



KDK:N TIETOARKKITEHTUURIRYHMÄN NIMITIETOPALVELUA KOSKEVA SELVITYS

Versio 1.0

Sisällys

1.	Johdanto.....	4
2.	Tavoitetila	5
3.	Nimitietopalvelun tietomalli.....	6
3.1.	Kuvailun nykytila KAM-sektoreilla	6
3.2.	Yhteismitallisuus ja yhteentoimivuus nimitietopalvelussa.....	6
3.3.	Kuvailun kohteet.....	8
3.4.	Toimijoiden yhteismitalliset kuvailutiedot	8
4.	Toimijoiden tunnusjärjestelmät.....	10
4.1.	OID-tunnusjärjestelmä.....	11
4.2.	ORCID JA ISNI - tunnusjärjestelmät.....	11
4.3.	Tunnusjärjestelmät ja Nimitietopalvelu.....	13
5.	Muiden tietovarantojen hyödyntäminen.....	13
5.1.	Julkishallinto.....	13
5.2.	Kansainvälisiä tietovarantoja.....	14
5.3.	Kansallisia tietovarantoja	15
6.	Lainsäädännöllisten esteiden poistaminen	16
7.	Nimitietopalvelun toteutuksesta	17
7.1.	Toteutusvaihtoehdoista.....	17
7.2.	Valmiin formaatin hyödyntäminen.....	18
7.3.	Oman formaatin kehittäminen.....	19
8.	Jatkoselvitysten ja –toimenpiteiden tarve	20
8.1.	Nimitietopalvelun arkkitehtuuri	21
8.2.	Muut selvitykset.....	21
9.	Lähteet	22
	Liite 2: SELVITYKSEN SANASTO	25

Tämä dokumentti käsittelee arkistojen, museoiden ja kirjastojen yhteistä Nimitietopalvelua. Nimitietopalvelulla tarkoitetaan tässä selvityksessä kulttuuriperintösektoreiden ja muun julkisen hallinnon organisaatioiden yhteistä, toimijoiden (luonnolliset ja juridiset henkilöt, yhteisöt, suvut) kuvailutietoja sisältävää metatietopalvelua. Yhteinen Nimitietopalvelu

- tehostaa KAM-sektoreiden toimintaa
- lisää tuottavuutta
- edistää julkisen sektorin yhteentoimivuutta
- edistää nimitiedon semanttista yhteismitallisuutta
- tehostaa kuvailuprosessia KAM-organisaatioissa
- vähentää päällekkäistä työtä toimijoiden metatietojen tuottamisessa koko julkishallinnossa
- tukee toimijoita kuvailevien metatietojen yhdenmukaistamisessa
- parantaa oleellisesti toimijoiden kuvailutietojen laatua, käytettävyyttä ja saatavuutta

Selvityksen on laatinut Kansallinen Digitaalinen Kirjasto (KDK) -hankkeen tietoarkkitehtuuriryhmä. Selvityksen tarkoituksena on tukea Nimitietopalvelun toteuttamista tunnistamalla ja tarkastelemalla toteuttamisen edellytyksenä olevia yhteistä sopimista edellyttäviä tietoarkkitehtuurikysymyksiä. Tietoarkkitehtuuriryhmän keskeisenä tavoitteena on ollut selvittää kulttuuriperintösektoreiden valmius yhteisen Nimitietopalvelun semanttisen yhdenmukaisuuden edellyttämään toimijoiden yhdenmukaiseen kuvailuun. Työryhmän työn tuloksena voidaan todeta, että tällainen valmius on olemassa. Selvityksen luvussa 3 esitetään tietoarkkitehtuuriryhmän näkemys Nimitietopalvelun tavoitetilaksi, jota on pyritty konkretisoimaan selvityksen liitteenä 1 olevalla luonnoksella tietomalliksi.

Tämän lisäksi selvityksessä tarkastellaan tunnusjärjestelmiin (luku 4), nimitietopalvelussa hyödynnettäviin tietovarantoihin (luku 5) sekä nimitietopalvelun toteuttamisen lainsäädännöllisten esteiden poistamiseen (luku 6) liittyviä kysymyksiä.

Kaikkiin selvityksessä tunnistettuihin yhteistä sopimista edellyttäviin tietoarkkitehtuurikysymyksiin liittyy lisäselvitystarpeita, jotka työryhmä on rajannut tämän selvityksen ulkopuolelle. Nämä lisäselvitystarpeet on esitetty yhteenvedonomaaisesti selvityksen luvussa 8. Nimitietopalvelun toteutuksen suunnittelu ja valmistelu ei sisälly tietoarkkitehtuuriryhmän toiminta-alueeseen, vaan se on organisoitava erikseen. Jatkovalmistelun tueksi selvityksen luvussa 7 kuitenkin avataan työryhmän visioiman tavoitetilän toteuttamisen keskeisiä edellytyksiä.

Selvityksen taustalla ovat olleet julkisen hallinnon yhteisen metatietopalvelun esiselvitys¹ sekä KDK:n kokonaisarkkitehtuurissa² mainitut yhteentoimivuuteen tähtäävät periaatteet ja linjaukset. Selvitystä on käsitelty myös julkisen hallinnon tietoarkkitehtuuriryhmässä.

¹ <https://www.avoindata.fi/data/fi/dataset/julkisen-hallinnon-yhteisen-metatietopalvelun-esiselvitys>

² <http://www.kdk.fi/index.php/fi/kokonaisarkkitehtuuri>

2. TAVOITETILA

Tässä luvussa esitetään tietoarkkitehtuuriryhmän näkemys Nimitietopalvelun tavoitetilasta. Luku toimii samalla yhteenvetona selvityksessä jatkossa yksityiskohtaisemmin tarkasteltuihin asiakokonaisuuksiin.

KDK:n tietoarkkitehtuuriryhmä visioi Nimitietopalvelun kulttuuriperintösektoreiden ja muun julkisen hallinnon organisaatioiden yhteisenä, toimijoiden (luonnolliset ja juridiset henkilöt, yhteisöt, suvut) kuvailutietoja sisältävänä metatietopalveluna. Tietovarantona Nimitietopalvelu sisältää toimijoiden erilaiset nimimuodot, kattavat kuvailutiedot, toimijoihin liittyvät suhteet sekä toimijoille että toimijakuvailuille annetut pysyvät tunnukset (tekstissä toimijatietovaranto/toimijatieto). Teknisenä toteutuksena Nimitietopalvelu koostuu yhteisestä toimijatietokannasta, sektorikohtaisista ja organisaatiokohtaisista tuotantojärjestelmistä sekä toimijatietokannan rajapinnoista ulkoisiin järjestelmiin kuten väestötietojärjestelmään tai kansainväliseen ISNI (International Standard Name Identifier) -tietokantaan. Tarjottavina palveluina Nimitietopalvelun voi nähdä jakautuvan nimiauktoriteettipalveluun, kuvailutietopalveluun sekä tunnuksien välityspalveluun.

Kulttuuriperintösektoreiden ja muun julkisen hallinnon organisaatioiden yhteinen Nimitietopalvelu tehostaa KAM-sektoreiden toimintaa ja lisää tuottavuutta, mutta ennen kaikkea se edistää julkisen sektorin yhteentoimivuutta ja toimijatietovarannon semanttista yhteismitallisuutta. Nimitietopalvelu tehostaa kuvailuprosessia KAM-organisaatioissa ja vähentää päällekkäistä työtä toimijoiden metatietojen tuottamisessa koko julkishallinnossa. Nimitietopalvelu mahdollistaa sen, että organisaatioiden ei tarvitse kuvailla kunkin erikseen samoja toimijoita, vaan kukin toimija kuvaillaan kertaalleen yhteisessä palvelussa, jota kaikki hyödyntävät ja johon voidaan poimia nimitietoja myös ulkoisista järjestelmistä. Yhteinen nimitietopalvelu parantaa oleellisesti toimijoiden kuvailutietojen laatua, käytettävyyttä ja saatavuutta. Finna-asiakasliittymään tai muuhun yhteiseen hakujärjestelmään liitettynä yhteinen Nimitietopalvelu mahdollistaa toimijoihin liittyvän aineiston löytämisen, vaikka nimistä käytettäisiin tiedonhaussa ja eri tuotantojärjestelmissä erilaisia versioita. Paikkansapitävä ja luotettava toimijatietovaranto tukee muidenkin kansallisten verkkopalveluiden rakentamista ja mahdollistaa kansainvälisten toimijatietokantojen yhteiskäyttöisyyden ja hyödynnettävyyden. Yhteinen Nimitietopalvelu myös purkaa julkisin varoin tuotetun tiedon käytön esteitä ja edistää valtionhallinnon ja EU:n avoimen tiedon linjauksia.

Toimijoiden kuvailutietojen semanttinen yhteismitallisuus Nimitietopalvelussa varmistetaan noudattamalla kuvailussa Resource Description and Access (RDA) -standardia. Muiden julkisen hallinnon organisaatioiden ja kansainvälisen yhteistyön tarpeisiin Nimitietopalvelusta tuotetaan toimijoiden tietoja EU:n Core Vocabularies -standardien (esim. EU Core Person, EU Core Business) määrittelemässä muodossa. ISNI-tietokantaan metatiedot tulee lähettää ISNI-standardin (ISO 27729³) määrittelemällä tavalla. Kirjastojen, arkistojen ja museoiden tuotantojärjestelmissä voidaan soveltaa jatkossakin sektorikohtaisia kuvailukäytäntöjä, jos ne eivät estä tai rajoita merkittävästi tietojen semanttista yhteismitallisuutta.

Tässä selvityksessä kulttuuriperintösektorin organisaatiot nähdään toimijatietovarannon keskeisiksi sisällöntuottajiksi. Kun jokin organisaatio luo uuden toimijakuvailun, se tallentaa Nimitietopalvelun tietomallissa pakolliseksi määritellyt kuvailutiedot ja mahdolliset muut tarvitsemansa tiedot yhteisesti

³ https://en.wikipedia.org/wiki/International_Standard_Name_Identifier

sovitulla tavalla. Käyttäjät voivat hyödyntää palvelusta tarvitsemiaan kuvailutietoja ja halutessaan täydentää niitä. Nimitietopalvelussa hyödynnetään standardirajapintojen kautta KAM-sektoreiden tietovarantojen lisäksi julkisen hallinnon perusrekistereiden tietovarantoja sekä kansainvälisiä toimijoiden nimitietovarantoja kuten kansainvälistä ISNI-tietokantaa⁴.

