

PYSYVÄT TUNNISTEET JULKISESSA HALLINNOSSA

YTI-SELVITYS 2018

Versio 1.01

SISÄLTÖ

PYSYVÄT TUNNISTEET JULKISESSA HALLINNOSSA	1
YTI-SELVITYS 2018	1
Johdanto	2
Tunnusten ehdottomat vaatimukset:	2
Tunnusten muita ominaisuuksia:	3
Tausta ja tavoitteet	6
Pysyvät tunnisteet Euroopan unionissa	7
Semantiikan ongelmat ja tunnisteiden muodostaminen	11
Tilanne Suomessa	15
Toiminnallisten tunnisteiden käyttö Suomessa	20
URN	20
DOI	20
Handle	20
Cool URI –osoitteiden käyttö tunnisteina	21
Koodistot ja sanastot	21
Yhteenveto	21
Työryhmän suosituksia	23
Tunnisteiden hallinnointi	23
Kansallisen tason hallinnointi	24
Kansallinen asianhallinnan tunnistepalvelu	25
Toiminnalliset tunnisteet	26
Tiedon linkittäminen	26
Määritelmät	27

Johdanto

Tiedon jäljitettävyyden ja luotettavuuden ovat ehdottomia edellytyksiä hyvälle hallinnolle. Siirtyminen digitaalisiin palveluihin ajankohtaistaa tiedon yksilöinnin tarpeen myös pitkällä aikavälillä ja järjestelmistä riippumattomalla tavalla. Jotta kansalainen voi luotettavasti seurata asiointinsa etenemistä ja tietojensa hallintaa, on tiedon oltava todennettavissa ja jäljitettävissä. Myös tietosisältöjen on oltava yksilöityjä ja niiden elinkaarta on hallittava suunnitelmallisesti. Tunnisteiden käyttö on tällöin välttämätöntä.

Yksittäinen tunnus identifioi aina jotakin: esimerkiksi julkaisun, asiakirjan, henkilön tai organisaation. Niitä koskevan informaation paikallistamisen ja saatavuuden takaamiseksi tarvitaan suunnitelmallisuutta ja yhteisiä pelisääntöjä – sekä pysyviä tunnisteita. Tässä selvityksessä käsittelemme tätä luotettavan yksilöinnin haastetta. Nykytilan kartoituksen ohella selvityksessä esitetään keinoja tunnistejärjestelmien ja tunnusten käytön tehostamiseksi.

Yleiskielessä käytetään vaihtelevasti termejä tunniste ja tunnus. Tässä raportissa käytetään määritelmää, jonka mukaan tunnistejärjestelmistä jaetaan tiettyjen määriteltyjen käytäntöjen mukaisesti yksittäisiä tunnuksia. Pysyillä tunnisteilla (persistent identifier) tarkoitetaan ainutkertaisia ja pysyviä tunnuksia jotka ovat toiminnallisia Internet-verkossa. Työryhmä katsoo, että pysyvät tunnisteet ovat tietohallintolain (634/2011) mukaisia yhteentoimivuuden kuvauksia ja määrittelyksiä.

Tunnusten ehdottomat vaatimukset:

Pysyvyys tarkoittaa, että samalle asialle tai kohteelle ei anneta uutta tunnusta. Pysyviä tunnuksia hallinnoidaan koko identifioitun kohteen elinkaaren ajan ja vielä sen päätyttyäkin: identifioitu kohde voidaan poistaa käytöstä, mutta *tunnusta ei tuhota eikä käytetä uudestaan*.

Pysyvyys on ehdoton vaatimus ja se vaatii sekä tunnistejärjestelmän että tunnusten dokumentointia. Tunnuksen pysyvyys on käytännössä lupaus siitä, että identifioitu kohde on löydettävissä ja hyödynnettävissä koko sen elinkaaren ajan.

Julkisen hallinnon tiedonhallinnan näkökulmasta tunnusten pysyvyys on keskeinen hyvän hallinnon edellytys, koska viranomaisen tulee huolehtia sekä tarjoamansa tiedon oikeellisuudesta, saavutettavuudesta että jäljitettävyydestä.

Ainutkertaisuus tarkoittaa, että kerran annettua tunnusta ei koskaan anneta uudestaan jollekin toiselle objektille. Tämä voidaan aikaansaada hallinnoimalla tunnistejärjestelmää joko keskitetysti tai hajautetusti.

Tunnusten muita ominaisuuksia:

Resoluutiolla tarkoitetaan identifioituun kohteeseen liittyvien palvelujen tarjoamista. Palvelu voi olla esimerkiksi Internet-resurssin tai kuvaavien metatietojen URL-osoitteen hakeminen tai metatietojen lähettäminen (jos ne on tallennettu resoluutiopalveluun, kuten ARK-tunnistejärjestelmässä).

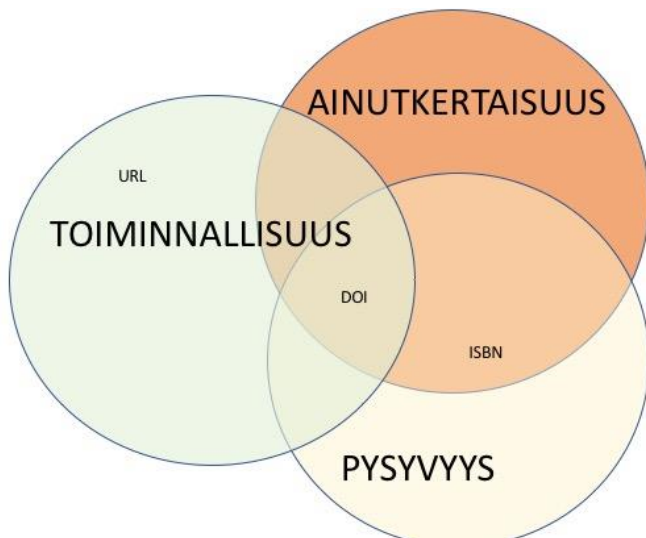
Resoluutiopalvelun antama URL-osoite ei välttämättä ole lopullisen kohdesivun URL-osoite, vaan osoitteessa käytetty protokolla kuten HTTP voi uudelleenohjauksella siirtää käyttäjän resurssin oikeaan URL-osoitteeseen.¹

Resoluutiopalvelu (resolveri) on toiminnallisten tunnisteiden resolvoinnista vastaava sovellus, kuten Handle system, jota käytetään sekä Handle- että DOI-järjestelmissä (Digital Object Identifier). URN-tunnuksella (Uniform Resource Name) ei ole yhtä yhteistä sovelluspakettia, vaan periaatteessa jokaisella URN-tunnuksia jakavalla taholla on oma resolverisovellus. Resoluutiopalvelu voi sisältää monenlaista älykkyyttä ja lisäpalveluita, joten on syytä valita tarpeita parhaiten palveleva luotettava palvelu.

Tunnuksen **semanttisuus** tarkoittaa, että sillä on tulkittavissa oleva rakenne. Esim. kirjan ISBN-tunnuksesta (International Standard Bibliographic Number) voi päätellä sen julkaisumaan ja kustantajan.

Tunnisteita kehittävät organisaatiot eivät ole yksimielisiä siitä, pitäisikö tunnisteiden olla semanttisia vai ei. Esimerkiksi kirjan ensipainoksen ISBN-tunnukseen sisältyvä kustantajatieto vanhenee, kun kirjan kustantaja muuttuu. Ei-semanttisten tunnusten resolvointi voi olla vaikeaa, koska tunnus ei anna mitään vinkkiä, jonka nojalla resoluutiopalvelu voidaan Internetistä löytää. Resolvoitavissa olevaa tunnistetta kutsutaan **toiminnalliseksi**. Perinteiset tunnisteet kuten kirjan ISBN eivät sellaisenaan ole toiminnallisia, vaikka niiden avulla voi tehdä esimerkiksi Google-hakuja. Mutta monet perinteiset tunnisteet voidaan esittää myös toiminnallisena tunnisteena, eli esimerkiksi URN- tai DOI-tunnisteena.

¹ Joe Wass, URLs and DOIs; a complicated relationship. Blog. Crossref. (4.11.2016.)
<https://www.crossref.org/blog/urls-and-dois-a-complicated-relationship/>



Kuva 1. Tunnisteita on erilaisia ja niihin liittyy erilaisia määritelmiä ja palveluita. Valittaessa tunnistejärjestelmää on syytä huolellisesti kartoittaa tarpeet ja suunnitella pitkäjänteisesti.

Tunnistejärjestelmän **katteella** tarkoitetaan sen avulla identifioitavien kohteiden joukkoa, joka on yleensä rajattu. ISBN-tunnuksilla voi identifioida vain kirjoja, ja ORCID-tunnuksilla (Open Researcher and Contributor Identifier) tutkijoita. Standardoiduissa tunnistejärjestelmissä kate määritellään yleensä jo itse standardissa. Katteen rajaaminen helpottaa tunnistejärjestelmän hallinnointia.

Merkittäviä tunnistejärjestelmiä joiden kate on rajaamaton, ovat esimerkiksi ISO:n OID, Handle, DOI, UUID (Universally Unique Identifier) sekä niin sanottu Cool URI. Niiden tehokas soveltaminen ja päällekkäisyyksien välttäminen edellyttää katteen rajaamista käytännön tasolla. Mitä tahansa edellä mainituista tunnistejärjestelmistä voitaisiin periaatteessa käyttää opiskelijoiden julkisena tunnisteena, mutta Suomessa voitaisiin yhteisesti sopia OID-pohjaisen opiskelijanumeron käyttämisestä opiskelijoiden julkisena tunnisteena. Rajausta ja yhteiset sopimukset estävät myös päällekkäisyyksiä, kuten sitä, että opiskelijoiden identifiointiin alettaisiin käyttää myös tutkijoille suunnattua ORCIDia.

Julkaisualalla standarditunnisteiden yhteinen kate on varsin hyvä mutta ei täydellinen, eikä kaikkia standardoituja tunnistejärjestelmiä ole otettu laajasti käyttöön. Esimerkiksi tieteelliselle artikkelille voidaan antaa DOI-tunnus, mutta artikkelissa olevalle kuvalle ei ole standarditunnistetta.

Organisaatioiden omat tunnistejärjestelmät eivät yleensä perustu standardeihin. Kehitettäessä ja käyttöönotettaessa paikallisia tunnistejärjestelmiä tulisi kuitenkin aina selvittää mahdollisuus käyttää standarditunnuksia.

Tunnuksilla on usein **käyttökonteksti**, eli niitä käytetään jossain ennalta sovitussa yhteydessä, kuten yksittäisellä hallinnonalalla tai sektorilla. Konteksti voi perustua lainsäädännössä annettuun

velvoitteeseen – esimerkiksi paikkatiedolle on INSPIRE-direktiivin vaatimuksesta käytettävä ns. Cool URI –tunnisteita.

Kontekstin rajausta helpottaa hallinnointia ja mahdollistaa tarkemman ohjauksen esimerkiksi tietomallien suhteen. Kontekstin rajausta ei itsessään kuitenkaan takaa yhteentoimivuutta vaan on syytä huolehtia esimerkiksi tunnistejärjestelmän perusteella tallennettavien metatietojen yhdenmukaisuudesta.

Tunnisteisiin voi liittyä **metatietomalli**. Tunnusta annettaessa identifioitu objekti pitää kuvailla joko standardissa tai muulla tavoin määritellyllä tavalla. Esimerkiksi ISSN-standardi edellyttää, että jokainen tunnuksen saava kausijulkaisu kuvaillaan, ja tiedot lähetetään kansainvälisen ISSN-keskuksen ylläpitämään ISSN-tietokantaan. ISSN –verkosto on laatinut metatietojen tallennuksesta tarkat ohjeet. Metatietomallin kattavuus vaihtelee tunnistejärjestelmästä riippuen. ORCID-tunnuksen saadakseen tutkijan on kerrottava vain etunimensä ja sähköpostiosoitteensa, mutta ISNI-toimijatunnisteen (International Standard Name Identifier) saaminen edellyttää sitä, että kansainvälisessä ISNI-tietokannassa on toimijaa koskevat metatiedot vähintään kolmen eri organisaation lähettämänä. Mitä kattavampi metatietomalli, sen tehokkaampaa on tiedonhallinta. ISNI-tietokannasta on helppoa löytää ”oikea” Albert Einstein tietokannan sisältämien julkaisutietojen avulla, mutta esimerkiksi ORCID-tietokannan Espanjassa opiskeleva Albert Einstein saattaa herättää epäilyksiä. Toisaalta vaatimus kattavien metatietojen tallentamisesta ja monimutkaiset tallennusohjeet tekevät tunnisteiden luomisesta työlästä. Lisäksi esim. toimijoiden tunnisteisiin voi liittyä henkilötietoja kuten syntymäaika, joita ei voida jakaa vapaasti, koska kyseessä on henkilöreisteri.

