mutta Internetin tähänastinen historia ei tue tätä ajatusta. Ja on tyhmää tehdä tunnistejärjestelmästä ehdoin tahdoin teknologiariippuvainen, jos se voidaan helposti tehdä myös teknologiariippumattomaksi.

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

URN-tunnisteen hajautettu rakenne on myös etu. Nimialueisiin perustuvan rakenteen ansioista muut pysyvät tunnisteet voivat käyttää URN-järjestelmää hyväkseen. Ensimmäiset askeleet tähän suuntaan on jo otettu: DOI-tunnus voidaan jo nyt esittää myös URN-tunnukseksi, ja DOI:n käyttämä Handle.Net -sovellus pystyy resolverimaan molemmat tunnukset. Esimerkiksi DOI-tunnuksesta <https://doi.org/10.23978/inf.65188> voidaan tehdä URN-tunnus <https://doi.org/urn:doi:10.23978/inf.65188>. Handle-tunnusten etuliitteeksi tulee todennäköisesti HDL.

Tekniikka eteni nopeammin kuin standardisointi, sillä URN-nimialueen "DOI" rekisteröinti DOI-tunnisteelle valmistui vasta maaliskuussa 2023 (<https://www.iana.org/assignments/urn-formal/doi>). Handle-tunnisteen nimialuerekisteröintiä ei ole tätä kirjoittaessa virallisesti edes aloitettu. ARK-tunnisteen osalta keskustelu nimialueen rekisteröinnistä on aloitettu yhteisön sisällä maaliskuussa 2023, mutta toistaiseksi ei ole tietoa siitä, johtaako tämä keskustelu käytännön toimiin.

PID-tunnisteiden käyttäjät tuntevat yleensä hyvin vain oman järjestelmänsä. Siksi muiden PID-tunnisteiden soveltajat eivät välttämättä ole täysin perillä siitä, miten URN-nimialueen voi rekisteröidä, ja mitä seurauksia siitä mahdollisesti olisi. Tiivistetysti voi sanoa, että haittoja ei ainakaan tule, koska nimialueen rekisteröinti ei maksa mitään, eikä rekisteröityä nimialuetta ole pakko ottaa käyttöön. Tällöin rekisteröinnin avulla vain varmistetaan, ettei jokin tunnistejärjestelmä ota käyttöön jollekin toiselle luontevasti kuuluvaan Namespace Identifier -tunnusta. Hyödyiksi voi laskea mahdollisuuden soveltaa URN-syntaksissa määriteltyjä ominaisuuksia kuten F-, Q- ja R-komponentteja, sekä sen, että URN-tunnuksina esitettyinä kaikki tunnukset ovat tunnistettavissa. Aika näyttää, pidetäänkö näitä etuja niin merkittävänä, että muiden PID-tunnisteiden esittämisestä URN-tunnisteena tulee yleinen käytäntö.

^[i] ks. https://en.wikipedia.org/wiki/Universally_unique_identifier

^[ii] <https://www.iana.org/assignments/urn-namespaces/urn-namespaces.xhtml>

^[iii] <https://en.wikipedia.org/wiki/Percent-encoding>

^[iv] <http://www.ietf.org/rfc/rfc2288.txt>

^[v] <http://verkkoarkisto.kansalliskirjasto.fi/va/>

^[vi] <https://fi.wikipedia.org/wiki/MD5>

^[vii] <https://www.rfc-editor.org/info/rfc9039>