Nimitietopalvelu edellyttää laintasoista valtuutusta henkilötietojen keräämiseen, luovuttamiseen ja julkaisemiseen. Käyttörajoitettu toimijatieto on Nimitietopalvelussa vain auktorisoitujen käyttäjien saatavilla. Siltä osin kuin lainsäädäntö sen sallii eli esimerkiksi yhteisöjen tietojen osalta, toimijatietovaranto tarjotaan käyttöön Finto-palvelun kautta ontologisoituna avoimena linkitettyinä datana.

3. NIMITIETOPALVELUN TIETOMALLI

3.1. KUVAILUN NYKYTILA KAM-SEKTOREILLA

Kirjasto-, arkisto- ja museosektoreilla on kullakin joko jo käytössä tai ainakin suunnitteilla usean organisaation yhteinen henkilöiden, sukujen ja yhteisöjen nimitietoja sisältävä toimijatietokanta. Tieteellisillä kirjastoilla on tuotannossa Asteri-tietokanta. Arkistosektorin AHAA-hanke suunnittelee kuvailujärjestelmän käyttäjien yhteistä toimijarekisteriä, ja Museo2015-hankkeessa on tarkoitus hankkia museosektorille yhteinen kokoelmahallintajärjestelmä, jonka osana on järjestelmän käyttäjille yhteinen toimijatietokanta. Jokaisella sektorilla on lisäksi toimijatietoja tallentavia organisaatioita, jotka eivät ole nykyisten tai suunniteltujen yhteisten ratkaisujen piirissä.

Sektoreiden välisessä kuvailuyhteistyössä on havaittu paljon yhtymäkohtia toimijakuvailuissa, jonka seurauksena on selvitetty yhteisen nimitietopalvelun edellytyksiä. Selvitysten perusteella tiedetään, että sektoreiden yhteinen nimitietopalvelu mahdollistaisi toimijatietojen hyödyntämisen kaikissa kirjastoissa, arkistoissa ja museoissa yhdenvertaisesti. Keskitetty ylläpito vähentäisi merkittävästi päällekkäistä työtä ja riippuvuutta organisaatioiden omista tuotantojärjestelmistä, joissa mahdollisuudet toimijatietojen käsittelyyn voivat olla rajalliset. Nimitietopalvelu tukee toimijoiden kuvailutietojen yhdenmukaistamista, jolloin ne ovat paremmin hyödynnettävissä yhteisissä hakujärjestelmissä kuten Finnassa. Yhteiset toimijatiedot varmistavat sen, että toimijaan liittyvät aineistot löytyvät riippumatta siitä, onko kuvailu tehty kirjastoissa, arkistoissa, museoissa tai jossakin muussa julkishallinnon organisaatiossa.

3.2. YHTEISMITALLISUUS JA YHTEENTOIMIVUUS NIMITIETOPALVELUSSA

Nimitietopalvelun suunnittelun lähtökohtana ovat olleet sektoreiden uudet, käyttöönottovaliheissa olevat kuvailukäytännöt ja -standardit. Näiden pohjalta on osana tätä selvitystä kehitetty toimijoiden kuvailutietojen yhteistuotantoon soveltuva tietomalli.

Osana selvityksen laatimista tuotettu Nimitietopalvelun tietomalli (liite 1) sisältää kattavasti kirjastojen, arkistojen ja museoiden toimijoiden kuvailussa tarvittavat metatiedot. Tietomallissa määritellään myös näiden metatietojen yhteismitallisuus ja yhteentoimivuus sekä laatuvaatimukset Nimitietopalvelun metatiedoille. Tietomallin lähtöolettamuksena on ollut, että nimitietopalvelu rakentuu keskitetynä ratkaisuna, jota voidaan ylläpitää reaaliajassa myös nimitietopalvelua käyttävien organisaatioiden luettelointijärjestelmistä käsin. Tietomalli sisältää ne kuvailukentät, jotka selvityksen aikana on

⁴ <http://isni.oclc.nl/>

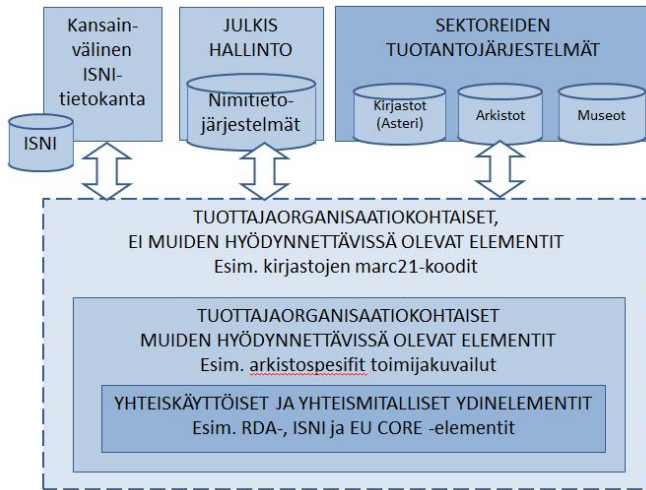
tunnistettu tarpeelliseksi. Tietomalliin tulee vielä jatkoselvitysten yhteydessä tehdä tarkennuksia, joiden tarkoituksena on ratkaista millaisella toteutus- ja ylläpitomallilla nimitietopalvelu kannattaa ja voidaan toteuttaa. Nämä enemmän tekniset ratkaisut määrittelevät sen, miten tietomallia tulee tarkentaa.

Yhteentoimivuus on edellytys nimitietojen tuotannon tehostamiselle: toimijan kuvailu luodaan vain kerran ja sitä rikastetaan ja siihen linkitetään muuta metatietoa yhteisessä tuotantoympäristössä. Yhteismitallisuus toteutuu yhteisten standardien ja metatietomääritysten kautta. Näitä ovat esimerkiksi RDA-kuvailusäännöt, ISNI-standardi ja EUn Core Vocabulary –määritykset. RDA-kuvailusäännöt toimivat semanttisen yhteismitallisuuden pohjana ja määrittelevät toimijatietokantaan tallennettavat kuvailuelementit ja sen millaista tietoa elementteihin sisällytetään. Julkishallinnon suosittelemien EUn Core Vocabulary –määritysten sisällyttäminen tietomalliin varmistaa toimijatietovarannon yhteentoimivuuden julkishallinnon muiden palveluiden kanssa. **Yhteismitallisuuden toteuttavien määritysten lisäksi tietomallissa on huomioitu arkistojen, museoiden ja kirjastojen käyttämien sektorikohtaisten metadataformaattien vaatimuksia mahdollisuuksien mukaan. Tällaisia formaatteja ovat ISAAR CPF ja EAC-CPF (arkistot), SPECTRUM (museot) ja MARC 21 auktoriteettitiedot-formaatti (kirjastot).**

Toimijatietovarannon suurimman kuvailuelementtien ryhmän muodostavat yhteiskäyttöiset ja yhteismitalliset ydinelementit. Niiden lisäksi tietomalliin sisällytetään joukko kirjastojen, arkistojen ja museoiden (l. tuottajaorganisaatiot) omia sektori- tai organisaatiokohtaisia elementtejä, joita muutkin Nimitietopalvelun käyttäjät voivat halutessaan hyödyntää.

Toimijakohtaisen identifioinnin ja kuvailun lisäksi Nimitietopalvelussa tunnistetaan ja kuvataan toimijoiden välisiä suhteita, minkä ansiosta Nimitietopalvelua voidaan pitää ontologiana. Suhteet kuvailaan suhdetyyppinä, joita voidaan lisätä toimijatietovarantoon RDA-kuvailusääntöjen asettamissa rajoissa sektoreiden tarpeiden mukaan. Suhteiden tunnistaminen ja kuvaaminen on oleellista tiedonhaun tukemiseksi yli sektorirajojen. Sekä kuvailu- että suhde-elementtien laadun varmistamiseksi elementteihin tallennettavia termejä ja ilmauksia vakioidaan yhteiseen Metatietosanastoon.

KAM-sektoreiden tuottamien toimijoiden metatietojen lisäksi Nimitietopalvelussa hyödynnetään muiden julkishallinnon toimijoiden tuottamia metatietoja tietosuojan sallimissa rajoissa. Keskeiset tällaiset metatietolähteet ovat Väestötietojärjestelmä ja Patentti- ja rekisterihallituksen rekisterit (Kaupparekisteri, Yhdistysrekisteri ja Säätiörekisteri). Niiden sovellettavuutta heikentää se, että toimijatietojen käyttötarkoitus ja sisältö on toinen kuin muistiorganisaatioissa.



Kuva 1 Nimitietopalvelun metatietovirtoja

3.3. KUVAILUN KOHTEET

Toimijalla tarkoitetaan luonnollista henkilöä, oikeushenkilöä tai näiden muodostamaa ryhmää.

Toimijat kuvaillaan kirjastoissa eri tavalla kuin museoissa ja arkistoissa. Kirjastojen kuvailussa kuvailun kohteena ovat toimijoiden julkiset identiteetit, joita voi olla yhdellä luonnollisella henkilöllä tai oikeushenkilöllä useita. RDA-kuvailusäännön ja ISNI-standardin määritelmien mukaisesti julkisia identiteettejä voivat olla mm. toimijan oikea nimi ja salanimi tai yhteisön aiempi ja myöhempi nimi. Museot ja arkistot puolestaan ovat kuvanneet toimijan yhtenä entiteettinä, joka sisältää kaikki toimijan identiteetit. Museoiden ja arkistojen kuvailukäytäntöjen kannalta julkinen identiteetti on uusi käsite.

Nimitietopalvelun tietomallissa pyritään ottamaan huomioon sekä arkistojen ja museoiden tapa kuvailla toimijat yhtenä kokonaisuutena että kirjastojen soveltama julkisten identiteettien kuvailu. Toimitettavan tietojärjestelmän pitäisi pystyä luomaan julkisten identiteettien kuvailut toimijan kuvailusta ja kääntäen yhdistämään julkisten identiteettien kuvailut toimijan kuvailuksi sektoreiden tarpeita varten.

3.4. TOIMIJOIDEN YHTEISMITALLISET KUVAILUTIEDOT

Liitteenä 1 olevassa Nimitietopalvelun tietomalliluonnoksessa esitetään kirjastojen, arkistojen ja museoiden kuvailukäytäntöihin sisältyvät toimijan kattavat kuvailutiedot. Tietomalli on edelleen keskenäinen ja edellyttää jatkotyöstämistä.

Alla on esitetty RDA:n, ISNI-tunnusjärjestelmän sekä EU:n Core Person ja Legal Entity-tietomallien mukaiset toimijan yksiselitteisen identifioinnin kannalta välttämättömät kuvailutiedot.

Taulukossa on pakolliset (tummennetut), ehdollisesti pakolliset (tummennettu kursivilla) ja vapaaehtoiset RDA ja ISNI-elementit (kursiivi).