URI (Uniform Resource Identifier) ja Cool URI –termeillä viitataan URL-osoitteisiin, joiden pitäisi toimia verkossa paitsi osoitteina, myös tunnisteina. Suomessa URI-tunnisteita sovelletaan esimerkiksi paikkatiedossa sekä Finlex-järjestelmässä lakien tunnisteina. Suljetuissa käyttöympäristöissä kuten yksittäisillä palvelimilla, voidaan URL –osoitteita pitää varsin pitkään samoina, mutta ongelmia voi silti syntyä varsin pian esimerkiksi sen vuoksi, ettei Internet-domaineja voi ostaa, vaan vain vuokrata. Jos domainin vuokraaja vaihtuu, uusi toimija ei välttämättä halua järjestää uudelleenohjausta dokumenttien vanhoista osoitteista uusiin.

Pitkällä aikavälillä teknologiariippuvuus voi muodostua kriittiseksi ongelmaksi toiminnallisille tunnisteille. Mikään tietotekninen ratkaisu ei tähän mennessä ole ollut pitkäikäinen, joten riittävän kaukana tulevaisuudessa linkitys voi perustua uusiin, aiemmat ratkaisut korvanneisiin protokolliin. Silloin esimerkiksi HTTP-linkkien takautuva ylläpito voi käydä hankalaksi.

URL-osoitteet eivät itsessään täytä pysyvän tunnisteen perusvaatimuksia, koska ne eivät aina ole pysyviä eivätkä ainutkertaisia. Dokumentit voivat kadota verkosta kokonaan, mihin viitataan **linkkimätä**-termillä (engl. Link Rot). Tunnistekäytön kannalta on ongelmallista jo sekin, että dokumentin osoite voi muuttua ja sama dokumentti voi olla saatavilla samaan aikaan useista eri osoitteista. **Sisältönyrjähdyksellä** (engl. Content Drift) tarkoitetaan tietystä URL-osoitteesta olevan dokumentin sisällön muuttumista, joka usein tapahtuu siten, etteivät käyttäjät tai dokumenttia hyödyntävät tahot saa tietoa muutoksesta.

Tausta ja tavoitteet

Digitalisoidaan julkiset palvelut on yksi pääministeri Juha Sipilän hallitusohjelman kärkihanke.

Tavoitteena on rakentaa julkiset palvelut käyttäjälähtöisiksi ja ensisijaisesti digitaalisiksi toimintatapoja uudistamalla. Digitalisaatio on hallituksen strategian läpileikkaava teema ja se nähdään yhtenä avainkeinoista julkisen hallinnon tuottavuusloikassa. Kärkihankkeessa sitoutetaan julkinen sektori automatisoimaan ja digitalisoimaan toimintatapansa. Tämä selvitys on osa **Yhteinen tiedon hallinta** -hanketta (YTI, VM075:00/2016) joka puolestaan on **Digitalisoidaan julkiset palvelut** -hankkeen toimenpidekokonaisuus. YTI:n keskiössä ovat tieto, tietosisällöt ja tiedon hyödyntäminen.

Tämän työryhmän työn tavoitteena on ollut tukea julkisen hallinnon digitalisointumiskehitystä ja tehostaa erityisesti toimintatapojen ja tiedon yhteentoimivuutta. Hallitusohjelmaan kirjattu tavoite kysyä samaa tietoa vain kerran niin kansalaiselta, yritykseltä kuin muilta viranomaisilta *vaatii tehokkaita yhteisiä tiedon hallinnan toimintatapoja, joista yksi on pysyvien tunnusten käyttö ja hallinta*. Pysyvien tunnusten käytännön kehittäminen edellyttää sekä yhteistä sopimista että tarvittavien tunnuspalveluiden (ml. hallintapalvelut) tuottamista.

Digitalisoinnin ja sähköisen palvelukehityksen kannalta on keskeistä, että *palvelun tuottamiseksi käsiteltävät monimuotoiset aineistot voidaan yksilöidä ja että liitetyt sisällöt säilyvät luotettavasti*. Tiedon yksilöinti on osatekijänä sekä tiedon laadussa että sen todistusvoimaisuudessa läpi tiedon elinkaaren ja yli organisaatorajojen sekä myös sen jälkeen, kun tietoa ylläpitävät järjestelmät ovat muuttuneet ja/tai aineisto siirtynyt säilytettäväksi. Yksiselitteinen ja luotettava digitaalisessa muodossa olevan *aineiston yksilöinti tukee merkittävästi julkisen hallinnon informaation välittämistä organisaatorajat ylittävissä prosesseissa*. Aineistojen hyödyntäminen mahdollistuu ja tehostuu kun ne voidaan luotettavasti yksilöidä tunnusten avulla. Tavoitetilan saavuttaminen edellyttää, että toimijoilla julkisessa hallinnossa on yhteiset toimintasäännöt tietoaaineistojen yksilöintiin ja yksilöinti integroituu tiedon muodostamisen, hallinnan ja ohjauksen prosesseihin. Oleellista on tiedon ja tietoaaineiston identifiointi sen syntyhetkellä, sillä näin sen jakelu on helpompaa ja luotettavampaa kuin tilanteessa, jossa se on jo ehtinyt kopiaitua.

Pysyvien tunnusten hallinta on osa julkisen hallinnon arvokkaiden tietoaaineistojen ja tietovarantojen käyttöä ja informaatioekosysteemin kasvupohjaa. Tiedon tehokas hyödyntäminen edellyttää siilotumien poistamista ja tiedon jakamista ja yhteistä hyödyntämistä.

Asetettavan työryhmän tehtävänä oli:

- Laatia kokoava nykytilaselvitys julkisessa hallinnossa käytössä olevista tunnus/tunnistejärjestelmistä.
- Määrittää yhteinen tunnus/tunnistekäytäntö julkiseen hallintoon siten, että se tukee ja voi toimia nykyisten eri hallinnonaloilla vakiintuneiden tunnusjärjestelmien rinnalla. Esityksen tulee tukea julkisen hallinnon toimijoiden mahdollisuutta yksilöidä tietoaaineistot tai muut yksilöinnin kohteet jo niiden syntyvaiheessa. Linjaukset ovat yksi tietopoliittisista toiminnoista toimintatapojen ja tiedon yhteentoimivuuden tehostamisessa.

- Laatia ehdotus tunnusten tuottamisen, hallinnan ja tulkinnan edellyttämistä vastuista ja vastuutahoista julkisen hallinnon yhteisen tunnus/tunnistekäytännön hallinnoimiseksi.
- Kuvata tarvittaessa tunnusten/tunnisteiden muodostamiseen ja/tai hyödyntämiseen liittyvät kehitystarpeet uusien palveluiden tuottamiseksi.

Henkilötunnus rajattiin työryhmän tehtävänannon ulkopuolelle.

Pysyvät tunnisteet Euroopan unionissa

Sähköisen tiedon vaivaton ja luotettava liikkuminen Euroopan unionissa on tärkeä digitaalisen yhteismarkkinoiden toimivuuden edellytys. Unionissa yhteentoimivuutta on edistetty ISA-hankkeissa². Tunnisteiden soveltamista on ohjeistettu paitsi rahaliikenteen, myös lainsäädännön, paikkatiedon ja datakatalogien osalta. Vuosina 2014 ja 2015 panostettiin tunnisteiden määrittelyyn ja hallintaan, mutta työ on edelleen ainakin osittain kesken. Ratkaisut tukevat linkitetyn datan tuottamista, mutta perusratkaisuna on Cool URI.



Kuva 2. Joinup-alustalle on kerätty yhteentoimivuutta edistävien hankkeiden tuloksia.

<https://joinup.ec.europa.eu/document/10-rules-persistent-uris>

Kuvassa 1 oleva ohjeistus liittyy URL-osoitteiden saamiseen mahdollisimman pitkäikäisiksi, jota voi parantaa muun muassa soveltamalla olemassa olevia tunnistejärjestelmiä.

Kehittämällä Cool URI -tunnisteiden hallintaa pyritään ratkomaan seuraavia ongelmia:

- Eri tunnisteiden jakaminen samalle asialle
- Identtisten tunnisteiden jakaminen eri asioille
- DOI, UUID ja muiden olemassa olevien tunnisteiden linkittäminen
- Sekava nimiavaruuksien käytäntö
- Palvelulupausten puute
- Sekava terminologia ja väärinkäsitykset

² https://ec.europa.eu/isa2/home_en

- URLien käytön aiheuttamat epäselvyydet³

Suuri osa EU-aineistoista on luonteeltaan sellaisia, ettei niille ole olemassa omaa standarditunnistetta. Silloinkin voidaan käyttää joko laaja-alaista järjestelmää kuten DOI:ta tai Handlea tai kehittää aineistotyyppille oma tunnistejärjestelmä kuten tuonnempaan kuvattava European Legal Identifier (ELI).

EU-tunnisteille ehdotettu Cool URI-syntaksi on seuraavanlainen:

`http://data.europa.eu/{namespace string}/{local id}/{version}/{language}`

”Namespace string” on tässä ilman semantiikkaa ja loppuosa on paikallisesti määriteltävissä. Sen suhteen on kuitenkin ehdotettu seuraavan laista ohjeistusta:

- Älä ilmaise omistajuutta
- Käytä olemassa olevia tunnisteita
- Käytä auto-increment -toiminnallisuutta varoen
- Ole varovainen, jos käytät versionumerointia
- Älä käytä kyselyjä/komentoja
- Ole varovainen, jos käytät tiedostopäätteitä
- Käsittele kieliversioita versioina
- Huomioi kirjainmerkistöt ja monikielisyys
- Käytä HTTP 303 uudelleenohjausta reaaliaikaisen maailman kohteille
- Käytä erityislaitteistoa
- Huolehdi varajärjestelmästä⁴

Tunnistekokonaisuutta hallinnoi EU:n Publications Office. Järjestelmä ei ole vielä kattava eikä ole selvää, mille aineistoille näitä tunnisteita voidaan / pitäisi soveltaa. Tunnistejärjestelmiä on 15⁵, mutta muutamat näyttävät poistuneen käytöstä (linkit niiden kotisivuille eivät enää toimi). Useiden Cool URI-pohjaisten ratkaisujen lisäksi listalla on myös DOI.

Joistakin yleisesti käytetyistä tunnistejärjestelmistä on annettu tarkempaa ohjeistusta. Esimerkiksi INSPIRE-direktiivin yhteydessä⁶ on linjattu (paikkatiedon) tunnistejärjestelmän vaatimuksia:

- Ainutlaatuinen maailmanlaajuisesti
- Pysyvä (dokumentoitava)
- Resolvoituva
- Luotettava
- Provenienssi hallittu
- Kustannustehokas ratkaisu
- Käyttö tuettu organisaation toimintaperiaatteissa
- Organisaation käyttämä tunnusjärjestelmä on hallinnoitu

³ SC8DI07171 D04.02.01.B: Common approach for the management of persistent URIs by EU institutions. Towards a common policy for the governance and management of persistent URIs by EU institutions. Study Report. (Versio 0.53, 2014) s. 1-2.

⁴ SC8DI07171 D04.02.01.B: Common approach for the management of persistent URIs by EU institutions. S.4-5.

⁵ <http://data.europa.eu/URI.html>

⁶ https://joinup.ec.europa.eu/sites/default/files/document/2016-02/are3na-pid_governance_final.pdf

- Yhteentoimiva
- Tarkoituksenmukaisesti karkeistettu
- Versioitu⁷

Tämä ohjeistus koskee erityisesti tunnistejärjestelmän hallinnointia ja se sisältää monia ominaisuuksia joita periaatteessa jokaisella järkevällä järjestelmällä tulee olla. Ohjeistuksen olemassaolo lisää myös yleistä luottamusta tunnistejärjestelmää kohtaan.

Datakatalogien kehittäminen on ollut tärkeä osa yhteistä tiedonhallintaa ja DCAT-metadatatamäärittystä⁸ käytetään yleisesti data-aineistojen kuvailuun. DCAT edellyttää Cool URI:n antamista kaikille ”metadataentiteeteille” mutta suosittelee olemassa olevien tunnisteiden kuten DOI:n käyttämistä data-aineiston tunnisteena (adm:identifier –metadataentiteetti). Konvertoitaessa DCAT-tietueita Dublin Core –formaattiin adm:identifier mapataan dct:identifier –kenttään.

Data-aineiston Cool URI tulee uudelleenohjata (HTTP 303) data-aineiston osoitteen asemesta sen metatietojen osoitteeseen ”kun mahdollista”.⁹ Käytännössä tämä merkitsee yleensä ohjausta ns. landing pagelle. Jos käytetään toiminnallisia tunnisteita, on tulevaisuudessa mahdollista tarjota käyttäjille erilaisia palveluita, joista yksi on metatietojen tarjoaminen.

EU:n sisällä on luotu tunnistejärjestelmät paikkatiedolle ja lainsäädännön tiedoille. Molemmat järjestelmät ovat Cool URI-pohjaisia. Edellistä koskeva tunnisteratkaisu on INSPIRE-käyttötapausten osalta määritelty INSPIRE-direktiivissä, jälkimmäisen osalta EU:n ohjaus ei ole yhtä vahvaa.