RDA	ISNI	EU Core Vocabulary (Person ja Legal Entity)
HENKILÖT		
	<i>toimijan tyyppi</i>	

RDA	ISNI	EU Core Vocabulary (Person ja Legal Entity)
henkilön ensisijainen nimi	toimijan nimi	Name; Full Name; Birth Name; Given Name; Patronymic Name
syntymäaika	<i>aika ja ajan tyyppi</i>	Date of Birth
kuolinaika	<i>aika ja ajan tyyppi</i>	Date of Death
<i>henkilön toimintaan liittyvä ajanjakso</i>	<i>aika ja ajan tyyppi</i>	
<i>henkilön arvonimi</i>		
<i>henkilön ammatti tai ala</i>		
henkilön tunnus		Identifier
<i>nimen täydellisempi muoto</i>		Full Name
<i>henkilön varianttiniimi</i>		Alternative Name; Name; Full Name; Birth Name; Given Name; Patronymic Name
<i>sukupuoli</i>		Gender
<i>syntymäpaikka</i>	<i>paikka ja paikan tyyppi</i>	Location: Country Of Birth, Place Of Birth
<i>kuolinpaikka</i>	<i>paikka ja paikan tyyppi</i>	Location: Country Of Death, Place Of Death
<i>henkilöön liittyvä maa</i>	<i>paikka ja paikan tyyppi</i>	
<i>asuinpaikka</i>	<i>paikka ja paikan tyyppi</i>	Jurisdiction: Citizenship, Residence
<i>henkilön osoite</i>		Address
<i>viiteryhmä tai -yhteisö</i>		
<i>henkilön käyttämä kieli</i>		
<i>henkilön toiminta-alue</i>		
<i>elämäkerralliset tiedot</i>		
	aineistotyyppi (tyypitys aineistolle, johon toimijalla on suhde)	
	rooli (joka toimijalla on suhteessa aineistoon)	
	ulkoinen linkki (sisältää tietoa aineistotyyppistä ja roolista)	
	<i>toimijaan liittyvän toisen toimijan ISNI-tunnus</i>	
	<i>toimijan julkisten identiteettien suhde toisiinsa</i>	
YHTEISÖT		
<i>yhteisön ensisijainen nimi</i>	toimijan nimi	Legal Name
		Company Status
		Company Type
<i>yhteisön toimintaan liittyvä laitos</i>		
<i>yhteisön tunnus</i>		Identifier, Legal Identifier
konferenssin ajankohta	<i>aika ja ajan tyyppi</i>	
konferenssin paikka	<i>paikka ja paikan tyyppi</i>	
konferenssin tms. numero (<i>ehdollinen</i>)		

RDA	ISNI	EU Core Vocabulary (Person ja Legal Entity)
<i>yhteisöön liittyvä paikka</i>	<i>paikka ja paikan tyyppi</i>	Location: Location related to the business
<i>yhteisön toimintaan liittyvä ajanjakso</i>	<i>aika ja ajan tyyppi</i>	
<i>perustamisaika</i>	<i>aika ja ajan tyyppi</i>	
<i>lakkauttamisaika</i>	<i>aika ja ajan tyyppi</i>	
<i>yhteisöön liittyvä laitos</i>		
<i>yhteisön varianttiniimi</i>		Alternative Name
<i>yhteisön tyyppi</i>	toimijan tyyppi	
<i>hallintoalueen tyyppi</i>		
<i>yhteisön käyttämä kieli</i>		
<i>yhteisön osoite</i>		Address: Address related to the business; Address: Registered Address
<i>yhteisön toiminta-alue</i>		Company Activity
<i>yhteisön historia</i>		
	<i>toimijaan liittyvän toisen toimijan ISNI-tunnus</i>	
	<i>toimijan julkisten identiteettien suhde toisiinsa</i>	
SUVUT		
<i>suvun ensisijainen nimi</i>	toimijan nimi	
<i>suvun tyyppi</i>	toimijan tyyppi	
<i>sukuun liittyvä aika</i>	<i>aika ja ajan tyyppi</i>	
<i>suvun tunnus</i>		
<i>sukuun liittyvä paikka</i>	<i>paikka ja paikan tyyppi</i>	
<i>suvun tunnettu jäsen</i>		
<i>periytyvä arvonimi</i>		
<i>suvun varianttiniimi</i>		
<i>suvun käyttämä kieli</i>		
<i>sukuhistoria</i>		
	<i>toimijaan liittyvän toisen toimijan ISNI-tunnus</i>	
	<i>toimijan julkisten identiteettien suhde toisiinsa</i>	

4. TOIMIJOIDEN TUNNUSJÄRJESTELMÄT

Nimitietopalvelussa jokaisella toimijalla tai toimijan julkisella identiteetillä tulee olla ainakin yksi pysyvä tunnus. Jos toimijalla on useita julkisia identiteettejä, hänellä voi tunnus jokaiselle niistä.

Toimijoille on useita tunnusjärjestelmiä. Kansallisten järjestelmien lisäksi on olemassa kansainväliset järjestelmät ISNI (International Standard Name Identifier, ISO 27729)⁵ ja ORCID (Open Researcher and Contributor ID)⁶. Tunnukset eivät ole toisensa poissulkevia – samalla tutkijalla voi olla kansallinen tunnus ja lisäksi sekä ISNI että ORCID.

⁵ <http://www.isni.org/>

⁶ <http://orcid.org/>

Nimitietopalvelun kannalta relevantteja tunnusjärjestelmiä ovat Opetushallituksen oppijanumero, ORCID ja ISNI. Henkilötunnus on rajattu tarkastelun ulkopuolelle, koska tietosuojasyistä sen soveltaminen olisi nimitietopalvelussa vaikeaa. Lisäksi palveluun tulee paljon ulkomaisia tai historiallisia toimijoita, joilla ei suomalaista henkilötunnusta ole.

4.1. OID-TUNNUSJÄRJESTELMÄ

Opetushallituksen hallinnoima oppijanumero perustuu OID:hen (Object Identifier, ISO/IEC 8824-1:2002). Sen avulla voidaan tunnistaa periaatteessa mitä tahansa entiteettejä, mutta sitä ei käytetä toimijoiden tunnistamiseen muistiorganisaatioiden järjestelmissä, tieteellisessä tutkimustoiminnassa tai tutkimusdatahankkeissa. OID-pohjaisen kansallisen oppijanumeron käyttäminen KDK:ssa ja ATT-hankkeessa toimijoiden ensisijaisena tai ainoana tunnuksena edellyttäisi muutoksia käytössä oleviin sovelluksiin ja metatietoformaatteihin, eivätkä OID-tunnukset tämän jälkeenkään soveltuisi kansainväliseen tiedonvaihtoon, sillä ulkomaisissa järjestelmissä käytetään ORCID- ja ISNI-standarditunnuksia tai maiden omia kansallisia tunnuksia. Lisäksi oppijanumero kattaa vain Suomessa opiskelevat henkilöt, ei yhteisöjä eikä ulkomaisia tutkijoita.

4.2. ORCID JA ISNI - TUNNUSJÄRJESTELMÄT

ISNI on ISO-standardiin ISO 27729 perustuva toimijoiden julkisten identiteettien standarditunnus. Järjestelmää hallinnoi International Standard Name Identifier International Agency eli ISNI-IA. Standardista vastaa ISO:n tekninen komitea 46:n alakomitea 9 eli ISO TC 46/SC 9 (Information and documentation – Identification and description). 15.3.2016 mennessä ISNI-tunnuksia oli noin 9 miljoonaa, joista 500.000 oli annettu yhteisöille ja 2.5 miljoonaa tutkijoille. ISNI IA:n toiminnassa on mukana esimerkiksi kirjastoja ja tekijänoikeusjärjestöjä.

ORCID on Yhdysvalloissa rekisteröidyn non-profit -organisaation ORCID Incorporatedin hallinnoima, tutkijoiden tunnistamiseen tarkoitettu tunnusjärjestelmä. ORCID:in taustavoimia ovat esimerkiksi suuret tieteelliset kustantajat ja yliopistot. Toisin kuin ISNI, ORCID ei ole standardi, joten ORCID Inc. llä on suuri valta ja vastuu järjestelmän kehittäjänä.

ORCID ja ISNI ovat teknisesti identtisiä, 16 merkin mittaisia numeerisia tunnuksia. Sovellus, joka tukee yhtä näistä tunnuksista, tukee siis hyvin pienin muutoksin myös toista. Päällekkäisyyksien estämiseksi (sama numero olisi sekä ORCID että ISNI) ORCID-tunnuksille on varattu osa ISNI-tunnusavaruudesta. Tunnusten jakelumenetelmät ja kate eroavat kuitenkin merkittävästi toisistaan. ORCIDia voivat periaatteessa hakea vain tutkijat, kun taas ISNI voidaan antaa myös yhteisöille. Tietosuojamääräysten vuoksi suomalainen tutkija voi saada ORCID-tunnuksen vain hakemalla sitä itse. Ketään ei voida pakottaa hakemaan tunnusta itselleen. ISNI-tunnukset luodaan keskitetysti kansainvälisessä ISNI-tietokannassa yhteistyössä mukana olevien organisaatioiden toimittamien metatietojen pohjalta. ISNI-tunnuksen saaminen ei siis edellytä tunnuksen saavan henkilön omaa aktiivisuutta tai edes lupaa. Toisaalta henkilöt eivät välttämättä tiedä että heillä on ISNI-tunnus.

Kansainvälisissä tieteellisissä lehdissä julkaisevat tutkijat voivat saada ORCID-tunnuksen kirjoittautuessaan kustantajan tietojärjestelmään. Riippuen kustantajan asettamista vaatimuksista nimitiedot voivat olla kattavat ja ainakin aluksi ajan tasalla, ja niitä täydentää tieto henkilön ko. kustantajan välityksellä julkaisemista aineistoista. Mutta jos henkilö tallentaa tiedot itse suoraan ORCID-tietokantaan, minimitaso on niukka: etunimi ja sähköpostiosoite. Siksi ORCID-metatietojen perusteella tutkijan yksiselitteinen tunnistaminen voi olla vaikeaa tai mahdotonta.

ISNI-tietokannassa olevien nimitietojen taso vaihtelee suuresti; osa tiedoista on erittäin laadukkaita, osa hyvin puutteellisia. Koska nimitietoja tuottavat kirjastot, tekijänoikeusjärjestöt ja muut tahot, tallennuksessa ei ole sovellettu yhtenäisiä periaatteita, joten esimerkiksi yhteisöjen nimen muutoksia on käsitelty eri tavoin.