URI-pohjaista European Legislation Identifier (ELI) -tunnistetta¹⁰ sovelletaan Suomessa Semanttinen FINLEX –palvelussa (data.finlex.fi/). Myös monet muut maat ovat ottamassa järjestelmää käyttöön. EU:n tavoitteena on, että ELI:llä identifioidaan kaikki Euroopassa virallisesti julkaistut oikeudelliset aineistot. ELI-järjestelmään kuuluu URI-tunnisteen lisäksi metadatatamäärittäminen (ontologia) juridisen informaation kuvailuun sekä metatietojen vaihtformaatti.

ELI on semanttinen tunnus. Se rakennetaan seuraavan mallin mukaan:

```
/eli/{jurisdiction}/{agent}/{sub-agent}/{year}/{month}/{day}/{type}/{natural identifier}/{level 1...}/{point in time}/{version}/{language}
```

Kukin jäsenvaltio on velvollinen rakentamaan tarvitsemansa ELI-tunnukset ottaen huomioon erityiset kielivaatimuksensa. Kaikki tunnuksen komponentit ovat valinnaisia, ja ne voidaan valita kansallisten vaatimusten pohjalta eikä niillä ole ennalta määriteltyä järjestystä. Jotta ELI-metatietoja voidaan vaihtaa, valittu URI-malli on dokumentoitava käyttämällä URI-mallimekanismia (ks. seuraava esimerkki)¹¹. Tekniikka perustuu täysin linkitetylle datalle (dcterms, FRBR, RDA, SKOS), sille on oma

⁷ Vasilescu A et al.

⁸ <https://www.w3.org/TR/vocab-dcat/>

⁹ Wendy Carrara et al., Towards an open government data ecosystem in Europe using common standards ISA 2, 06 June 2017.

¹⁰ <http://eur-lex.europa.eu/eli-register/about.html>

¹¹ Neuvoston päätelmät eurooppalaisen lainsäädäntötunnuksen (European Legislation Identifier (ELI)) käyttöönotosta (2012/C 325/02)

[http://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=CELEX:52012XG1026\(01\)&from=EN](http://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=CELEX:52012XG1026(01)&from=EN)

ontologiansa ja tietovaranto CELLAR, jossa SPARQL Endpoint.¹² Järjestelmässä ei oteta vastuuta sisältöneuvottelusta eli se ei erottele ihmisille tai koneille tarkoitettua sisältöä toisistaan. Ohjeet URI:n muodostamiselle ovat seuraavat:

- 1) Käytä ihmisymmärrettäviä viittauksia
- 2) Älä käytä organisaationimiä tms muuttuvia osia
- 3) Tee tunnuksesta toimiva (ks. alla)

Toimivuus tarkoittaa tässä yhteydessä sitä, että `./eli/{typedoc}` palauttaa listan pyydetystä dokumenttityypin dokumenteista, ja `./eli/{typedoc}/{year}` listan tietyn vuoden vastaavat dokumentit ja niin edelleen.¹³

Paikkatiedon osalta Suomessa on kansallinen JHS 193 –suositus.¹⁴ Sen mukaan URI-pohjainen paikkatiedon tunnus muodostetaan seuraavasti, kansallisella kontekstilla:

`http://paikkatiedot.fi/{tunnustyyppi}/{aineistotunnus}/{paikallinen tunnus}`, jossa

- `tunnustyyppi` = yksilöitävän kohteen tyyppi
 - paikkatietokohteilla `tunnustyyppi` = so
 - reaali maailman kohteilla `tunnustyyppi` = id
 - käsitteillä `tunnustyyppi` = def
- `aineistotunnus` = Paikkatietohakemiston aineistotunnus
- `paikallinen tunnus` = tietoa tuottavan järjestelmän tietokantaan perustettu yksilöivä paikallinen tunnus.

Paikkatietokohteen versiotunnusta käytetään kohteen elinkaarisääntöjen mukaisesti. Versiotunnus ei ole pakollinen. Ellei elinkaarisääntöjä ole määritetty, voi versiotunnukseksi olla esimerkiksi kohteen luontipäivämäärä. (esimerkiksi ISO 8601 -standardin mukainen aikaleima: 2014-01-19 T12:38:31+03:00 tai 2014-01-19).

Reaali maailman kohdetta voi vastata yksi tai useampi paikkatietokohde tai -käsite. Nämä liitetään dokumentaatiotunnukseen. Reaali maailman kohteen tunnukselle tehdään sitä perustettaessa HTTP 303 edelleenohjaus tähän dokumentaatiotunnukseen.

Publications office on piakkoin julkaisemassa selvityksen, joka on tehty arkistojen julkasemista varten. Siinä arvoitiin muun muassa seuraavia asioita:

- käyttötarkoitus ja käyttäjien tarpeet
- laadunvalvonta ja seuranta
- palvelut

¹² ELI ontologia: http://publications.europa.eu/mdr/eli/documentation/ELI_Ontology-v1.1.htm, tekniset ohjeet: serialisation http://eur-lex.europa.eu/eli-register/technical_information.html

¹³ Katso tarkemmin myös ELI implementation methodology. Good practices and guidelines (2015)

¹⁴ JHS 193 Unique identifiers of the geographic information

<http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/193/full>

URI-muodostus:

http://docs.jhs-suositukset.fi/jhs-suositukset/JHS193_liite1/JHS193_liite1.html

- tunnettuus ja näkyvyys

Selvityksessä päädyttiin ehdottamaan yhdistettyä formaattia:

<https://data.europa.eu/doi/10.2760/376572>

Kentän koodi muuttunut

Pysyvät tunnisteet ja linkitetty data

Linkitetyn datan kehittäminen on kohdannut haasteita. Edes pysyvät tunnisteet eivät ole täysin varma ratkaisu linkittämiseen. Esimerkiksi oikeuksien hallintaan käytetyn "info" URI –skeeman¹⁵ kehittäminen ja tuki on lopetettu jo vuosia sitten. Pysyvien tunnisteiden rinnalla sovellettu Cool URI –ratkaisu on mahdollista saada suljetuissa ympäristöissä toimivaksi ainakin joksikin aikaa, mutta koko verkon tasolla on laajoissa tutkimushankkeissa osoitettu URL-osoitteiden olevan epäluotettavia. Jatkuva tekninen kehitys ja elektronisten aineistojen käyttöön ja säilyttämiseen vaikuttavat hallinnolliset muutokset eivät pitkällä aikavälillä ainakaan paranna Cool URI:en tilannetta.¹⁶

Cool URI on helpolta vaikuttava tapa yksilöidä sähköisiä tietoaaineistoja. Käytännön kokemus on kuitenkin osoittanut, että niiden pysyvyyttä uhkaavat muun muassa seuraavat tekijät:

- Organisaation nimenmuutos tai verkko-osoitteen menetys
- Resurssin siirtyminen tai muuttuminen (esim. yhdistyminen)
- Metatietojen muutos
- Resurssin häviäminen¹⁷

W3C ja muut toimijat ovat julkaisseet teknisiä ohjeita URLien soveltamisesta^{18 19}, mutta niiden noudattaminen ei ratkaise sitä hallinnollista ongelmaa, että vain pieni osa verkon dokumenteista on alun perinkään tarkoitettu pitkäaikaiseksi.

Riippumatta verkon dokumenttien elinkaaresta, URI:in sisältyvä laaja tunnisteiden määritelmä on ongelmallinen, koska se hyväksyy myös kelvottomia tunnisteita. Hyvä esimerkki tästä on sähköpostiosoite, joka ei täytä mitään pysyvän tunnisteiden kriteereitä. Samalla ihmisellä voi olla useita sähköpostiosoitteita ja sama osoite voi eri aikoina kuulua eri henkilöille. Sama ongelma koskee myös tiedostoja: sama tiedosto voi löytyä useista osoitteista ja samassa URL-osoitteesta voi eri aikoina olla eri dokumentti. URL-osoitteesta ei yleensä voi nähdä, mitkä niistä ovat "cool" ja mitkä ei (lukuun ottamatta verkkoarkisto-osoitteita, jotka kestävät niin pitkään kun arkisto on olemassa tai käytettävissä). Tuntematta Cool URI:n antaneen organisaation palvelulupausta ja toimintatapaa on vaikea tietää, voiko osoitteeseen luottaa vai ei.

¹⁵ <http://info-uri.info/>

¹⁶ Tutkimusta on paljon, tässä muutamia esimerkkejä: Martin Klein et al., Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. Published: December 26, 2014 <https://doi.org/10.1371/journal.pone.0115253>. Julie McMurry, Lilly WInfree, Melissa Haendel: Bad identifiers are the Potholes of the Information Superhighway. Herbert Van de Sompel & Michael L. Nelson, Reminiscing About 15 Years of Interoperability Efforts. D-Lib Magazine. November/December 2015 DOI: 10.1045/november2015-vandesompel

¹⁷ Vasilescu A et al.

¹⁸ <https://www.w3.org/TR/chips/>

¹⁹ <https://www.w3.org/TR/ldp-bp/>

Standardoituuihin tunnistejärjestelmiin kuten kirjan ISBN-tunnukseen perustuvien toiminnallisten tunnusten käyttö selventäisi tilannetta käyttäjien kannalta, koska standarditunnisteet ovat pysyviä, ainutkertaisia ja niillä on käyttöä ohjaavat yleisesti tunnetut säännöt. Sähköisten aineistojen pitkäaikaissäilytysjärjestelmien ja -prosessien suunnittelu ja kehittäminen on lisännyt tarvetta standarditunnisteiden käytölle mutta myös osoittanut niiden tärkeyden.

Pitkäaikaissäilytyksessä on tärkeää taata aineistojen eheys ja luotettavuus pitkällä aikavälillä, jopa satojen vuosien ajan. Tällöin on identifioidava paitsi tallennettavat objektit eri versioineen, myös arkistointiin osallistuvat toimijat kuten pitkäaikaissäilytysjärjestelmän operaattorit ja heidän käyttämänsä ohjelmistot sekä säilytettäviin objekteihin kohdistuvat tapahtumat kuten migraatiot.

Arkistoitua aineistoa voi kuvailla neljällä eri tasolla: teokset (esim. ”Talousarvioesitys vuodelle 2018”), teosten representaatiot (esim. budjettikirjan PDF-versio) sekä tiedostojen mahdollisesti sisältämät bittijonot. Kaikkia näitä tasoja ei ole pakko käyttää kuvailua tehtäessä. Esimerkkinä on tässä mainittu julkaisu, mutta vastaavaa jaottelua voidaan soveltaa myös esimerkiksi asiakirjoihin tai data-aineistoihin.

Kuvailun kaikilla tasoilla tulee soveltaa pysyviä tunnisteita, joiden avulla esimerkiksi teosten ja niiden representaatioiden (esimerkiksi ”Talousarvioesitys 2018” ja sen EPUB-versio) väliset suhteet voidaan kuvata koneluettavalla tavalla.

Pitkäaikaissäilytyksessä sovelletaan tunnisteina yleensä aina standardeihin perustuvia pysyviä tunnuksia. Esimerkiksi CSC:n ylläpitämässä PAS-palvelussa Cool URI –tunnisteita ei voi käyttää muun muassa sen vuoksi, ettei niiden pysyvyydestä ja ainutkertaisuudesta ole takeita.

Useat organisaatiot ovat luoneet ohjeita tunnisteiden luomiselle ja käytölle. Australian kansallisen datapalvelun (ANDS) ohjeistus on hyvä esimerkki selkeästä ohjeistuksesta.²⁰ ANDS suosittelee ehdottomasti automatisoitua linkkien tarkistamista. Tunnusten on oltava URI-syntaksimäärittelyn mukaisia eli niissä ei tule käyttää RFC 3986:ssa kiellettyjä erikoismerkkejä koodaamattomina. Perusohjeina ANDS suosittelee semanttisten tunnusten välttämistä ja kehottaa antamaan tunnisteita ”löysästi” esimerkiksi roolin ennemmin kuin yksilön mukaan (”webmaster” eikä ”Matti Möttönen @konttori”) tai tietosisällön ennemmin kuin ilmentymän / manifestaation mukaan (jolloin tunnistetta ei tarvitse muuttaa, kun kohteen formaatti, käyttökonteksti tai sijainti muuttuu). Nimien uudelleenkäyttöä ei saa sallia, tunnusten sen sijaan tulee antaa siirtyä kontekstien välillä. Tärkeitä kysymyksiä ANDS:n mukaan ovat:

- Onko objekti itsessään pysyvä?
- Julkaistaanko tunniste?
- Tuleeko se siirtymään palvelinten välillä?
- Tullaanko siihen viittaamaan muualla?
- Millainen on tietomalli, onko sillä väliä?
- Onko objekti vakaa (historiatieto, muutoshistoria)
- Onko objekti hallinnassa/hallussa?
- Milloin tunniste luodaan?

²⁰ ANDS Guides and resources. Persistent identifiers: Expert level. <http://www.ands.org.au/guides/persistent-identifiers-expert>

Hyvän tarkistuslistan tarjoavat myös tutkijat Nicholas, Ward ja Blinco (2009):

1. Selvitä mitä teillä voi tunnistaa, millä ja miten nämä suhtautuvat toisiinsa > tietomalli
2. Selvitä mitä teillä tullaan tunnistamaan (mikä on tärkeää) > sisällytää tunnisteet tietomalliin
3. Varmista, että tunnisteet säilyvät (erota tunnisteiden hallinta tiedonhallinnasta)
4. Varmista mekanismit tunnisteiden ylläpitämiseksi (määrittele työvuot)
5. Varmista tunnisteiden vakaa käyttö (hyödynnä niitä!)
6. Integroi pysyvät tunnisteet järjestelmiin (määrittele pidit osaksi tiedonhallintaa)
7. Luo pitkän aikavälin kestävyys (määrittele sitoutumisenne, kommunikoi se)²¹

Toiminnallinen tunnistejärjestelmä edellyttää käytännössä aina erillisen järjestelmäkohtaisen resoluutiopalvelun käyttämistä. Käyttäjille on periaatteessa mahdollista tarjota erilaisia palveluita, joista yksi on tunnisteiden uudelleenohjaus identifioitujen kohteiden URL-osoitteeseen. Teknisesti se tarkoittaa lisäaskelta, joka kuitenkin mahdollistaa sen, että organisaation identifioimien kohteiden tunnisteet ovat riippumattomia sen omasta tai muiden organisaatioiden hallinnoimista verkkotunnuksista (eli ns. domain-nimistä), joita käytettäisiin Cool URI:ssa. Identifioitujen resurssien verkkotunnuksen sijaan PID-tunnuksissa käytetään tunnistejärjestelmän resolverin, joka voi olla esimerkiksi urn.fi tai doi.org. URN-järjestelmässä verkkotunnuksesta voidaan tulevaisuudessa luopua, ja käyttää pelkkää tunnusta.

Verkkotunnuksen nimominen tunnistejärjestelmään on ongelma, koska verkkotunnuksia ei voi ostaa, vaan vain vuokrata. Verkkotunnus on myös aina protokollasidonnainen. Käyttäjän näkökulmasta ongelmaksi muodostuu se, että on *mahdotonta tietää milloin URI on tarkoitettu pysyväksi eli milloin siihen voi luottaa, ellei asiaa voi todentaa jostakin ulkoisesta lähteestä*. Tällainen ulkoinen lähde voi olla esimerkiksi FINLEX-dokumentaatio tai tieto W3C:n URI-käytänteistä.

Tästä syystä on suositeltavaa, että mikäli Cool URI-tunnuksia on pakko käyttää, niiden muodostaminen tehdään tarkoin määritellysti, dokumentoidusti ja hallitusti. Pysyviksi tarkoitettujen Cool URI-tunnuksien nimiavaruudet tulisi rekisteröidä ja hallinnoida keskitetysti ja niihin sisältyvä normaalia verkkotunnusten rekisteröintiä täydentävä semantiikka on valittava ja ylläpidettävä huolella. On esimerkiksi mieluummin kuvattava tietosisältöjä kuin hallinnollisia rakenteita tai tiedon omistajuutta. Verkkotunnusten rekisterin tulisi olla avoin ja koneluettava, jotta sovellukset pääsevät tarvittaessa tarkistamaan niitä koskevia tietoja.

Näitä ratkaisuja ei voida soveltaa verkossa yleensä eikä kaikkiin URL-osoitteisiin. Cool URI-mallia tulisi käyttää rajatuissa ja hyvin säädellyissä yhteyksissä vain silloin, kun se ei ole vältettävissä. Muulloin on suositeltavaa käyttää toiminnallisia tunnisteita, erityisesti rakennettaessa kansallisia tai kansainvälisiä linkitetyn datan kokonaisuuksia ja erityisesti silloin, jos identifioitava aineisto on säilytettävä pitkäaikaisesti.

Uudelleenohjaus voidaan tehdä keskitetysti esimerkiksi identifiers.org -tai Wikidata-palveluissa. Niissä halutun palvelun verkkotunnus (domain-nimi) ja aineiston tunnus tallennetaan erikseen. Jälkimmäiseen

²¹ Nick Nicholas, Nigel Ward, Kerry Blinco: A Policy Checklist for Enabling Persistence of Identifiers D-Lib Magazine January/February Volume 15 Number 1/2 2009. ISSN 1082-9873 <http://www.dlib.org/dlib/january09/nicholas/01nicholas.html>

sisältyy myös kontekstin tunnus tai nimiavaruus. Rakenteisuus helpottaa tunnuksen hallinnointia ja monipuolistaa palvelutarjontaa. Esimerkki metadataan linkittämisestä REST-pohjaisen rajapinnan kautta:

<http://identifiers.org/rest/metadata/reactome:R-HSA-446203>

Tässä esimerkissä on tietojärjestelmän sisäisestä aineistotunnuksesta (accession number) luotu Cool URI, joka palauttaa yhdessä nimiavaruuden määrittelyn kanssa tietojärjestelmän REST-rajapinnan kautta tunnistetun kohteen metatiedot. Linkki ei välttämättä ole täysin pysyvä koska REST-rajapinta tulee aikanaan muuttumaan, mutta jos reactome-tietokantapalvelun osoite muuttuu, muutetaan myös identifiers.org-palvelun tarjoamaa uudelleenohjausta reactomeen.

Taulukko 1.

	Resolvointi	Semantiikka	Hallinnointi	Käyttökohde rajattu	Käyttökonteksti määritelty	Metatietomalli	Globaalisti uniikki
URN:NBN	kyllä	-	hajautettu	ei	ei	ei	kyllä
URN:ISBN	kyllä	kyllä	keskitetty	kyllä	kyllä	kyllä	kyllä
ELI (uri)	ei	kyllä	hajautettu	kyllä	kyllä	ei	kyllä
Handle	kyllä	-	hajautettu	ei	ei	ei	kyllä
Paikkatieto (uri)	osittain	kyllä	keskitetty	kyllä	kyllä	ei (ontologia)	kyllä
Lajitieto (uri)	osittain	kyllä	keskitetty	kyllä	kyllä	ei (ontologia)	kyllä
PURL	kyllä	ei	keskitetty	ei	ei	ei	kyllä
ORCID	kyllä	ei	keskitetty	kyllä	kyllä	kyllä	kyllä
ISNI	kyllä	ei	keskitetty	kyllä	kyllä	kyllä	kyllä
OID	ei	kyllä	keskitetty	ei	ei	ei	kyllä
UUID	ei	ei	-	ei	ei	ei	(kyllä)
DID	ei	kyllä	hajautettu	ei	ei	ei	kyllä

Esimerkkejä pysyvien tunnisteiden eri ominaisuuksista. Tunnistetyyppi tulee valita tarkoituksen ja vaatimusten mukaisesti.

Tilanne Suomessa

Julkisessa hallinnossa käytetään hyvin monia eri tunnistejärjestelmiä. Velvoite tunnuksen käyttöön on usein lakisääteinen. Tunnistejärjestelmän valintaan vaikuttavat paitsi lainsäädäntö, sekä yksilöitävä kohde että tunnuksen antaja. Yleisimpiä yksilöitäviä tietoaaineistoja tai –kohteita ovat asiakirjat, käsiteltävät asiat, julkaisut, organisaatiot, kiinteistöt, toimijat, rakennukset ja paikat.

Lakisääteisiä tunnistejärjestelmiä ovat muun muassa arvonlisäverotunniste, yritys- ja yhteisötunniste, kiinteistötunnus, rakennustunnus, kunnanosan ja korttelien numerot, huoneistotunnus ja määräalatumus. Järjestelmät perustuvat usein kansalliseen tai kansainväliseen standardiin, mutta ne voivat olla myös organisaatiokohtaisia. Varsinkin standarditunnisteilla on yleensä jakelua hoitava kansallinen vastuuorganisaatio. Sen mandaatti voi perustua mm. lainsäädäntöön, ohjeistukseen tai vakiintuneeseen käytäntöön. Jotkin tunnistejärjestelmät (esim. UUID) on suunniteltu niin, ettei mitään keskitettyä ohjausta ja koordinoitua tarvita.

Monia tunnistejärjestelmiä on käytetty Suomessa vuosikymmenien ajan, mutta toiminnallisia tunnisteita vain toistakymmentä vuotta. Toiminnalliset tunnisteet kehitettiin vasta 1990-luvun lopulla, kun Internetin käyttö yleistyi ja kaupallistui. Keskeiset URN-standardit kehitettiin 1995-1997 ja Handle system tuli tuotantoon jo 1994, vaikka sen standardointiyritykset jatkuivatkin 2000-luvulle. Handleen perustuva DOI-järjestelmä lanseerattiin vuonna 2000 ja ARK 2001. URN-tunnisteen syntaksi on päivitetty nykyisen URI-syntaksin tasolle vuonna 2017, muut PID-tunnistejärjestelmät eivät tässä suhteessa ole täysin ajan tasalla. Käytännön tasolla tästä ei ole merkittävää haittaa.

DOI-järjestelmää käytetään laajasti tieteellisten artikkeleiden ja tutkimusaineistojen identifiointissa. Handle merkittävin soveltamisalue ovat julkaisuarkistot. URN-järjestelmää käytetään muun muassa kansalliskirjastojen digitoitujen ja syntysähköisten aineistojen tunnistamiseen, mutta järjestelmä on laajenemassa julkaisuista muihin aineistoihin (merenkulku, autoteollisuus).

Toiminnallisiin tunnisteisiin perustuvia PID-tunnuksia on annettu satoja miljoonia, ja omilla vahvuusalueillaan niillä on vakaa asema myös Suomessa. Mutta vaikka PID-tunnisteita on paljon, täälläkin URL-osoitteita on paljon enemmän, mutta emme tiedä miten monet niistä ovat myös tunnuksia.

Cool URI –pohjaisia tunnistejärjestelmiä ajaa erityisesti W3C. Tim Berners-Leen kirjoitus Cool URIs don't change²² loi Cool URI –toimintamallin perustan. Artikkelin ilmestyessä 1997 URN-tunnisteet ja URL-osoitteet erotettiin Internet-standardeissa vielä toisistaan, mutta vuonna 2005 ilmestynyt Internet-standardi RFC 3986 eli URI Syntax²³ korvaa ne Uniform Resource Identifier -käsitteellä.

RFC 3986 ei rajaa mitenkään sitä, millaiset URI:t ovat tunnisteita. Kaikki käy, ja niinpä esimerkiksi henkilön sähköpostiosoitetta voisi käyttää hänen tunnuksenaan, mikä ei ole suositeltava ratkaisu. W3C on siksi laatinut ohjeistusta siitä, millaisia Cool URI –tunnisteiden tulisi teknisesti olla²⁴. Näitä linjauksia

²² Tim Berners-Lee, Cool URIs don't change <https://www.w3.org/Provider/Style/URI>

²³ RFC 3986, <https://www.ietf.org/rfc/rfc3986.txt>

²⁴ Esim. <https://joinup.ec.europa.eu/document/10-rules-persistent-uris>.

on sovellettu muun muassa edellä kuvatuissa EU-ohjeissa siitä, mitä ominaisuuksia Cool URI – tunnisteissa ei saisi käyttää.

Vaikka kiellettyjä ominaisuuksia ei olisi käytetty, URI-tunnisteesta voi olla vaikea nähdä, onko sen edes tarkoitus olla ”cool”. Sitäkin voi olla vaikea selvittää, kuka tunnistejärjestelmää hallinnoi eli huolehtii siitä, että Cool URI -linkki ei hajoa. Käyttäjän pitäisi vain tietää, mitkä URI:t ovat ”cool”. Suomessa luotettavia ovat esimerkiksi lakien ja paikkatiedon URI:t. Paikkatiedon osalta tunnistearkkitehtuuri määrätään INSPIRE-direktiivissä, FINLEX soveltaa omaa tunnisteratkaisua. Laeille on tulossa EU-tason URI-pohjainen ELI-tunniste. Se on Cool URI –pohjainen ja sen rakenne poikkeaa FINLEXin nykyisestä ratkaisusta. Tätä raporttia kirjoitettaessa ei ole tietoa mitä FINLEXin nykyiselle tunnisteelle tulee tapahtumaan.

Tätä selvitystä varten toteutettiin kysely, jolla haluttiin toisaalta kartoittaa nykytilaa tunnisteiden käytössä ja toisaalta selvittää sitä, millaisia yhteisiä määrittelyitä ja tukipalveluita julkishallinnon toimijat arvioivat tarvitsevänsä. Kysely toteutettiin marraskuussa 2017 webropol-kyselynä, joka lähetettiin valtionhallintoon, suurimpiin kuntiin sekä yliopistoille. Vastauksia saatiin 96 kappaletta, joista 17 kunnista ja 79 valtiolta ja yliopistoista. Vastaajat olivat useimmiten tietohallinnon, asiakirjahallinnon tai kirjaamon henkilökuntaa. Vastauksia sallittiin tarkoituksella useampia samalta organisaatiolta – kunhan vastaava osasto, yksikkö tai muu taho oli selkeästi määritetty – ja tällaisia vastauksia myös saatiin.