Kansainvälinen ISNI-keskus ei tarkista, kuuluuko jokin julkinen identiteetti tieteellisen tai taiteellisen luovan työn tekijälle. Käytännössä ISNI-tunnus voidaan hakea kenelle tahansa luonnolliselle tai juridiselle henkilölle tai niiden julkiselle identiteetille, jolle kirjastot, arkistot tai museot tarvitsevat pysyvän tunnuksen. Myöskään ORCID-järjestelmässä ei tarkisteta, kuuluuko julkinen identiteetti tutkijalle tai edes luonnolliselle henkilölle.

ISNI-tietokannassa on jo nyt paljon suomalaisten henkilöiden ja yhteisöjen julkisten identiteettien tietoja, ja etenkin monilla muusikoilla on jo ISNI koska Gramex Finland tuottaa tietokantaan nimitietoja. Tunnuksia ei kuitenkaan ole riittävästi KDK- ja ATT-hankkeiden tarpeita varten, koska tietokanta ei ole kattava ja siinä on virheitä. On paljon tapauksia, joissa kahden tai useamman suomalaisen julkisen identiteetin tiedot ovat yhdistyneet. Toisaalta samalla toimijalla voi myös olla kaksi tai mahdollisesti jopa useampia ISNI-tunnuksia. ISNI-järjestelmän kattava käyttöönotto Suomessa tutkijoiden tunnistamisessa edellyttää Kansalliskirjaston tallentamien nimitietojen päivittämistä ISNI-tietokantaan, sekä ISNI-tietokannassa olevien ongelmien riittävän kattavaa korjaamista suomalaisten toimijoiden osalta. Lisäksi Nimitietopalvelun tietokantaan on tallennettava sieltä puuttuvien tutkimusyhteisöjen ja tutkijoiden nimet, ja sovittava siitä, miten näitä tietoja pidetään ajan tasalla.

ORCID-tunnuksia luodaan yksi kerrallaan sitä mukaa kuin tutkijat niitä hakevat, eikä tunnuksia voida päivittää ohjelmallisesti nimitietokantaan johtuen ORCID-metatietojen vaihtelevasta ja usein heikosta laadusta. Mutta jos tiedot riittävät henkilön identifiointiin, tunnus voidaan lisätä nimitietopalveluun. Nykyisessä laajuudessaan ORCID on nimitietopalvelun kannalta kuitenkin toistaiseksi varsin marginaalinen järjestelmä, josta on apua tutkijoiden tunnistamisessa vain jos ORCID-kannan metatiedot ovat riittävän kattavat ja ajan tasalla.

ISNI Member -organisaationa Kansalliskirjasto voi jo nyt hakea ISNI-tunnuksia kaikille suomalaisille toimijoille, joista on nimitietoja Asteri-tietokannassa. ISNI-tietokantaan tehtävällä suomalaisten nimitietojen eräpäivityksellä tunnukset saadaan helposti suurelle osalle Asterissa olevia toimijoita. ISNI-tietokannan metatietojen laadun vaihtelun vuoksi joidenkin toimijoiden tiedot on kuitenkin korjattava käsin ennen kuin tunnus voidaan antaa. Tämän jälkeen suomalaisten toimijoiden ISNI-tunnukset ja muita tietoja voidaan poimia ISNI-tietokannasta käyttäen ISNI-tietokannan tarjoamaa ohjelmistorajapintaa.

Eräpäivitysten ohella ISNI-tunnuksia voidaan hakea toimijoille yksitellen hintaan 20 senttiä / tunnus. Tämä oikeus on kuitenkin vain ISNI Registration Agency - eli RAG-organisaatioilla ja uusitun ISNI-sopimuksen mukaan myös Federated Registration Agency -yhteenliittymien jäsenillä. Kansalliskirjasto on toistaiseksi valmistellut vain RAG-vaihtoehtoa, mutta myös Federated RAG voi tulla Suomessa kyseen, jos ISNI-tunnuksia tarvitsevia organisaatioita löytyy riittävästi.

Kansainvälisen yhteistyön tärkein tavoite on toimijoiden yksiselitteinen tunnistaminen ISNI- tai ORCID-tunnuksen avulla. Näiden standarditunnusten saaminen suomalaisille toimijoille on työvoimavaltainen prosessi, joka edellyttää ohjelmistokehityksen lisäksi myös paljon henkilötyötä. Kansalliskirjasto on laatinut OKM:lle hakemuksen ISNI-järjestelmän kattavasta käyttöönotosta Suomessa. Siinä alustavasti kuvattua ISNI-tunnuksen käyttöönottoprosessia ei toisteta tässä. Vastaavasti OKM on laatinut

selvityksen ORCID-järjestelmän käyttöönotosta⁷. Molemmat tunnukset voidaan jo nyt tallentaa muutamisiin muistiorganisaatioiden taustajärjestelmiin kuten kirjastojärjestelmiin.

4.3. TUNNUSJÄRJESTELMÄT JA NIMITIETOPALVELU

Nimitietopalvelun kannalta tunnusten käytettävyydessä on merkittävä ero. ISNI-tunnukseen on aina liityttävä toimijan identifioiva metatieto. Mutta koska kansainvälinen ISNI-tietokanta on uusi palvelu, nämä metatiedot voivat olla virheellisiä tai puutteellisia. Siksi ISNI-järjestelmän käyttöönotto saattaa edellyttää ISNI-tietokannassa olevien nimitietojen manuaalista tarkistamista ja tarvittaessa korjaamista. Tämä voi vaatia merkittävän määrän henkilötyötä, jos ongelmaticetueita on paljon.

ORCID-tietokannasta ei välttämättä löydy riittäviä nimitietoja tunnuksen saaneista henkilöistä. Minimitiedot (etunimi ja sähköpostiosoite) eivät riitä muistiorganisaatioille sen selvittämiseen, kenelle ORCID-tunnus kuuluu. Lisäksi sama henkilö voi hakea itselleen joko vahingossa tai tarkoituksella toisenkin tai jopa useampia ORCID-tunnuksia.

Tässä selvityksessä lähtökohtana pidetään, että yhteisöjen tunnistamiseen tulisi soveltaa ISNI-tunnusta. Henkilöiden osalta ORCID-järjestelmää voidaan soveltaa tutkijoihin, mutta ISNI on paras vaihtoehto silloin, kun toimija ei ole tutkija tai ei halua tai pysty hankkimaan ORCID-tunnusta itse. Samalla henkilöllä voi olla sekä ORCID että ISNI, jolloin ne on molemmat tallennettava nimitietopalveluun. Palvelun tulee sisältää myös mahdolliset tuplatunnukset sekä virheelliset tunnukset.

Linkitetyn datan tuottaminen edellyttää, että toimijan kuvailuille⁸ annetaan niin sanottu pysyvä tunnus eli PID, jota voidaan soveltaa linkkien luontiin esimerkiksi Wikipediasta kuvailuun. Finto-järjestelmässä nimitietojen PID-tunnuksena sovelletaan URN:NBN-tunnusta. Esimerkiksi Nokian kauppalan Finto-kuvailun URN on <http://urn.fi/URN:NBN:fi:au:cn:18181A>, Nokian kaupungin <http://urn.fi/URN:NBN:fi:au:cn:39666A> ja Nokia-yhtiön <http://urn.fi/URN:NBN:fi:au:cn:52993A>.

5. MUIDEN TIETOVARANTOJEN HYÖDYNTÄMINEN

5.1. JULKISHALLINTO

Arkistolain jatkotyöryhmän toukokuussa 2015 julkaistussa muistiossa⁹ on luonnos Laiksi tiedonhallinnasta ja tietojen käsittelystä julkishallinnossa. Lakiehdotuksen 3 §:ssä esitetään, että ellei laissa säädetystä veloitteesta tai erityisistä syistä muuta johdu, viranomaisten on käytettävä toiminnassaan seuraavia yleiseen käyttöön tarkoitettuja rekistereitä:

Väestörekisterikeskus

- Väestötietojärjestelmä¹⁰

⁷ <https://confluence.csc.fi/display/tutki/ORCID-selvitys+suomeksi>

⁸ Tietueilla voi olla tietokannan sisäiset tunnukset jotka eivät välttämättä ole pysyviä. Kuvailun tunnus sen sijaan on pysyvä ja sen tulisi säilyä formaattikonversioissa ja tietokannasta toiseen kopioitaessa.

⁹ Laki tiedonhallinnasta ja tietojen käsittelystä. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2015:12. http://minedu.fi/OPM/Julkaisut/2015/arkistolain_jatkotyoryhma.html?lang=fi

¹⁰ <https://evertti.vrk.fi/>

Väestötietojärjestelmän tiedot eivät tätä kirjoitettaessa ole julkisia. Niiden käyttö edellyttää aina Väestörekisterikeskuksen käyttäjäorganisaatiolle myöntämää, väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista annetun lain (661/2009) mukaista tietolupaa sekä käyttäjäorganisaation sitoutumista tietoluvan ehtoihin. Tietolupaa väestötietojärjestelmän tietojen käyttöön saamiseksi haetaan kirjallisesti Väestörekisterikeskukselta. Keskeinen hakemuksessa selvitettävä asia on tietojen käyttötarkoitus.

Väestötietojärjestelmän tietoja käsiteltäessä on noudatettava tietosuojaa ja tietojenkäsittelyn turvallisuutta koskevaa lainsäädäntöä ja viranomaisten määräyksiä.

Patentti- ja rekisterihallitus

- Yhdistysrekisteri <http://www.prh.fi/fi/yhdistysrekisteri.html>
- Säätiörekisteri <http://www.prh.fi/fi/saatiorekisteri.html>
- Kaupparekisteri <http://www.prh.fi/fi/kaupparekisteri.html>
- Yritys- ja yhteisötietojärjestelmä <https://www.ytj.fi/>

Muistiorganisaatioista ainakin Kansalliskirjasto käyttää jo nyt PRH:n rekistereitä hyväksi yhteisöjen nimitietojen tallennuksessa. PRH:n kokoamat yhteisöjen perustiedot ovat maksutta haettavissa Virre-palvelusta¹¹. Tietueita ei kuitenkaan voi siirtää eränä, vaan vain yksitellen. Kuvailut ovat kattavia, esimerkiksi yrityksistä kerrotaan mm. kaikki edeltäjät.

Lakiluonnos ei kuvaa näiden rekistereiden soveltamistapaa. Käytännössä toimijatietojen manuaalinen haku ja poiminta Nimitietopalveluun ei ole riittävän tehokasta, vaan tarvitaan automatisoituja eräpäivityksiä. Niiden tekoon tarvitaan jaettu tiedonsiirtoväylä kuten X-Road, molempien järjestelmien ymmärtämä tiedonhaun / poiminnan rajapintamääritys ja yhteinen toimijatietojen vaihtoformaatti.

Toistaiseksi ei ole tietoa siitä, mitä vaihtoformaattia esimerkiksi PRH:n järjestelmät tukevat, mutta EU Core Legal Entity on mahdollinen vaihtoehto. Jos se valitaan vaihtoformaatiksi, Nimitietopalveluun tulee voida siirtää tässä muodossa yritysten, yhteisöjen, yhdistysten ja säätiöiden tiedot Patentti- ja rekisterihallituksen ylläpitämistä tietojärjestelmistä eräpäivityksinä julkishallinnon palveluväylää käyttäen. Tarvittavat formaattikonversiot (PRH:n rekistereiden sisäisestä formaatista EU Core Legal Entity -muotoon ja siitä edelleen Nimitietopalvelussa sovellettuun formaattiin) tulisi laatia yhteistyössä PRH:n kanssa muunnosten laadun optimoimiseksi.

Myös väestötietojärjestelmää tulee voida käyttää henkilöiden tietojen eräpoimintaan. Nämä poiminnat eivät olisi välttämättömiä kirjastoille, koska ne tarvitsevat kuvailutyössään yleensä vain luovan työn tekijöiden tietoja. Mutta esimerkiksi arkistoilla on tuhansia ”tavallisten” ihmisten arkistoja ja museotkin tarvitsevat hyvin erilaisten toimijoiden tietoja.

5.2. KANSAINVÄLISIÄ TIETOVARANTOJA

Kansainvälisen yhteistyön avulla voidaan tehostaa toimijatietojen tallennusta, mutta välttämätöntä kansainvälisyys on silloin, kun tieteellisen tai taiteellisen luovan työn tekijöille hankitaan julkisesti jaettavissa ja sovellettavissa olevia tunnuksia.

Kirjastosektori on ottamassa toimijan kuvailussa käyttöön ISNI-tunnusjärjestelmän. Tunnuksen saaminen edellyttää toimijoiden tietojen tallentamista kansainväliseen ISNI-tietokantaan. ISNI-tunnuksen

¹¹ <https://virre.prh.fi/>

lisäksi kirjastot voivat käyttää hyväksi myös ORCIDia tai muita julkisen identiteetin tunnuksia. Arkisto- ja museosektorit eivät ole määritelleet yhteistä tai suositeltavaa tunnusjärjestelmää, vaan niiden taustajärjestelmissä hyväksytään erilaisten järjestelmien käyttö. Esim. arkistosektorilla tämä on ollut välttämätöntä, koska viranomaisten ja muiden yhteisöjen aineistot muodostuvat näiden toimijoiden käytännön toiminnan tarpeisiin, jotka määrittelevät myös tarkoituksenmukaiset tunnusjärjestelmät.

Varsinkin kirjastot ovat jo pitkään tehneet kansainvälistä yhteistyötä toimijatietojen tallennuksessa. OCLC:n¹² VIAF¹³ (The Virtual International Authority File) perustuu WorldCat-yhteisluetteloon¹⁴ lähetettyjen bibliografisten tietueiden sisältämiin toimijatietoihin. VIAF-tietokannassa on paljon suomalaisten luonnollisten ja juridisten henkilöiden nimiä, jotka ovat peräisin kansallisbibliografiasta ja ulkomaisista kirjastotietokannoista. Tietueet voivat olla korkeatasoisia, esimerkiksi Alvar Aalosta on tallennettu runsaasti nimenmuotoja eri kirjoitusjärjestelmissä kuten koreaksi (알바르 알토) ja arabiaksi (ألفار ألتو). Tietueisiin sisältyy sekä ISNI-tunnus (jos se on olemassa) että VIAF-tietokannan sisäinen tunnus. Alvar Aallon ISNI on 0000 0001 0913 6025 ja VIAF-tunnus 714 10602.

Kattavien nimitietokantojen ohella on olemassa erikoisalojen palveluita, kuten Getty Foundationin ylläpitämä ULAN eli Union List of Artist Names¹⁵, jonka kuvailutiedot ovat poikkeuksellisen rikkaat; tietokannan avulla voi selvittää esimerkiksi kuvataiteilijoiden opettajia ja oppilaita ja löytää taiteilijoista kertovia lähdeteoksia.

5.3. KANSALLISIA TIETOVARANTOJA

Yksittäiset muistiorganisaatiot ja muut toimijat kuten tekijänoikeusjärjestöt ovat tuottaneet omiin tarpeisiinsa merkittäviä tietyn toimijaryhmän kuvailutietovarantoja, joiden tulisi olla Nimitietopalvelun hyödynnettävissä. Tämä tukisi julkisen hallinnon kokonaisarkkitehtuurin tavoitteiden mukaista päälekkäistyön vähentämistä.

Kulttuuriperintösektoreiden merkittäviä toimijoiden kuvailun metatietovarantoja ovat esimerkiksi

- Arkkitehtuurimuseon kattavat kuvailutiedot arkkitehteista,
- Valokuvataiteen museon kattavat kuvailutiedot valokuvaajista ja valokuvaamoista
- Kansallisen audiovisuaalisen instituutin kattavat nimitiedot elokuva-alan toimijoista
- Suomen Taiteilijaseuran Kuvataiteilijamatrikkeli¹⁶
- Arkistolaitoksen kattavat kuvailutiedot valtion viranomaisista
- Yksityisten arkistojen kuvailutiedot oman hankinta-alueensa yhteisö- ja henkilötoimijoista

Tekijänoikeusjärjestöillä on kattavia rekistereitä omien vastuualueidensa toimijoista, kuten Sanastolla ja Kopiostolla kirjailijoista ja Gramex Finlandilla muusikoista. Toistaiseksi kansallista yhteistyötä ja toimijoiden nimitietojen vaihtoa on näiden organisaatioiden ja muistiorganisaatioiden välillä verraten vähän. Kansalliskirjasto toimittaa julkaisujen tietoja Kopiostolle ja Sanastolle, ja niihin sisältyvät myös tiedot tekijöistä, mutta varsinaisia nimitietoja ei toimiteta.

¹² <http://www.oclc.org/>

¹³ <https://viaf.org/>

¹⁴ <http://www.worldcat.org/>

¹⁵ <http://www.getty.edu/research/tools/vocabularies/ulan/>

¹⁶ <http://www.kuvataiteilijamatrikkeli.fi/fi/etusivu.html>

Nimitietokanta sekä toimijoiden kuvailussa tehdyt sääntömuutokset (kuten kaikkien toimijoiden mainitseminen kuvailussa korkeintaan kolmen toimijan mainitsemisen sijaan) luovat pohjaa muistiorganisaatioiden välisen yhteistyön tiivistämiselle. Tiedon siirto muille edellyttää kuitenkin sekä teknisiä että juridisia selvityksiä.

Yhteistyön lisäämiselle on käytännön tarve muun muassa siksi, että Gramexin järjestelmään tallennetut muusikoiden nimitiedot päätyvät kansainvälisen yhteistyökumppanin kautta ISNI-tietokantaan. Samojen julkisten identiteettien nimitietojen lähettäminen Viola-tietokannasta tulee aiheuttamaan ongelmia, koska Kansalliskirjaston ja Gramexin kuvailun periaatteet ovat erilaiset.

Toimijatietojen vaihdon tekniset edellytykset on selvitettävä järjestelmäkohtaisesti. Esimerkiksi Saaston ja Kopioston järjestelmät pystyvät ottamaan vastaan kirjastojen ISO 2709 -vaihtomuotoa, mutta Gramex Finlandin sovellukseen, jota ollaan uudistamassa, ei ole standardirajapintoja.

6. LAINSÄÄDÄNNÖLLISTEN ESTEIDEN POISTAMINEN

Informaatio- ja tietoteknologiaoikeuden professori Tomi Voutilainen toteaa laatimassaan palveluarkkitehtuuria tukevan lainsäädännön kehittämistarpeiden tutkimusraportissa¹⁷, että palveluarkkitehtuurin kehittäminen edellyttää myös keskustelua voimassa olevan lainsäädännön toimivuudesta sekä keskustelua uusien ratkaisujen käyttöönoton edellyttämästä sääntelyn kehittämisestä. Nimitietopalvelun kannalta keskeinen lainsäädännöllinen este on henkilötietolainsäädäntö, joka estää henkilötoimijoiden kuvailutietojen luovuttamisen eli jakamisen ja julkaisemisen, ellei siihen ole henkilön suostumusta tai ellei jokin laki anna oikeutta tietojen luovuttamiseen. Muistiorganisaatiot ovat lähettäneet marraskuussa 2015 tietosuojalautakunnalle yhteisen hakemuksen siitä, että ne voisivat keskenään jakaa henkilötietoja virkatehtäviensä hoitamista varten.

Kulttuuriperintösektorilla jokainen kirjasto, arkisto ja museo on erillinen rekisterinpitäjä, jolla tulisi olla laintasoinen valtuutus henkilötietojen jakamiseen ja julkaisemiseen. Voutilaisen mukaan palveluarkkitehtuurin hajautuminen tiedonhallinnan tasolla lähtee julkisuuslakiin sisältyvästä jyrkästä viranomaisten erillisyyperiaatteesta sekä henkilötietolain viranomaiskohtaisista tulkinnoista rekisterinpitäjyydestä. Sinänsä henkilötietolaki sallisi Voutilaisen mukaan yhteisen rekisterinpitäjyyden silloin, kun henkilörekisterissä olevia tietoja käytetään samaan ennalta määriteltyyn käyttötarkoitukseen.

Kansalliskirjasto, Kansallisarkisto ja Museovirasto ovat kesällä 2015 käynnistäneet yhteistyön, jonka tavoitteena on aikaansaada Nimitietopalvelun edellyttämä laintasoinen valtuutus henkilötietojen luovuttamiseen ja julkaisemiseen. Valtuutusta tarvitaan yhtäältä henkilötietojen luovuttamiseen KAM-organisaatioista toimijatietokantaan ja toimijatietokannasta KAM-organisaatioille ja muille julkishallinnon organisaatioille. Toisaalta henkilötietojen julkaiseminen edellyttää laintasoinen valtuutusta sekä Nimitietopalvelulle että sitä hyödyntäville organisaatioille. Tässä selvityksessä pidetään lähtökohtana, että Nimitietopalvelun koko tietosisältöä saavat hyödyntää vain auktorisoidut käyttäjät. Kaikkia Nimitietopalveluun kerättyjä aineistojen tunnistamisen ja valinnan kannalta välttämättömiä henkilötietoja

¹⁷ Voutilainen, Tomi: Palveluarkkitehtuuria tukeva lainsäädäntö. Helsinki: Valtiovarainministeriö, 2014. (VM:n julkaisuja 22/2014). <http://vm.fi/dms-portlet/document/354409>

ei siis julkaistaisi avoimessa verkossa. Nimitietopalvelun, sitä koskevan laintasaisen valtuutuksen perusteella, julkaistavissa olevien henkilötietojen jakelu ja julkaiseminen hoidetaan erillisenä palveluna Finton kautta.