Vastaukset jakautuivat vastaajatahoittain seuraavasti:

Valtionhallinnosta vastauksia oli yhteensä 64

Kunnilta saatiin 17 vastausta

Yliopistoilta saatiin 15 vastausta

Osastoittain jaoteltuna vastauksia saatiin

Asiakirjahallinnosta 29

Tietohallinnosta 27

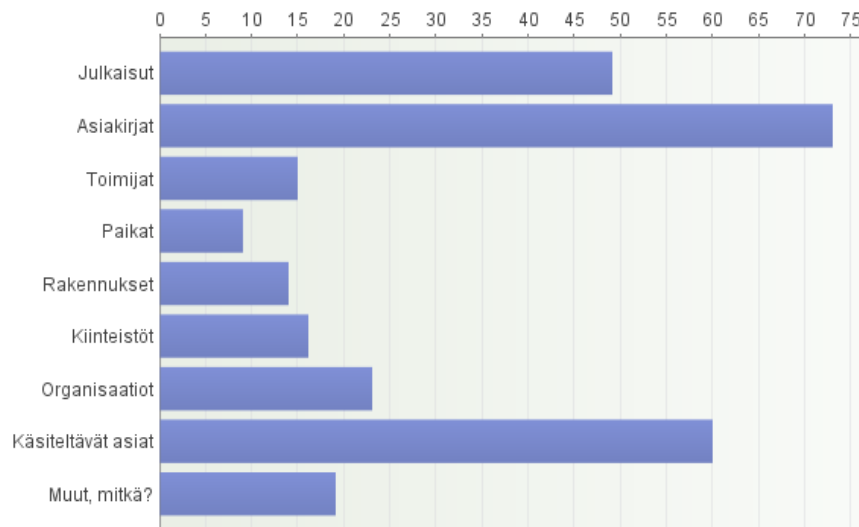
Kirjastoista/kirjastotoimesta 15

Näiden lisäksi vastauksia oli merkitty kirjaamolle 9 kpl sekä valintana ”Muu” 16 kpl. Näistä merkittävä osa pudonnee yllä lueteltuihin kategorioihin. Kirjaamo-vastausten suhteen tätä ei oltu tarkennettu, Muu-vastauksissa oli usein kyse siitä, että vastaajatahoja oli useampia, esimerkiksi ”hallinto”, ”kaikki vaihtoehdot” tai ”koko virasto”.

Tunnisteita annettiin useimmiten asiakirjoille ja käsiteltäville aisoille. Yleisessä käytössä olivat mm. Y-tunnus ja arvonlisäverotunniste. Esimerkiksi hovi oikeuksien, tilastokeskuksen ja Eviran rekistereissä käytettiin useita tunnisteita. Periaatteessa kaikkien perusrekistereissä käytettyjen tunnistejärjestelmien tulee olla muidenkin organisaatioiden hyödynnettävissä.

Kysymyksessä ”Millaisille tietoaaineistoille tai -kohteille annatte pysyviä ja ainutlaatuisia tunnisteita?” yleisin vastaus oli asiakirjat, sen jälkeen käsiteltävät asiat ja julkaisut. Alla olevassa kuvassa vastaukset on jaoteltu tarkemmin. Kategoria Muu jakautui henkilöihin sekä tietokohteisiin liittyviin tunnisteisiin.

Huomioitavaa on, että henkilöiden suhteen varsinaiset henkilötunnukset oli rajattu kokonaan pois kyselystä. Myös erilaiset tekniset tunnisteet rajattiin kyselyn ulkopuolelle.



Kuva 3. Mitä tunnistimia teillä on käytössänne? (x2 kysymystä)

Vastaajat katsoivat lähes kaikkien käytössään olevien tunnistimien olevan pysyviä ja ainutkertaisia (kyllä-ei: 80-4). Kommenttien perusteella esimerkiksi diaarinumerot katsottiin täyttävän kriteerit, riippumatta numeron rakenteesta. Vaikka päällekkäisten diaarinumeroiden antamisen todennäköisyys on pieni, työryhmä haluaa tässä kiinnittää huomiota siihen, että *tunnusten ainutkertaisuutta ei voi arvioida vain organisaation sisäisestä näkökulmasta*. Elektronisia aineistoja tullaan pitkäaikaissäilyttämään keskitetyissä järjestelmissä, eivätkä eri organisaatioiden antamat sisäiset tunnukset saa mennä PAS-järjestelmässä päällekkäin.

Lähes samoin lukemin (82-5) vastattiin, että organisaation pysyviä tunnistimia voi käyttää myös tiedon hakemiseen. *Koska tiedonhakujärjestelmät ovat kansallisia (Finna) tai kansainvälisiä (esim. Google), tunnistimien on oltava ainutkertaisia ainakin kansallisella tasolla mutta mieluummin globaalisti.*

Useiden kommenttien perusteella virasto-/organisaatiokohtaista diaarinumerointia pidettiin riittävän hyvänä ja yksilöivänä ja siitä pois siirtymistä riskinä. Tämä näkyy seuraavan kysymyksen vastauksissa, jossa lukemin 44-31 oltiin sitä mieltä, että asianhallinnassa ei ole tarvetta kansallisesti määritellylle standarditunnistimelle. Jos yhteisesti määriteltyn diaaritunnukseen kuitenkin päätettäisiin siirtyä, sen käyttöönotto tulisi tehdä hallitusti niin, ettei se aiheuttaisi ylimääräistä vaivaa organisaatioissa.

Kun pysyvät tunnistimet liitettiin tarjottaviin tietotuotteisiin, kyllä-ei -vastausten suhde jonkun verran supistui lukemiin 61-24. Kommenttien perusteella käsite tietotuote ei ollut selvä kaikille vastaajille.

Kun kysyttiin mitä nämä yllä mainitut tunnisteet ovat, niin lukumääräisesti eniten niistä liittyi asianhallintaan: diaarinumerot, asiattunnisteet/-numerot sekä asiakirjojen tunnisteet. Kirjastotoimessa ja yliopistoissa käytettiin yleisimmin ISBN, ISSN, DOI, URL, URN, Handle sekä ISIL-tunnisteita. Segmenttikohtaisesti saatiin organisaatioiden substanssiin, esimerkiksi ajoneuvoihin, liittyviä tunnisteita.

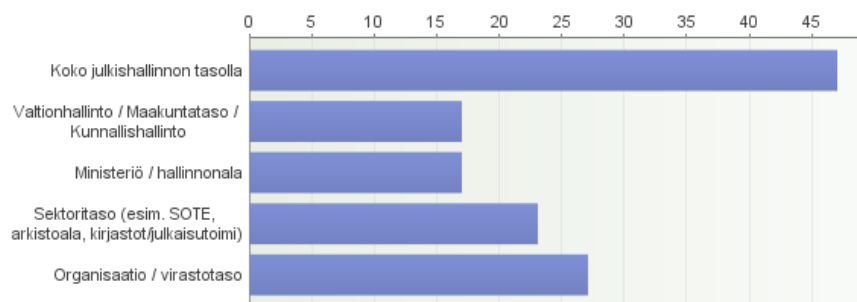
Edelleen kysyttiin, sovelletaanko tunnisteita myös muissa tietojärjestelmissä organisaation sisällä ja/tai sen ulkopuolella. Molemmissa tapauksissa kyllä voitti ein, sisäisessä käytössä lukemin 57-25 ja ulkopuolisessa käytössä lukemin 46-34. Vastausten eroa voi pitää aika pienenä suhteutettuna erityisesti diaariin liittyviin vastauksiin ylempänä. Jonkun verran on vastausten perusteella käytössä rajatussa yhteiskäytössä olevia diaarinumeroiteja.

Suurin osa vastaajista ilmoitti, että jokin sisäinen tai ulkopuolinen taho vastaa tunnusten hallinnoinnista. Kirjastoalan tunnistejärjestelmien hallinnointivastuu oli pääasiassa Kansalliskirjastolla. Organisaation sisällä hallinnoitiin yleisimmin (organisaatiokohtaisia) asian- ja asiakirjahallinnan tunnisteita. Kansalliskirjaston ohella laajasti käytettyjä tunnistejärjestelmiä hallinnoivat esim. PRH, THL ja VRK, minkä lisäksi on toimialakohtaisesti muita kansallisesti merkittäviä tahoja (esim. kiinteistötunnuksia hallinnoiva MML).

Tunnistejärjestelmien soveltamisohjeita oli käytettävissä 74 vastaajalla. Toisin sanoen 22 organisaatiolla ei ole kirjattua tunnistejärjestelmän soveltamisohjetta, mitä voi pitää verraten korkeana lukuna. Osa tunnistejärjestelmistä on niin yksinkertaisia, ettei niiden soveltamiseen tarvita ohjeita, mutta on myös mahdollista että ohjeistusta ei ole hankittu tai sitä ei ole olemassakaan.

Hieman yli puolet vastaajista, 56, piti toiminnallisten tunnisteiden käytön lisäämistä tarpeellisena, jotta linkitys verkossa oleviin kohteisiin olisi luotettavampaa. Ei-vastausten kommentteissa oli hyvin vähän selkeitä perusteluita, ehkä tarvetta ei osattu nähdä.

Seuraavaksi kysyttiin millä tasolla tunnisteita pitäisi julkishallinnossa hallinnoida ja miten. Vaihtoehtoista sai valita kuinka monta halusi. 11 vastaajaa ei valinnut mitään tiettyä tasoa, mutta näidenkin perusteluissa nähtiin tarve hallinnalle jollain tasolla riippuen tunnuksesta ja käyttötarkoituksesta. Alla olevan kuvan palkkidiagrammi sisältää kaikki valinnat vastausvaihtoehdoittain.



Kuva 4 Millä tasolla tunnisteita pitäisi hallinnoida?

Vastaajista 60 valitsi vain yhden vaihtoehdon, ja näistä tasan puolet eli 30 kallistui julkishallinnon tason kannalle. Valtionhallinnon, maakunnan tai kunnallishallinnon taso sai 9 ääntä, pelkän ministeriötason valitsi 4, sektoritason 8 ja organisaatiotason 9 vastaajaa. Loput 25 valitsivat 2-5 vaihtoehtoa. Vastausten perusteluissa toistuu tarkoituksenmukaisuuden tarve.

Keskeisiksi hallinnoitaviksi asioiksi nousevat melko tasaväkisesti käytettävät tunnistejärjestelmät (66) sekä tunnistettavat tietoaaineistot ja kohteet (62). Yksittäisten tunnistejärjestelmien soveltamisohjeet sai 42 ääntä.

Kysymykseen ”minkälaista tukea tai palveluita toivoisitte tunnistejärjestelmien valintaan tai tunnisteiden hallinnointiin” tuli 42 vastausta. Koulutuksen ja ohjeistuksen tarve tulee esiin useimmissa vastauksissa. Toinen toistuva teema on toiminnan ja hallinnoinnin selkeys ja luotettavuus. Tunnisteiden käyttämisestä ei selvästikään haluta lisätyötä tai vaivaa.

Viimeisenä kohtana pyydettiin vapaita kommentteja. Näitä tuli 20 vastaajalta. Varsin monet vastaajat pitivät kysymyksiä vaikeina tai tulkinnanvaraisina. Tämä voi johtua siitä, että organisaation sisälläkin eri yksiköt käyttävät eri tunnistejärjestelmiä erilaisiin tarpeisiin ja erilaisin taustatiedoin. Esimerkiksi Opetushallitus antaa tunnuksia opiskelijoille, hakemuksille, hauille, hakukohteille ja koulutusohjelmille. On vaikeaa laatia kyselyä, jonka tulkinta olisi helppoa kaikille toimijoille ja joka olisi samalla myös riittävän kattava.

Asiakirjahallinnossa käytetään laajalti diaarinumeroita, jotka ovat vastaajien mukaan yleensä pysyviä ja ainutkertaisia. Numeron rakenne vaihtelee, se voi olla organisaatiolyhenne/tehtäväryhmä/kirjaamisvuosi eli esimerkiksi PRH/28/2017, tai juokseva numero/tehtäväryhmän tunnus/diari vuosi eli 121/31/2016. Saatujen tietojen perusteella on vaikeaa arvioida luotettavasti diaarinumeroiden ainutkertaisuutta, mutta esimerkiksi jälkimmäisen mallin mukaiset numerot eivät välttämättä ole ainutkertaisia koska tehtäväryhmiä koordinoidaan vain organisaatiokohtaisesti, ei kansallisella tasolla.

Kyselyyn vastanneista organisaatioista 31 katsoi, että asianhallintaan olisi hyvä saada kansallinen standarditunnus; 44 vastaajaa oli sitä mieltä, ettei sitä tarvita. Yhdessä kommentissa todettiin, että eri virastojen käyttämät diaaritunnukset ovat olleet tunnistettavia ja toimivia. *Diaarinumeroiden riittävää ainutkertaisuuden astetta tulisi kuitenkin arvioida tulevaisuuden tarpeiden kuten keskitetyn pitkäaikaissäilytyksen, eikä nykytilanteen pohjalta.*

Merkittävä osa julkishallinnon käyttämisestä tunnistejärjestelmistä perustuu OID-järjestelmään. Sen käyttö puolestaan nojautuu JHS-suositukseen JHS 159 OID-yksilöintitunnuksen soveltaminen julkishallinnossa²⁵. OID:n etuna on tunnuksiin sisäänrakennettu globaali ainutkertaisuus, ja siksi sitä on helppo käyttää esimerkiksi digitaalisessa muodossa olevan aineiston välittämisessä Kansallisarkistoon.

JHS 159:n mukaan Suomen OID-juureen (1.2.246) voidaan perustaa esimerkiksi henkilötunniste (1.2.246.21), huoneiston tunniste, kaava, tonttijako-, rajoitusalue- ja rakennuskieltotunnukset, rakennustunnukset, sähköinen asiointitunnus (SATU), kansainväliset ja kansalliset koodistot sekä yritys- ja yhteisötunnus (1.2.246.10). Lisäksi voitaisiin perustaa OID-pohjaiset asiakas-, asiakirja-,

²⁵ JHS 159 Use of the ISO OID Object Identifier in public administration (2006) <http://www.jhs-suositukset.fi/suomi/jhs159>

henkilönumero-, huone-, laite-, koodisto-, laskutus-, logistiikka-, näyte-, ohjelmistotunnus jne. Käytännössä OID-tunnuksen soveltaminen on edennyt nopeimmin sosiaali- ja terveysalalla.

Tunnisteita tulisi 18 vastaajan mielestä hallinnoida koko julkishallinnon tasolla. Huomattavasti pienempi vastaajamäärä tuki erilaisia hajautettuja ratkaisuja; esimerkiksi ministeriötason hallinnointia kannatti vain kaksi vastaajaa. Vaikka käytettyjen tunnusjärjestelmien moninaisuus tekee keskitetyn ohjauksen haasteelliseksi, se on kuitenkin paras vaihtoehto ainakin toiminnallisia tunnisteita koskevaan ohjaukseen, joka on jossakin määrin riippumatonta käytetyistä tunnusjärjestelmistä.

Toiminnallisten tunnisteiden käyttö Suomessa

URN

Kansalliskirjasto koordinoi URN-tunnisteen käyttöä Suomessa ja on osallistunut standardin kehittämiseen IETF:ssä. URN-syntaksi määritellään vuonna 2017 julkaistussa Internet-standardissa RFC 8141²⁶.

URN-tunnisteita on jaettu Suomessa julkaisuille ja muille aineistoille jo yli vuosikymmenen ajan. Käytetyt nimialueet ovat URN:NBN (kansallisbibliografian tunnus) sekä URN:ISBN (ISBN-tunnus). Tunnisteita sovelletaan esimerkiksi kirjastojärjestelmissä, julkaisuarkistoissa (virallisena tunnisteena, silloin kun Handle on sisäinen tunniste) sekä PAS-järjestelmässä.

Kyselyyn vastanneista organisaatioista URN-tunnisteita käytti 25. URN:NBN-tunnisteita käyttävät myös useat muut suomalaiset organisaatiot kuten CSC ja Yhteiskuntatieteellinen tietoaarkisto aineistojensa identifiointiin. Niille on jaettu URN:NBN:fi-nimiavaruudesta omia aliavaruuksia, joita ne hallinnoivat.

DOI

DOI-tunnistetta käytetään muun muassa Tieteellisten seurain valtuuskunnan journal.fi-palvelussa tieteellisten artikkeleiden tunnisteena sekä esimerkiksi data-aineistojen tunnisteena silloin, kun aineisto tallennetaan DOI-tunnusta soveltaviin kv. palveluihin kuten DataCiteen. Kyselyyn vastanneissa organisaatioissa DOI-käyttäjää oli noin 10.

Kansallista DOI-koordinaattoria ei toistaiseksi ole nimetty, mutta CSC on kiinnostunut ottamaan tämän vastuun. CSC on perustamassa kansallista DataCite-konsortiota, jossa DOI-tunnisteita jaetaan etupäässä tutkimusaineistoille. DOI perustuu handle-palveluun, mutta sillä on tarkoin määritelty tietomalli ja selkeä arkistointi, kate ja resolvoiva vaatimus.

Handle

Handle-tunnistetta käytti vastanneista organisaatioista noin 10. Sitä sovelletaan kaikissa DSpace-ohjelmistoa soveltavissa julkaisuarkistoissa järjestelmän sisäisenä tunnisteena. Monissa julkaisuarkistoissa, kuten Helsingin yliopiston Heldassa, se on myös varsinainen tunniste.

Handle systemillä ei ole kansallista koordinaattoria, vaan jokainen organisaatio vastaa itse omasta Handle-tunnisteen käytöstään.

²⁶ RFC 8141 <https://tools.ietf.org/html/rfc8141>

Kyselyvastausten perusteella muiden kuin edellä mainittujen toiminnallisten tunnisteiden käyttö Suomessa on vähäistä. Niiden hyödyntäjiksi ilmoittautui vain kaksi organisaatiota.

Cool URI –osoitteiden käyttö tunnisteina

Kyselyn perusteella Cool URI –pohjaisia tunnistejärjestelmiä käyttää noin 35 organisaatiota. Jaettujen tunnusten määrästä ja tunnistetuista kohteista ei saatu kyselyssä tietoa. Tunnistejärjestelmät voivat olla joko paikallisia tai keskitetysti koordinoituja ja hallinnoituja kuten FINLEXin tunnisteet, ELI-tunniste ja paikkatietojen tunniste.

Cool URI –tunnisteita jakavista organisaatioista ei ole olemassa keskitettyä rekisteriä. Emme tiedä, mille aineistoille Cool URI –tunnisteita annetaan. Toistaiseksi ei ole myöskään organisaatiota, joka tarjoaisi keskitetysti Cool URI –pohjaisia tunnistepalveluita ja niihin liittyvää neuvontaa.

VRK aikoo tarjota Cool URI –palveluita (työnimi URI uudelleenohjauspalvelu) domainissa <http://uri.suomi.fi>.²⁷ Palvelu on tarkoitus ottaa käyttöön yhteentoimivuusalustan työkaluissa (sanastot, tietomallit, koodistot). Teknisesti kyse on HTTP-protokollaan perustuvasta edelleenohjauksesta eikä varsinaisesta resoluutiopalvelusta. Suomi.fi on domainina suojattu, mutta mikäli sen varaan halutaan rakentaa pysyviä tunnisteita, on syytä varmistua huolellisesti tunnisteiden hallinnasta ja laadusta, jotta luotettavuus (ja brändi) eivät kärsi (puhumattakaan hallinnon ja kansalaisten oikeusturvan tärkeydestä tässä yhteydessä). Palvelulupaus, vastuut, kate ja käyttötarkoitus sekä pitkäaikainen suunnittelu ja avoin dokumentaatio ovat keskeisiä.

Koodistot ja sanastot

Koodistojen ja käsitteiden luotettava hallinnointi ja luotettavuus vaativat pitkän aikavälin versionhallintaa. Hallinnointi on syytä toteuttaa pysyvien tunnisteiden avulla. THL ja VRK käyttävät koodistopalvelua koodistojensa säilyttämiseen ja julkaisemiseen. Koodisto ei siis yksinään tarjoa yksilöiviä tunnisteita, eli jos koodistoa käytetään, niin yksilöivää tunnistetta ei välttämättä käytetä. Koodistolla itsellään on sen sijaan oltava yksilöivä tunniste, jos koodiston sisältö ei ole pysyvä. Esimerkiksi THL:llä on tautiluokitus, joka muuttuu ajassa, ja sen jokaisella versiolla pitää olla oma pysyvä tunnus. Koodistot on yksilöity (OID), koodiston sisällä on koodiarvo (mm. kuntakoodi). Organisaatioilla on omat pysyvät tunnukset eikä samaa tunnusta ei anneta toiselle, vaikka organisaatio loppuisi. Koodisto vaatii siis myös nimiavaruuden hallintaa.

Yhteenveto

Selvityksen mukaan organisaatioissa on käytössä huomattava määrä erilaisia tunnistejärjestelmiä ja ne jakavat pysyviä tunnuksia, mutta ne eivät läheskään aina ole koneymärrettäviä, toiminnallisia tai globaalisti ainutkertaisia (ts. ne eivät varsinaisesti täytä kaikkia pysyvän tunnisteiden kriteereitä). Tunnistejärjestelmien tekniset puutteet voivat aiheuttaa ongelmaa, kun tietoa jaetaan tai siirretään organisaatioiden välillä ja kun palveluita halutaan tuottaa sähköisesti ja kansalaislähtöisesti.

Organisaatioiden tyytyväisyys nykyisiin organisaatiokohtaisiin diaarikäytäntöihinsä on oire tiedon siiloutumisesta. Kun tietoa jaetaan ja käytetään tehokkaasti organisaatiorajojen yli ja tarjotaan myös kansalaisille avoimesti (mm. avoimena datana) hallinnoidut, globaalisti toiminnalliset ja yhteiskäyttöiset

²⁷ katso <https://confluence.csc.fi/display/VRKYTI/URI+uudelleenohjauspalvelu>

tunnisteet ovat elintärkeitä. Julkisen hallinnon organisaatiot voivat käyttää erilaisia tai rakenteeltaan poikkeavia tunnisteita samanlaisille / samansisältöisille asioille. Näiden tunnisteiden yhtenäistäminen tai tunnisteiden avulla tapahtuva linkitetyn avoimen datan saumaton yhteismitallinen tuottaminen ei käytännössä ole mahdollista, joten tunnusten linkittäminen ja näiden toisiinsa liittyvien tunnisteiden hallinta on ensiarvoisen tärkeää.

Tunnisteiden linkittäminen ja hallinta edellyttävät siihen tarkoitettuja palveluita. Julkisesti dokumentoidut ja yhteisesti sovitut tavat nimetä asioita ja viitata niihin ovat keskeinen osa digitalisoitumisen toimeenpanoa. Kun tietoa tarjotaan verkossa avoimesti tai sitä jaetaan viranomaisten kesken, selkeät palvelulupaukset ja käytännöt semantiikan ja laadun osalta ovat tärkeitä sekä yhteentoimivuuden että palvelukehityksen näkökulmasta. Myös kansallisesti keskitetty pitkäaikaissäilytys vaatii säilytettävän aineiston globaalia yksilöintiä. Yksilöintiä ei tarvita vain asiakirjoille ja reaali maailman asioille vaan myös pelkästään digitaalisille objekteille (ns. born-digital aineisto) sekä erilaisille abstrakteille käsitteille.

Mikäli tämän hetkiset suunnitelmat toteutuvat nykymuodossaan, Suomessa on jatkossa ainakin kolme kansallista pysyviä tunnuksia tarjoavaa palvelua (Kansalliskirjasto ja URN-tunnukset, CSC ja DOI-tunnukset sekä VRK ja Cool URI:t). Tämän lisäksi useat tahot käyttävät Handlea tai Cool URI –osoitteita ilman keskitettyä tukea tai hallintaa.

Pysyvien tunnusten pitää säilyä toiminnallisina periaatteessa ikuisesti, mikä ei tule olemaan ilmaista. Kolmen rinnakkaisen kansallisen järjestelmän malli tarkoittaisi suurella todennäköisyydellä päällekkäisiä teknisiä ratkaisuja ja tuki- ja hallinnointiorganisaatioita mikäli suunnittelusta ja ohjauksesta ei huolehdita kansallisesti. Tilanne ei ole tarkoituksenmukainen eikä kustannustehokas mikäli kansallisena tavoitteena on pyrkiä sekä keskittämään palveluita että toteuttamaan niitä yhdenmukaisesti. Hallinnointi edellyttäisi vähintäänkin selkeää vastuunjakoja ja tunnisteiden muodostamiskriteereitä palveluntarjoajien kesken.

Tarvitaan yleistä keskustelua millainen painoarvo annetaan toisaalta pysyville tunnisteille, toisaalta Cool URI –pohjaisille ratkaisuille. Cool URI –mallissa on tiettyjä etuja (ei tarvita erillistä resoluutiopalvelua) mutta myös merkittäviä rajoituksia, jotka on otettava huomioon jatkosuunnittelussa ja päätöksenteossa.

Pysyvän tunnuksen valinnan tulisi perustua kattavaan tilanteen ja tarpeiden arviointiin. Esimerkiksi Hollannissa käytettävä PID-valitsin²⁸ voitaisiin sopeuttaa Suomeen, jotta organisaatiot voivat tehdä tunnistejärjestelmävalintansa asiapohjalta. Aiemmin yllä on esitetty useita erilaisia tarkistuslistatyyppejä esimerkkejä, joita voisi hyödyntää. Myös Ruotsissa toteutettu Checklista: Beständiga identifierare²⁹ voisi toimia apuna.

Tunnisteet tietohallintolain mukaisia yhteentoimivuuden kuvauksia ja määrittäjiä.