Nimitietopalvelua koskevan lainsäädäntöhankkeen käynnistäminen edellyttää opetus- ja kulttuuriministeriöltä sekä valtiovarainministeriöltä aloitteellisuutta suhteessa oikeusministeriöön, jonka toimialaan Nimitietopalvelun käyttöönoton edellyttämä lainsäädäntöhanke kuuluu. Myös palveluarkkitehtuuria tukevan lainsäädännön kehittämistarpeiden tutkimusraportissa esitetään, että valtiovarainministeriö käynnistäisi toimet kansallista palveluväylää ja siihen liittyviä palveluja koskevan lainsäädännön valmistelemiseksi. Nimitietopalvelu tulisi ottaa huomioon myös valtiovarainministeriön valmisteluvastuulla olevassa tiedonhallinnasta ja tietojen käsittelystä julkishallinnossa annetun lakiesityksen (opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2015:12) jatkotyöstämisessä tiedonhallintalainsäädännöksi osana hallituksen kärkihankkeita.

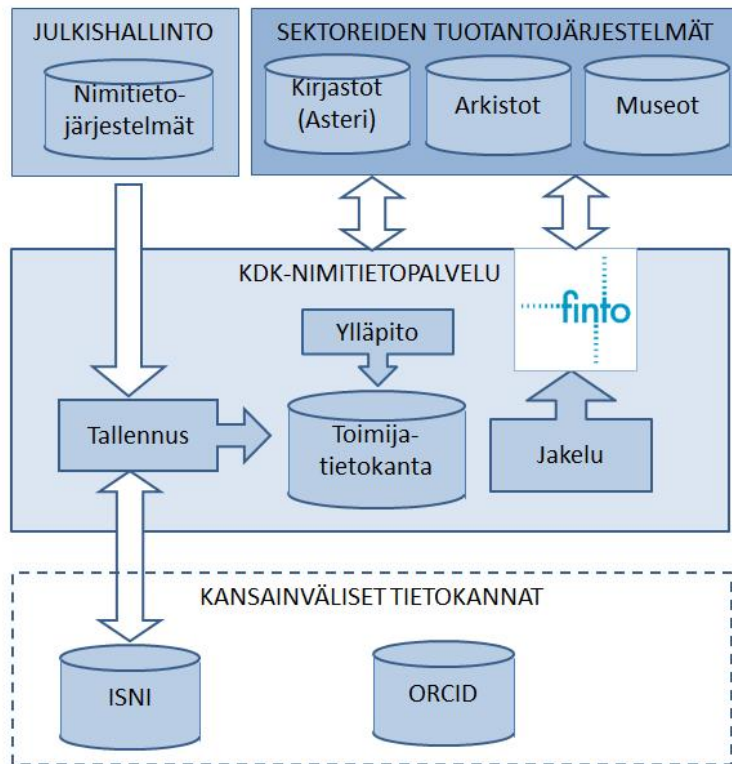
7. NIMITIETOPALVELUN TOTEUTUKSESTA

7.1. TOTEUTUSVAIHTOEHDOSTA

Tässä luvussa tarkastellaan Nimitietopalvelun tekniseen toteutukseen liittyviä reunaehtoja toimijatietovarannon tietokantarakenteen, käyttöliittymän sekä tietovarannon ja sektoreiden tuotantojärjestelmien välillä vaadittavien konversioiden näkökulmasta.

Tässä selvityksessä Nimitietopalvelun on teknisenä toteutuksena ymmärretty koostuvan yhteisen toimijatietovarannon tallennukseen, ylläpitoon ja jakeluun käytettävästä toimijatietokannasta sekä sektori- ja organisaatiokohtaisista tuotantojärjestelmistä. Tietoarkkitehtuuriryhmän selvityksen luvussa 2 visioimassa tavoitetilassa Nimitietopalvelun toimijatietovaranto muodostuu toimijoiden erilaisista nimimuodoista, toimijoiden kattavista kuvailutiedoista, toimijoiden välisten suhteiden kuvauksista sekä toimijoille ja toimijakuvailuille annetuista pysyvistä tunnisteista.

Nimitietopalvelua voidaan päivittää joko eräajoina tuotantojärjestelmistä poimituilla tiedoilla tai suoraan erillisen asiakassovelluksen avulla. Asiakassovellus on välttämätön esimerkiksi silloin, kun palvelun sisältämiin nimitietoihin tehdään tietueiden yhdistämisen tai erottamisen kaltaisia vaativia teknisiä korjauksia. Toimijoiden kuvailutietojen jakelun KDK:n ulkopuolelle julkishallintoon on tässä selvityksessä nähty tapahtuvan Finto-palvelun avulla, joka on tässä roolissaan Nimitietopalvelu-kokonaisuuden osa. Tavoitetilan mukainen järjestelmäarkkitehtuuri on esitetty kuvassa 2.



Kuva 2 Nimitietopalvelun järjestelmäarkkitehtuuri

Kulttuuriperintösektoreiden yhteisistä Nimitietopalveluista ei ole olemassa kansainvälisiä malleja. Tietoarkkitehtuuriryhmän tavoitetila, joka laajentaa Nimitietopalvelun kulttuuriperintösektoreiden yhteiseksi toimijatietovarannoksi, lisää teknisen toteutuksen haasteellisuutta, koska eri sektorit käyttävät eri formaatteja ja kuvailua on tähän asti tehty eri periaattein.

Keskeinen tekninen kysymys toteutuksen osalta on, onko Nimitietopalvelun tavoitetilan mukaisen toimijatietokannan tallennukseen, ylläpitoon ja jakeluun käytettävissä jonkin sektorin olemassa olevaa formaattia (skeemaa) vai pitääkö palvelulle luoda oma formaatti, jotta se palvelisi kaikkia sektoreita tasapuolisesti. Mikäli toimijatietokanta on mahdollista rakentaa valmiin formaatin pohjalta, tulee tietoarkkitehtuuriryhmän visioiman tavoitetilan toteuttaminen olemaan merkittävästi yksinkertaisempaa kuin jos Nimitietopalvelun tietomallin toteuttava formaatti suunnitellaan ja toteutetaan itse.

7.2. VALMIIN FORMAATIN HYÖDYNTÄMINEN

Käytännössä vaihtoehtoja toimijatietokannan valmiiksi formaatiksi on kaksi: kirjastojen MARC 21 Auktoriteettitiedot sekä arkistojen Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF). MARC 21 –pohjaisia MARCXML - ja MADS -formaatteja ei ole otettu huomioon. Society of American Archivists –organisaation tekemän formaattivertailun¹⁸ perusteella MARC 21:n ja EAC-CPF:n keskinäinen yhteismitallisuusaste on korkea. MARC 21:stä soveltaa valtaosa Suomen (ja maailman) kirjastoista, EAC-CPF on käytössä arkistoissa kansainvälisesti ja kansallisesti arkistolaitoksessa. Formaatti on tulossa käyttöön arkistojen yhteisessä AHAA-palvelussa.

¹⁸ http://www2.archivists.org/standards/DACS/appendices/appendix_c_crosswalks

Nimitietopalvelun tietomallissa on tehty alustavaa vastaavuuksien määrittelyä EAC-CPF formaattiin, mutta tämä määrittely täytyy jatkovalmistelussa systematisoida ja lisäksi on määriteltävä myös tietomallin MARC 21 -vastaavuus. Vasta tämän jälkeen voidaan arvioida formaateista puuttuvien tietoelementtien perusteella, miten suuri tarve toimijatietokannan oman formaatin laadintaan on. Valmiiden formaattien soveltavuuteen vaikuttaa myös niiden yhteismitallisuuden aste tuotantojärjestelmien formaattien ja niiden soveltamiskäytänteiden kanssa.

Konversioiden osalta toimintamalli on se, että eri sektoreilla tuotetut toimijoiden tietovarannot muunnetaan valittuun formaattiin ennen (erä)päivitystä yhteiseen toimijatietokantaan. Tämä edellyttää kahdelta sektorilta tarvittavien konversioiden (omasta formaatista tai formaateista toimijatietokannan formaattiin ja takaisin) rakentamista. Konversioiden häviöllisyyden aste riippuu tuotantojärjestelmissä käytössä olevista formaateista sekä toimijatietokantaan siirrettävien tietoelementtien määrästä. Osa formaattikonversioista saataisiin valmiina, mutta muutamia jouduttaisiin rakentamaan. Suurin haaste on se, että formaattien väliset konversiot ovat yleensä aina häviöllisiä. Se, paljonko tietoja tosiasiallisesti menetetään, riippuu toimijatietokannan ja tuotantojärjestelmien formaattien semanttisen yhteismitallisuuden lisäksi siitä, miten formaatteja sovelletaan käytännön luettelointityössä. Konversio-ohjelmien saatavuudesta ei ole tietoa, joten tämäkin pitää selvittää osana toteutuksen suunnittelua.

Oleellisin tekijä on kuitenkin Nimitietopalvelun käyttöön soveltuvien ohjelmien saatavuus. Jos toimijatietokanta on mahdollista rakentaa valmiin formaatin pohjalta, ohjelmiston jatkokehittämistarve olisi luultavasti varsin vähäinen. Järjestelmän perustaksi voitaisiin valita jokin olemassa oleva valitun formaatin mukainen (kirjasto/arkisto-)ohjelmisto. Toimijatietokannan luontiin, eräpäivityksiin ja tietokannan huoltoon voitaisiin käyttää valmisohjelmia, joita jo nyt sovelletaan valitussa formaatissa olevien toimijatietojen käsittelyyn.

Toimijatietojen tallennus suoraan toimijatietokantaan ja tietokannan manuaaliset korjaukset (esimerkiksi samaa julkista identiteettiä koskevien tietueiden yhdistäminen) voidaan tehdä joko sovelluksen omalla luettelointisovelluksella, tai erikseen laadittavalla asiakasohjelmalla.

Jos käytetään sovelluksen omaa luettelointiohjelmistoa, yhden sektorin luetteloijilla olisi jo tarvittava formaattikoulutus, mutta muiden pitäisi se hankkia. Jos taas laaditaan oma ohjelmisto, se voisi tukea periaatteessa kaikkia sektorikohtaisia formaatteja tai tarjota oman, formaattiriippumattoman tallennuslujan. Tällöin kaikki sektorit olisivat tallennuksen kannalta tasaveroisessa asemassa vaikka toimijatietokannan formaatti olisin yhden sektorin ”oma”. Erillisen luettelointisovelluksen rakentaminen lisäisi palvelun kehittämis- ja ylläpitokustannuksia.