²⁸ PID-guide <http://www.ncdd.nl/en/pid-wijzer/>

²⁹ Checklista: Beständiga identifierare. Digisam.

http://www.digisam.se/images/checklistor/Checklista_bestandiga_identifierare.pdf

Työryhmän suosituksia

Tunnisteiden hallinnointi

Viranomaisten tulee soveltaa ainutkertaisia ja pysyviä tunnistejärjestelmiä, erityisesti silloin kun tietoa jaetaan verkossa tai säilytetään pitkäaikaisesti, mutta myös tiedon hallinnoinnissa yleensä.

Tunnistejärjestelmien hallinnointia on kehitettävä siten, että tietovarantojen ylläpitäjät suhteuttavat käyttämiensä tunnistejärjestelmien omaisuudet identifioitujen kohteiden elinkaareen ja jakelutarpeisiin. Julkishallinnossa tulisi soveltaa yhdenmukaisia tunnistejärjestelmien soveltamiskäytänteitä esimerkiksi pitkäaikaissäilytettävälle aineistoille, koska niiden osalta tunnuksen elinkaari voi olla jopa pidempi kuin identifioitujen aineiston.

Työryhmän suosittelee, että:

- 1) Jos aineistolle on jo annettu ainutkertainen ja pysyvä tunnus, käytä sitä
- 2) Jos jollekin aineistotyyppille on olemassa standardoitu tunnistejärjestelmä, käytä sitä
- 3) Hyödynnä olemassa olevia, avoimia ja laajasti käytettyjä toimittajariippumattomia teknologioita
- 4) Jos valmista tunnistejärjestelmää ei ole ja luot organisaatiokohtaisen ratkaisun, suunnittele se huolellisesti, vältä turhaa semantiikkaa ja suunnittele tunniste niin, että sitä voidaan soveltaa myös pysyvänä toiminnallisena tunnisteena
- 5) Huolehdi antamiesi tunnusten ainutkertaisuudesta ja pysyvyydestä
- 6) Noudata soveltamasi tunnistejärjestelmän ohjeita ja käytäntöjä. Jos järjestelmä on oma, luo sille selkeät käyttöohjeet
- 7) Kun identifioitua kohteesta tulee uusi versio, anna sille uusi tunnus ja huolehdi linkityksestä edellisiin versioihin tai – jos ne on poistettu – niiden metatietoihin
- 8) Älä käytä tunnuksia uudelleen tai tuhoa niitä
- 9) Jos käytät Cool URI –tunnuksia, tee niistä selkeitä ja hallinnoi niitä huolellisesti
- 10) Dokumentoi ja julkaise tunnusten jakelun käytännöt ja periaatteet
- 11) Toimi vastuullisesti³⁰

Lähtökohtaisesti tietoa arkistoidaan, jotta se olisi jatkossa paitsi oman organisaation, myös muidenkin hyödynnettävissä. Käytettävyyden parantamiseksi arkistoitujen aineistojen tunnisteet ja muu metatieto tulee myös kyetä tarjoamaan muiden käytettäväksi. Näin tieto linkittyy, sen ylläpito helpottuu ja turhaa redundanssia voidaan välttää. Jos on olemassa riski, että paikallisen tunnusjärjestelmän tunnukset eivät ole ainutkertaisia, tunnuksen rakennetta tulisi muuttaa esimerkiksi lisäämällä siihen ainutkertaisuuden takaava etuliite, minkä jälkeen paikallinen tunniste on helposti muutettavissa pysyväksi toiminnalliseksi tunnisteeksi. Jos aineistolla on jo ainutkertainen tunnus, *sitä tulee käyttää, ja tästä syystä paikallisten tunnisteiden kehittämisessä on otettava muiden organisaatioiden mahdollinen käyttö huomioon.*

Tunnisteiden tulisi lisäksi olla toiminnallisia (resolvoitavia) siten, että ne aina viittaavat johonkin ja niiden edustamasta asiasta aina saa tietoa. Toiminnallinen tunnus mahdollistaa myös erilaisten palvelujen

³⁰ Osa suosituksista perustuu julkaisuun Julie A. McMurphy et al., Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. PLoS Biology. June 29, 2017 <https://doi.org/10.1371/journal.pbio.2001414>

tarjoamisen ja parantaa siten mm. linkitetyn datan laatua. Tunnistejärjestelmänä tulisi käyttää Suomessa yleisesti käytettyjä järjestelmiä kuten DOI:ta, Handlea tai URN-tunnistetta. Niiden etuna on valmis resoluutiopalvelusovellus (Handle server, URN-tunnisteen osalta esimerkiksi Kansalliskirjaston URN-resolveri ja siihen perustuva palvelu julkishallinnolle). Organisaatiokohtaisia tunnisteratkaisuja tai sovelluskehitystä tulisi välttää.

Tunnisteen soveltamisen periaatteiden (esim. tunnisteiden jakelu- ja mahdollinen korjausprosessi, tunnistettavan aineiston valinnan periaatteet jne.) tulee olla yhdenmukaisia ja avoimesti saatavilla. Lisäksi tulee huomioida, että tunnisteen elinkaari voi olla pidempi kuin sen identifioiman resurssin.

Kansallisen tason hallinnointi

Pysyvien tunnisteiden kansallinen käyttö edellyttää yhdenmukaista hallinnointia, joten myös tunnistejärjestelmien luotettavuudesta ja selkeydestä on huolehdittava yhdenmukaisesti. Niillä tulisi olla käyttöohje sekä selkeä käyttötarkoitus ja soveltamisala. *Kaikille yleisesti käytetyille toiminnallisille tunnisteille tulisi nimetä kansalliset vastuutahot*, ja näiden toimijoiden tarjoamat palvelut tulisi määritellä esim. sopimuksin. Näitä palveluita voivat olla esimerkiksi osallistuminen tunnistejärjestelmän (standardin) kansalliseen ja kansainväliseen kehittämiseen ja hallinnointiin, keskitetyn resoluutiopalvelun tarjoaminen sekä koulutus ja viestintä.

Kun organisaatio ottaa käyttöön toiminnallisen tunnistejärjestelmän asia tulisi dokumentoida kyseisen järjestelmän kansallisen vastuutahon ylläpitämään käyttäjärekisteriin tai mahdolliseen myöhemmin perustettavaan kansalliseen pysyvien tunnisteiden käyttäjärekisteriin. Tällöin pysyvien tunnusten hallinta ja organisaatioiden toiminnan seuranta ja tuki helpottuisivat. Lisäksi menettely tukisi kansallista yhteentoimivuuden ohjausta mm. sen vuoksi, että julkisessa hallinnossa tulee varautua pysyvien tunnusten pitämiseen toiminnallisina sen jälkeenkin, kun aineistosta vastanneen organisaation toiminta päättyy. Pysyviä tunnisteita käyttävien organisaatioiden rekisteröinti ja tarkoituksenmukaisesti resursoituiden tunnistejärjestelmäkohtaiset kansalliset vastuorganisaatiot helpottaisivat kuvattujen ongelmatilanteiden ratkaisemista.

Koska tunnisteet on katsottava yhteentoimivuuden kuvauksiksi ja määrittäviksi tunnisteiden kansallinen hallinnointi kuuluu nykyisen lainsäädännön mukaisesti valtiovainministeriön vastuulle (tietohallintolaki). Tunnistejärjestelmien hallinnoinnin jatkuvuuden näkökulmasta on tarkoituksenmukaista, että nykyiset vastuorganisaatiot jatkavat toistaiseksi tehtävissään mutta laajentuneiden vastuualueiden (koko julkinen hallinto, ei vain oma sektori) vaikutukset niiden toiminnalle tulee selvittää.

Selvitys ei ota yleisesti kantaa siihen, mille aineistoille pitäisi antaa pysyvät toiminnalliset tunnisteet. Näiden (elektronisten) aineistojen osuus voi olla vain murto-osa organisaation kaikista aineistoista. Jos organisaatiolla on vastuu joidenkin aineistojen pitkäaikaisesta säilyttämisestä, niille tulisi antaa pysyvät tunnisteet dokumentoidun suunnitelman mukaisesti.

Kansallisten koodistopalveluiden käsitteille ja termeille on kuitenkin syytä luoda toiminnalliset tunnisteet, jotta linkitetyn datan laatu ja datan hyödyntäminen voidaan taata silloinkin, jos näitä järjestelmiä ylläpitävät organisaatiot vaihtuvat tai niiden nimet muuttuvat. Kun linkit palveluihin perustuvat pysyviin tunnisteisiin resoluutiopalvelun tarjoama linkitys toimii myös ennakoimattomissa muutostilanteissa ja silloinkin, kun nimipalveluun perustuva uudelleenohjaus ei toimi tai sitä ei voi käyttää.

Kun koodistopalveluiden sisältämiin käsitteisiin voidaan viitata pysyillä tunnisteilla julkishallinnon palveluissa käytetyt käsitteet säilyvät yhdenmukaisina myös silloin, jos käsite päivitetään.

Työryhmän näkemyksen mukaan keskitettyjen koodistopalveluiden kuten Kansalliskirjaston ylläpitämän Finton käyttö yhtenäistäisi julkisen hallinnon eri palveluissa käytettävää käsitteistöä. Käsitteistöjen ylläpidon keskittäminen poistaisi käsitteistöjen ylläpitovastuut yksittäisiltä organisaatioilta ja mahdollistaisi rahoituksen ja hallinnollisen työn keskittämisen. Keskitetty käsitteistöjen hallinnointi ja käyttö pysyviä tunnisteita hyödyntäen olisi kokonaistaloudellisesti edullisempaa, kuin niiden hajautettu ylläpito ja kehittäminen. Lisäksi palvelujen ja palveluissa käytettyjen kuvailutietojen laatu ja luotettavuus paranisi, kun kaikki kansallisesti merkittävät erikoisalojen sanastot saataisiin kansalliseen käyttöön yhteisen palvelun kautta.

Työryhmä suosittaa seuraavaa:

- Elektronisten aineistojen perinteiset tunnukset esitetään myös toiminnallisina pysyvinä tunnuksina linkitetyn datan luotettavuuden parantamiseksi
- Pysyvät tunnisteet katsotaan tietohallintolain mukaisiksi yhteentoimivuuden kuvauksiksi ja määrityksiksi (Tietohallintolaki 634/2011, 4§)
- Määritellään kansallinen pysyvien tunnisteiden ohjauksesta vastaava taho ja kehitetään kansallista tunnisteiden hallinnan koordinoitua myös operatiivisella tasolla päällekkäisyyksien välttämiseksi
 - Tietohallintolain mukaan yleinen ohjaus kuuluu valtiovarainministeriölle
- Keskitetään pysyvien tunnisteiden hallinnointia nykyisiin hallinnointia suorittaviin organisaatioihin ja tiivistetään kansallista yhteistyötä
 - Keskittäminen tulee arvioida tapauskohtaisesti ja sen tulee tukea palvelukehittämistä
- Sovelletaan ensisijaisesti vain kansainvälisiin standardeihin perustuvia tunnistejärjestelmiä
- Käynnistetään kansallinen pysyvien tunnisteiden JHS-suosituksen valmistelu
- Perustetaan pysyvä kansallinen toimielin nykyisen kansallisen tietohallinnon ohjausrakenteen puitteissa, joka vastaa tunnistejärjestelmiin liittyvästä yhteistyöstä, kehittämisestä ja ohjauksesta. Toimielimen tehtäviin voisi kuulua esimerkiksi seuraavia kokonaisuuksia:
 - Kansallinen koordinointi
 - Käytettävät tunnistejärjestelmät
 - Tunnistettavat kohteet
 - Tunnistejärjestelmien soveltaminen
- Selvitetään mahdollisuudet kehittää ja ottaa käyttöön pysyvät tunnisteet viranomaisen asian- ja asiakirjanhallinnassa

Kansallinen asianhallinnan tunnistepalvelu

Asianhallinnassa käytetään organisaatiokohtaisia tunnistejärjestelmiä, joiden pysyvyydestä ja ainutkertaisuudesta ei ole täydellistä varmuutta. Yhteentoimivuuden ja kansalaisten oikeusturvan parantamiseksi sekä viranomaisten välisen tiedonvaihdon ja asiakirja-aineistojen pitkäaikaissäilytyksen helpottamiseksi on tarpeen luoda asianhallinnalle toiminnallinen kansallinen tunnistejärjestelmä sekä sen edellyttämä (kansallinen) resoluutiopalvelu. Se voidaan toteuttaa joko keskitetysti tai yhteen liitettyjen resoluutiopalvelimien verkostona, jolloin resoluutiota koskeva informaatio kuten linkit

pysyvistä tunnisteista URL-osoitteisiin voidaan hajauttaa. Tunnistejärjestelmä voisi soveltuvin osin perustua olemassa oleviin organisaatiokohtaisiin asioiden ja asiakirjojen tunnistejärjestelmiin, koska paikallisesti hallinnoidusta sisäisestä tunnuksesta on melko yksinkertaista tehdä ainutkertainen pysyvä tunnus kuten URN.