Kansainvälisistä muistiorganisaatioiden tietokannoista poimitut nimitiedot olisivat yhden sektorin osalta suoraan käytettävissä, muille olisi valmiina käytettävissä konversion apuvälineitä. Toimijatietovarannon Finto-jakeluun tarvittaisiin vain yksi konversio, joka olisi MARC 21 - SKOS -vaihtoehdossa jo valmiiksi käytettävissä; EAC-CPF:lle se tulisi rakentaa. Konversiot toimijatietokannan formaatista Fintossa käytettävään väljästi SKOS-pohjaiseen formaattiin on mahdollista saada häviöttömäksi, koska Finton SKOS-versio on vapaasti muokattavissa. Kansainvälinen yhteistyö esimerkiksi ISNI-tunnusten haussa voitaisiin toteuttaa suoraan Nimitietopalvelun toimijatietokannasta. EU Core Vocabulary -konversio tulisi rakentaa.

7.3. OMAN FORMAATIN KEHITTÄMINEN

Jos Nimitietopalvelun tietomalli ei edellyttäisi yhdenkään tai korkeintaan vain joidenkin uusien meta-tietoelementtien lisäämistä olemassa olevaan formaattiin tai jonkin/joidenkin metatietoelementtien soveltamistavan muuttamista, olisi periaatteessa mahdollista käyttää olemassa olevaa formaattia ja siihen perustuvaa sovellusta. Jos kuitenkin osoittautuu, että sen enempää MARC 21 kuin EAC-CPF –formaatti eivät, mahdollisine muutoksineenkaan, mahdollista tietoarkkitehtuuriryhmän vision mukaista tavoitetilaa, on Nimitietopalvelulle kehitettävä oma formaatti tai tarkistettava Nimitietopalvelun tavoitetilaa.

Oma formaatti olisi kansallinen ratkaisu, jollaista ei olisi käytössä missään muualla. Siihen sisältyisivät kaikki tässä selvityksessä tuotetun tietomallin mukaiset yhteiset ydinelementit sekä mallin sektorikohtaiset tietoelementit. Tällöin kaikki jaettavaksi tarkoitetut tiedot olisivat Nimitietopalvelussa kaikkien käytettävissä.

Tässäkin ratkaisussa tietueet tulee voida tallentaa toimijatietokantaan joko eräajoina tai suoraan. Tietokannan tietueiden tulee olla yhdistettävissä, erotettavissa ja muokattavissa. Nimitietopalvelun oman formaatin ja sovelluksen ylläpidon vaativuus riippuisi siitä, miten usein sektorikohtaiset formaatit ja kuvailukäytänteet päivittyvät tai vaihtuvat. Nimitietopalveluun voisi myös tulla mukaan uusia tahoja kuten tutkimusdataa tallentavia organisaatioita, jotka soveltavat muita kuin KAM-organisaatioiden kuvailuformaatteja. Näiden formaattien uusien tietoelementtien mahdollinen lisääminen Nimitietopalveluun tulisi toteuttaa kyseisten formaattien asiantuntijoiden voimin; mutta tämä edellyttää sitä että heille annetaan ensin Nimitietopalvelun formaatin koulutus.

Nimitietopalvelun toteutus edellyttäisi erikseen suunniteltavien ja toteutettavien formaattikonversioiden lisäksi uuden, pelkästään Nimitietopalvelua varten rakennettavan sovelluksen sekä toimijatietokannassa tapahtuvaan toimijatietojen tallennukseen, muokkaukseen ja tietueiden yhdistämiseen tarvittavan luettelointisovelluksen rakentamista ja ylläpitoa. Koska toimijatietokanta sisältäisi kaikki ne tietoelementit, joita eri sektoreilla on ylipäätään mahdollista tallentaa, kuvailussa voisi periaatteessa käyttää myös sektoreiden olemassa olevien sovellusten omia luetteloinnin asiakasohjelmistoja, jos ne saataisiin toimimaan yhdessä Nimitietopalvelun palvelinsovelluksen kanssa esim. standardirajapinnan kautta. On kuitenkin todennäköistä, että rajapinta ja valmisohjelmat eivät tukisi kaikkea tarvittavaa toiminnallisuutta.

Toimijatietokannan ylläpitäjät pitäisi kouluttaa oman formaatin käyttäjiksi, jos toimijat kuvaillaan Toimijatietokantaan. Siinäkin tapauksessa, että tiedot tallennetaan tuotantojärjestelmään, kuvailijoiden pitää tietää miten tiedot konvertoituvat paikallisesta yhteiseen formaattiin ja päinvastoin. Eri sektorien tuotantojärjestelmien ja toimijatietokannan välillä on tehtävä tarvittavat formaattikonversiot toimijatietokannan formaattiin ja takaisin.

Finto-jakelun edellyttämä konversio Nimitietopalvelun omasta formaatista Finton laajennettuun SKOS-formaattiin pitäisi toteuttaa ja ylläpitää. Toimijatietokanta voisi toimia ISNI-tunnusjärjestelmän suomalaisen Registration Agency:n tietokantana, mutta ISNI-konversio jouduttaisiin rakentamaan erikseen, kuten myös EU Core Vocabulary –konversio. Kaikki näiden konversioiden kannalta oleelliset tietoelementit sisältyisivät nimitietopalvelun formaattiin.

8. JATKOSELVITYSTEN JA –TOIMENPITEIDEN TARVE

Tässä selvityksessä on tarkasteltu Nimitietopalvelun toteuttamisen edellytyksenä olevia kulttuuriperintösektoreiden ja julkisen hallinnon organisaatioiden tietoarkkitehtuurikysymyksiä ja esitetty tietoarkkitehtuuriryhmän näkemyksiä niiden ratkaisemiseksi. Tässä luvussa esitetään yhteenvetona jatko selvityksiä ja -toimenpiteitä edellyttävät kysymykset.

8.1. Nimitietopalvelun arkkitehtuuri

MARC 21:N JA EAC-CPF:N SOVELTUVUUS NIMITIETOPALVELUN FORMAATIKSI

Tarvitaan selvitys siitä, onko Nimitietopalvelun tietomalli mahdollista toteuttaa MARC 21- tai EAC-CPF -formaateilla. Tämä selvitys on keskeinen edellytys Nimitietopalvelun toteutussuunnittelun käynnistämiseksi. Lisäksi tarvitaan selvitys siitä, miten yhteismitallinen kumpikin formaatti on kansallisten luettelointikäytäntöjen kanssa eri sektoreilla ja organisaatioissa. Jos toteutus ei ole mahdollista tukeutuen jompaankumpaan näistä formaateista, on tietoarkkitehtuuriryhmän visioiman Nimitietopalvelun toteutusta varten tuotettava oma formaatti.

VALMIIDEN KONVERSIO-OHJELMIEN SAATAVUUS

Standardien keskinäinen semanttinen yhteismitallisuus on varsin hyvä, mikä luo edellytykset konversioiden teolle. On olemassa ainakin yksi palvelu joka mahdollistaa konversion MARC 21 -formaattista EAC-CPF:ään (<http://socialarchive.iath.virginia.edu/dev/>).

Tulisi selvittää olemassa olevan konversiotyökalun laatu sekä kartoittaa, onko olemassa valmisohjelmistoja EAC-CPF - MARC 21 -konversion tuottamiseen.

SKOS-FORMAATIN TARKASTELU TOIMIJATIETOKANNAN NÄKÖKULMASTA

Toimijatietokannan tietomalli tulee mapata Finton käyttämään laajennettuun SKOS-formaattiin ja mikäli Finton formaatti vaatii lisälaajennusta tietomallin takia, tulee tätä tarkastella yhteistyössä Finton asiantuntijoiden kanssa.

KIRJASTOJEN TOIMIJAN JULKISEEN IDENTITEETTIIN PERUSTUVAN SEKÄ ARKISTOJEN JA MUSEOIDEN TOIMIJA-ENTITEETTIIN PERUSTUVAN KUVAILUN YHDISTÄMISEN EDELLYTYKSET NIMITIETOPALVELUSSA

Toimijat kuvaillaan kirjastoissa eri tavalla kuin museoissa ja arkistoissa. Kirjastojen kuvailussa kuvailun kohteena ovat toimijoiden julkiset identiteetit, joita voi olla yhdellä luonnollisella henkilöllä tai oikeushenkilöllä useita. Museot ja arkistot puolestaan kuvailevat toimijan yhtenä entiteettinä, joka sisältää kaikki toimijan identiteetit. Näiden erityyppisten kuvailujen yhdistäminen automaattisesti toimijatietokannassa edellyttää myös erojen tarkempaa analyysia ja mallintamista.

8.2. Muut selvitykset

KANSALLISEN ISNI-INFRASTRUKTUURIN KEHITTÄMINEN

ISNI-tunnuksien käytön tehostamiseksi Suomeen tulisi perustaa joko ISNI-tunnusjärjestelmän Registration Agency tai Federated Registration Agency. Vaihtoehtojen edut ja mahdolliset haitat ja niiden vaikutus käytännön toimintaan Suomen oloissa tulee analysoida, ja tehdä selvityksen pohjalta esitys jatkotoimiksi.

ULKOISTEN TIETOVARANTOJEN HYÖDYNTÄMINEN NIMITIETOPALVELUSSA

Nimitietopalvelu voisi hyödyntää Väestötietojärjestelmää sekä PRH:n yhteisörekistereitä samoin kuin yksittäisten kulttuuriperintöorganisaatioiden keräämiä toimijatietoja. Tämä edellyttää kuitenkin paitsi juridista selvitystyötä myös näiden tietovarantojen sisältämien toimijatietojen yhteismitallisuuden tarkempaa määrittelyä yhteistyössä asianomaisten organisaatioiden kanssa.

Tietojen vaihdon toteuttaminen edellyttää yhteisiä toimijatietojen vaihtoformaatteja (kuten EU Core Person) samoin kuin automatisoitujen eräpäivitysten edellyttämiä teknisiä määrittelyjä ja yhteyksiä.

LAINSÄÄDÄNNÖLLISTEN ESTEIDEN POISTAMINEN

Henkilötoimijoiden kuvailutietojen kerääminen, jakelu ja julkaiseminen edellyttävät joko tietosuojalautakunnan lupaa tai laintasoista valtuutusta. Laintasoinen valtuutus edellyttää joko Nimitietopalvelun sisällyttämistä osaksi jotakin muuta lainsäädäntöhanketta tai erillisen lainsäädäntöhankkeen perustamista. Valtiovarainministeriöllä ja OKM:llä on keskeinen rooli tietopalvelun lainsäädännöllisten esteiden poistamisessa.