Resoluutiopalvelun toteutustavalla ei ole vaikutusta tunnusten jakeluun ja hallinnointiin; se voisi edelleen olla hajautettua vaikka resoluutiopalvelun ylläpito olisi keskitetty.

Käytettävistä tunnistejärjestelmistä tulee sopia erikseen. Valittavasta järjestelmästä riippumatta tunnisteiden on oltava tarkoituksenmukaisesti hallinnoitu, ja sen käyttö tulee ohjeistaa. Järjestelmässä sovellettavien metatietomallien ja teknisten rajapintojen on oltava julkisia ja standardeihin perustuvia.

Kansallinen viranomaisen asianhallinnassa käytettävä organisaatioriippumaton yhteinen pysyvä tunnistejärjestelmä mahdollistaisi asiakirjojen ja muiden resurssien linkittämisen asioihin ja asioiden saumattoman käsittelyn eri viranomaistahoissa. Menettely parantaisi organisaatorajat ylittävien hallinnon prosessien yhteentoimivuutta ja tukisi kansallisen sähköisen asian- ja asiakirjahallinnan hallittua yhtenäistämistä sekä helpottaisi kansalaisen asiointia viranomaisen kanssa. Viranomaisen asianhallinnan pysyvä tunnistetehostaisi hallinnon toimintaa ja olisi konkreettinen toimenpide jolla edistetään julkisen hallinnon digitalisaatiota. Kansallisen asianhallinnan kehittäminen edellyttäne jatkossa myös kansallisen asiakirjatunnisteen kehittämistä ja käyttöönottoa.

Toiminnalliset tunnisteet

Perinteiset tunnisteet tulisi tallentaa kuvailutietoihin myös toiminnallisina tunnuksina. Niiden luominen on yleensä helppoa, esimerkiksi ISBN 978-951-25-2971-1 voidaan esittää URN-tunnuksena <http://urn.fi/URN:ISBN:978-951-25-2971-1>. Pysyvistä tunnuksista saatavan linkin avulla asiakkaat saavat tunnistetun kohteen (tässä tapauksessa e-kirjan) tai siihen liittyvää muuta aineistoa käyttöönsä riippumatta sen verkko-osoitteesta.

Tutkimusdatan ja tieteellisten artikkeleiden toiminnallisena tunnisteena käytetään kansainvälisesti DOI-tunnistetta, ja Suomessakin sen käyttö on aloitettu. Kansalliskirjasto on käyttänyt kokoelmiensa digitoinnissa URN-tunnistetta. Näiden tunnistejärjestelmien käyttöä Suomessa tulisi laajentaa esimerkiksi asiakirjoihin ja muihin aineistoihin, jotka voidaan tulevaisuudessa siirtää keskitetysti ylläpidettävään pitkäaikaissyntetysjärjestelmään.

Tieteellisen ja taiteellisen luovan työn tekijöille on annettava ainutkertaisten ja pysyvien tunnisteiden avulla voidaan helpottaa avoimen linkitetyn datan luontia ja huolehtia siitä, että tieteellinen meriitti ja tekijänoikeuskorvaukset voidaan kohdentaa oikein. Tutkijoille suositellaan ORCID-tunnusta, taiteellisen luovan työn tekijöille sekä tutkimusorganisaatioille ISNI-tunnusta, joka voidaan antaa myös tutkijoille, jotka eivät halua / pysty hakemaan ORCID-tunnusta.

Avoidata.fi:n on syytä ottaa huomioon EU:n ohjeistus tunnisteiden osalta ja huolehtia toiminnallisiin tunnuksiin perustuvan linkityksen mahdollistamisesta.

Tiedon linkittäminen

Laadukkaan ja pitkäaikaisen linkitetyn datan tuottamista pysyvien tunnisteiden avulla on edistettävä mm. koulutuksen ja ohjauksen avulla. Eteneminen on helpointa aloilla, jossa voidaan hyödyntää olemassa olevia kansainvälisiä malleja. Esimerkiksi paikkatietoa voidaan ja tulee julkaista linkitettynä datana EU:n INSPIRE-direktiivin mukaisesti.

Linkitettyä dataa tukevien palveluiden kuten Yhteentoimiva.suomi.fi, Finton ja paikkatieto.fi:n tarjoamien rajapintojen ja pysyvien tunnisteiden avulla poimittavissa olevan tiedon laatu tulee varmistaa.

Tarkoin rajattuja linkitetyn datan kokonaisuuksia (kuten paikkatieto, juridinen informaatio) voidaan rakentaa käyttäen kyseisen yhteisön sääntöjen mukaisia Cool URI –tunnisteita mikäli sääntöjä ja muodostamiskäytäntöjä ohjataan ja hallinnoidaan tarkoituksenmukaisesti. Käyttökohteet ja –kontekstit on määriteltävä hyvin ja palvelulupausten ja toimintaperiaatteiden on oltava dokumentoituja ja julkisia. Käyttövarmuuden takaamiseksi tulisi selvittää mahdollisuus lohkoketjuteknologian käyttöön sekä tulevan DID-standardin³¹ hyödyntämiseen³².

Nykyisiä resoluutiopalveluja tulee kehittää älykkäämmiksi, koska tämä mahdollistaa linkkien toiminnallisuuden parantamisen. Pysyvän tunnisteiden avulla käyttäjä voisi saada paitsi identifioitun aineiston, myös siihen liittyvää metatietoja tai esimerkiksi ilmoituksen siitä, että aineisto on saatavilla myös muissa tiedostoformaateissa

Resoluutiopalvelujen kehittäminen on pitkäjänteistä työtä, koska ohjelmistokehityksen ohella se edellyttää selkeää vastuutahojen ja roolien määrittelyä ja tunnistejärjestelmien standardien kehittämistä sekä kansallisenä että kansainvälisenä yhteistyönä.

Määritelmät

Tunnus

Ainutkertainen merkkijono joka joko yksinään tai yhdessä muiden metatietoelementtien kanssa pysyvästi identifioi resurssin, organisaation, henkilön, paikan tai muun entiteetin. Esimerkiksi kirjan ISBN-tunnus tai OID-pohjainen oppijanumero. Tunnusta voidaan käyttää toiminnallisen tunnisteiden osana. Vrt rekisteritunnus, käyttäjätunnus.

Toiminnallinen tunniste / Persistent Identifier (PID)

Koneellisesti tulkittava ainutkertainen ja pysyvä tunniste, joka on resoluoitavissa verkossa. Tunnisteeseen voi sisältyä koneellisesti käsiteltäviä, resoluution ”älykkyyttä” lisääviä elementtejä. Tällä hetkellä käytettävissä olevat resoluutiopalvelut ovat tunnistejärjestelmäkohtaista.

Yleisesti käytettyjä toiminnallisia tunnisteita ovat ARK (Archival Resource Key), DOI (Digital Object Identifier), Handle, PURL (Persistent URL) ja URN (Uniform Resource Name). Niistä suosituimmat ovat DOI ja Handle, joiden ohjelmistoympäristö on sama (DOI on tietyn tyyppinen Handle) mutta hallinnollisesti palvelut ovat hyvin erilaiset. DOI-tunnusten käyttöä valvotaan paljon Handle-käyttöä tarkemmin.

³¹ Decentralized Identifiers (DIDs) v0.7. Data Model and Syntaxes for Decentralized Identifiers (DIDs). Draft Community Group Report 09 December 2017. <https://w3c-ccg.github.io/did-spec/#dfn-dlt>

³² Lukasz Bolikowski Aleksander Nowiński Wojtek Sylwestrzak, A System for Distributed Minting and Management of Persistent Identifiers. International Journal of Digital Curation, 2015, Vol. 10, <http://dx.doi.org/10.2218/ijdc.v10i1.368>

Erillisen resoluutiopalvelun käyttö ei vapauta aineiston hallinnoijaa ylläpitovastuusta, sillä aineiston omistajan on aina huolehdittava siitä, että toiminnallinen tunnus linkittyy verkossa oikeisiin URL-osoitteisiin.

URN, Uniform resource name

1. Internet Engineering Task Forcen standardoima PID-järjestelmä.
2. Resurssin pysyvä tunnus, jonka tulee olla tekniikkariippumaton ja riippumaton resurssin sijainnista

URL, Uniform resource locator

käytetyn yhteysprotokollan (esim. HTTP) mukainen verkko-osoite. Käytännössä linkki ja resurssin sijainti verkossa (tai verkkoarkistossa).

URL-uudelleenohjaus (URL redirection)

tekniikka jolla sama verkkosivu voidaan asettaa tarjolle useammasta kuin yhdestä URL-osoitteesta. Uudelleenohjaus voi olla joko pysyvä tai väliaikainen.

URI Uniform resource identifier

Joko resurssin pysyvä tunnus tai verkko-osoite.

Cool URI

URI (URL-osoite), joka toimii tunnisteena eli jonka on tarkoitus olla pitkäaikainen. Cool URIa tulee hallinnoida siten, että jos dokumentin osoite muuttuu, käyttäjä uudelleenohjataan uuteen osoitteeseen. Cool URI:n erottaminen muista URL-osoitteista voi olla haasteellista tai mahdotonta, koska URL:stä ei suoraan näe miten sitä hallinnoidaan.

CURIE (Compact URI)

Lyhennysmerkintä jossa URI:ssa oleva (pitkä) nimiavaruus korvataan lyhenteellä. Esim. dct:title on CURIE-lyhenne URI:sta <http://purl.org/dc/title/>. CURIE-lyhenteen voi tehdä kaikista URI-osoitteista riippumatta protokollasta ja myös PID-tunnisteet kuten URN tai DOI voidaan haluttaessa esittää CURIEna.

Pitkäaikaissäilytys

Pitkäaikaissäilytyksellä tarkoitetaan elektronisen aineiston käytettävyyden takaamista vähintään 10 vuotta ja enimmillään pysyvästi. Koska teknologiat muuttuvat säilytysjakson aikana pelkkä bittien säilyttäminen (mikä ei sekään välttämättä ole helppoa) ei riitä, vaan dokumenttien ymmärrettävyyden säilyttäminen voi edellyttää migraation kaltaisia lisätoimia. Kansallisen digitaalisen kirjaston hankkeessa kehitetty, bittien säilytykseen keskittyvä PAS-ratkaisu sekä sitä soveltava, ymmärrettävyyden takaamiseen keskittyvä PAS-palvelu mahdollistavat elektronisten aineistojen kustannustehokkaan pitkäaikaissäilytyksen. PAS-ratkaisun ja –palvelun teknisestä toteutuksesta vastaa CSC, ja niitä

kehitetään yhdessä muistiorganisaatioiden ja Avoin tiede ja tutkimus –hankkeeseen osallistuvien organisaatioiden kanssa³³.

Pitkäaikaissäilytyksen metatiedot

Pitkäaikaissäilytyksen metatiedot ovat tietoja, joita säilytettävistä aineistoista täytyy tai voidaan tallentaa. PAS-järjestelmiin voidaan tallentaa esimerkiksi arkistoitujen tiedostojen tarkistussummat sekä tiedot kaikista aineistoihin kohdistuneista toimenpiteistä.

Säilytettävän resurssin pysyvä tunnus on yksi keskeisistä pitkäaikaissäilytyksen metatietojen elementeistä, koska sen avulla PAS-järjestelmään tallennettu resurssi kuten julkaisu tai asiakirja on helposti löydettävissä, ja saman resurssin versiot esimerkiksi eri tiedostoformaateissa voidaan luotettavasti linkittää toisiinsa. Viimeistään silloin kun organisaatiot päättävät ottaa käyttöön keskitetyn PAS-palvelun, niiden on varmistuttava siitä, että arkistoitavassa aineistossa käytetyt tunnisteet ovat pysyviä ja ainutkertaisia.

Linkkimätä ja sisältönyrjähdys

Linkkimätä tarkoittaa tilannetta, jossa aineistoa ei enää löydy verkosta (HTTP 404 –virhe) eikä edes verkkoarkistosta. Jos linkin takaa löytyvän aineiston sisältö on muuttunut merkittävästi tai kokonaan, resurssi kärsii sisältönyrjähdyksestä. Pysyvien tunnisteiden on tarkoitus olla immuuneja näille ongelmille. Käytännössä myös pysyvä tunniste voi lakata toimimasta, jos sitä ei hallinnoida asianmukaisesti.

Linkitetty data

Datan linkittämisen teknologioissa käytetään paljon erilaisia tunnisteita edustamaan erilaisia entiteettejä. Tietoa linkittämään käytetään usein RDF-tietomallia. Tietomallit kuvataan ontologioina ja yhteisesti viitatus resurssit SKOS-tietomallin mukaisena datana. Linkitetty avoin data hyödyntää perinteisesti myös ns. Cool URI-tunnisteita. Niihin perustuvat, URL-osoitteisiin tai niiden uudelleenohjaukseen perustuvat linkit ovat usein osoittautuneet varsin lyhytikäisiksi, mistä on aiheutunut laatuongelmia.

³³ <http://www.kdk.fi/fi/pitkaikaissailytykys>