9. LÄHTEET

Arkistolain jatkotyöryhmän muistio – Laki tiedonhallinnasta ja tietojen käsittelystä. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2015:12. http://minedu.fi/OPM/Julkaisut/2015/arkistolain_jatkotyoryhma.html?lang=fi

Isabelle Chave & Claire Sibille-de Grimoüard (2014): Towards the Development of a National Archival Authority File in France: An Approach to Implement EAC-CPF, *Journal of Archival Organization*, 12:1-2, 98-117, DOI: [10.1080/15332748.2015.1000207](https://doi.org/10.1080/15332748.2015.1000207). Elektroninen julkaisu, käytettävissä osoitteessa <http://dx.doi.org/10.1080/15332748.2015.1000207>

Voutilainen, Tomi: Palveluarkkitehtuuria tukeva lainsäädäntö. Helsinki: Valtiovarainministeriö, 2014. (VM:n julkaisuja 22/2014). Elektroninen julkaisu, käytettävissä osoitteessa <http://vm.fi/dms-portlet/document/354409>

versio	pvm	tekijä	muutoksen kuvaus
0.1	30.1.2015	Jaana Kilkki	Muistio käsittelyyn Anne Kauhanen-Simanaisen kanssa
	10.2.2015	Jaana Kilkki	Täydennystä tietoarkkitehtuuriryhmän kokouksen 9.1.2015 jälkeen
0.2		Marja-Liisa Seppälä Juha Hakala	RDA:n kuvausta täydennetty Lisätty tunnistejärjestelmien kuvaukset.
0.25	2015-03-23	Juha Hakala	Lisätty luvun 6 alku.
	2015-05-26	Leena Furu-Kallio	Lisätty Core Vocabulary tekstiä lukuun 2.
	2015-06-29	Leena Furu-Kallio	Lisätty kommentteja ja muokattu tekstiä kauttaaltaan, lisätty tekstiä lukuun 1 ja 5
	2015-07-07	Jaana Kilkki	sisällysluettelo; lukujen rakenteen muokkaus; tekstin täydennystä
	2015-07-13	Jaana Kilkki	Palaverissa 10.7. sovittu rakenteen muokkaus. Lukujen 1. ja 6. muokkausta.
0.5	2015-07-15	Juha Hakala	Palaverissa sovittu luvun 2. muokkaus. Kuvat puuttuvat. Arkistolain jatkotyöryhmän muistiossa mainitut rekisterit siirretty luvusta 5 lukuun 2, lisätty lukuun 5 muita kansallisia rekistereitä ja muutettu luvun otsikko.
0.51	2015-07-21	Juha Hakala	Kaikki tehdyt muutokset hyväksytyt, lisätty Jaana Kilkin muutosehdotukset.
	2015-07-22	Leena Furu-Kallio	Lisätty museoita koskevaa tekstiä lukuun 2.1.1.
	2015-07-23	Leena Furu-Kallio	Muutettu dokumentin järjestys 23.7. palaverissa sovitun mukaiseksi
	2015-07-24	Marja-Liisa Seppälä	Lisäkommentit rakenteesta
	2015-05-27	Marja-Liisa Seppälä	Tietomalliluvun päivitys
0.6	2015-08-11	Markus Merenmies	KDK:n dokumenttipohja
0.7	2015-08-27	Juha Hakala	Muutoksia lukuihin 2 ja 4
0.76	2015-09-02	Juha Hakala	Luvun 4 kuvat päivitetty ja tekstiä muokattu. Lukua 8 muokattu.
0.8	2015-09-07	Jaana Kilkki	Lisätty luku 2. Stilisoitu tekstiä kauttaaltaan. KDK-ohjausryhmälle ja JH-TA-ryhmälle esiteltävä versio.
0.83	2015-09-11	Juha Hakala	Lisätty uusi vaihtoehto optioksi kolme, vanha kolmosvaihtoehto muutettu vaihtoehdoksi 4.
	2015-11-23	Juha Hakala	Tehty 19.11. pidetyssä kokouksessa sovittuja muutoksia sekä muita pienempiä viilauksia.
	2015-11-24	Markus Merenmies	Luvun 3.2 kuvat yhdistetty.

0.9	2015-12-07 2015-12-21 2015-12-23	Markus Merenmies/Leena Furu/Juha Hakala/Seija Saari	Merkittäviä muutoksia dokumentin lukuihin 3 ja 4. Luvun 1 yhteyteen on lisätty sanasto. Lisäyksiä ja tekstikorjauksia myös muissa luvuissa.
0.93	2016-01-05 2016-01-07 2016-01-08	Seija Saari Jaana Kilkki Jaana Kilkki	Poistettu värejä sisältävä vaihtoehtojen vertailutaulukko; se on erillisenä tiedostona saatavilla. Pieniä täydennyksiä ja muutoksia. Merkitty punaisella. Luvun 3 täydentäminen ja muokkaaminen. Luvun 9 uudelleenkirjoituksen aloittaminen.
0.94	2016-01-11	Markus Merenmies/Seija Saari	Tehtyjen lisäysten ja muutosten (7. ja 8.1.2016) hyväksyminen. 1. luvun yhden kappaleen tekstin muotoileminen. Luvun 2 (käsitteet) siirtäminen erilliseksi liitteeksi "sanasto". Muiden lukujen numerointi muuttui yhdellä alaspäin.
0.96	2016-01-19	Ahlfors / Furu / Herrala / Seppälä	Tiivistetty rajusti koko dokumenttia. Siirretty painopistettä tavoitetilan kuvaamisen, jotta tekninen toteutus ei korostuisi liikaa. Poistettu toteutusvaihtoehdot 1 ja 2, koska ne eivät johda kohti tavoitetilaa. Muutettu nimi toimijatietopalveluksi. Yhtenäistetty käytettyjä termejä sekä selvennetty niitä sanastossa.
0.97	2016-02-2	Jaana Kilkki	Päivitetty kokouksessa 21.1.2016 sovitulla tavalla
0.98	2016-03-01	Jaana Kilkki	Kommenttikierroksen jälkeiset päivitykset
0.98	2016-03-03	Juha Hakala	Stilistisiä korjauksia. Korjattu ISNI-osuutta (SRU-rajapinta)
0.99	2016-03-07	Markus Merenmies	Julkaisukuntoon viimeistely ja kuvien uudistus
0.995	2016-03-16	Juha Hakala	Muutoksia tulee ISNI-järjestelmän kuvaukseen
1.0	2016-03-18	Markus Merenmies	Julkaisuversion paketointi

LIITE 2: SELVITYKSEN SANASTO

Nimitietopalveluselvityksessä käytettyjen keskeisimpien käsitteiden, lyhenteiden ja sanojen merkitykset kuvataan ohessa.

auktoriteettitietokanta

Järjestelmä, joka sisältää auktorisoituja (kontrolloituja) nimimuotoja, jotka toisaalta sitovat yhteen erilaiset nimimuodot (esim. erikieliset tai eri kirjoitusasut) ja toisaalta auttavat erottelemaan saman nimiset henkilöt toisistaan tai saman käsitteen eri merkitykset (esim. johtaminen yrityksessä, musiikissa ja sähkötekniikassa tai Saarijärvi kaupunkina tai sukunimenä).¹⁹

entiteetti

Yleisnimitys sille, minkä kielellinen ilmaus voi hahmottaa (Tieteen termipankki)²⁰.

Kohde; kuvailun kohde; kuvailukokonaisuus (RDA)²¹.

Esimerkiksi kuvailun käsitelmälleissä kuvailun kohteet esitetään entiteetteinä. RDA-kuvailusäännöissä aineistoon sisältyvät entiteetit ovat

Aineistoentiteetit: teos, ekspressio, manifestaatio ja kappale (entiteettiryhmä 1).

Toimijaentiteetit: henkilö, yhteisö ja suku (entiteettiryhmä 2). Henkilö viittaa yksilöön tai yksilön (yksin tai yhdessä yhden tai useamman muun yksilön kanssa) luomaan identiteettiin. Yhteisö tarkoittaa organisaatiota tai ryhmää henkilöitä ja/tai organisaatioita, jotka tunnetaan tietyllä nimellä ja jotka toimivat tai voivat toimia yhtenä kokonaisuutena. Suku tarkoittaa kahta tai useampaa henkilöä, joita yhdistää sukulaisuussuhde, avioliitto, adoptio, rekisteröity parisuhde tai vastaava oikeudellinen asema, tai jotka muuten esiintyvät perheenä

Aihe-entiteetit: käsite, paikka, tapahtuma ja objekti (entiteettiryhmä 3).²²

EU Core Vocabulary²³

EU:n Core Vocabulary sisältää metadatat määrityksiä. Se koostuu tätä kirjoittaessa neljästä eri sanastosta: Core Person Vocabulary, Core Registered Organisation Vocabulary, Core Location Vocabulary ja Core Public Service Vocabulary.

”Person”-sanasto käsittää esimerkiksi henkilön nimen, sukupuolen, syntymäpäivän ja syntymäpaikan. ”Registered Organisation”-sanasto käsittää esimerkiksi kansallisesti rekisteröidyn organisaation (yri-tyt, yhdistys tms., ”legal entity”) tunnusteen ja tiedon siitä, millä alalla organisaatio toimii. ”Location”-

¹⁹ Käsitteen määritelmä on kopioitu KDK-kokonaisarkkitehtuurin (versio 3.0, 26.10.2015) sanastosta.

²⁰ <http://tieteentermipankki.fi/wiki/Kielitiede:entiteetti>

²¹ RDA-tiedotussivut: [<https://wiki.helsinki.fi/display/RDA/RDA-tiedotussivut>]; RDA-tiedotussivut/Miksi RDA?: [<https://wiki.helsinki.fi/pages/viewpage.action?pagelD=136518480>].

²² RDA-tiedotussivut: [<https://wiki.helsinki.fi/display/RDA/Terminologiaa>]. Lisätietoja myös metatietosanas-
tosta: [<https://wiki.helsinki.fi/display/Metatietosanas/Metatietosanasasto>].

²³ A core vocabulary is a core concept, represented in a particular representation language, such as an RDF or XML schema. Ks. <https://joinup.ec.europa.eu/category/glossary/core-vocabulary>

sanasto käsittää esimerkiksi osoitteen ja tiedon alueellisesta sijainnista. "Public service"-sanasto käsittää esimerkiksi tiedot julkisen sektorin tarjoamista palveluista.²⁴

Finto

Suomalainen sanasto- ja ontologiapalvelu, joka mahdollistaa sanastojen julkaisun ja selailun. Palvelu tarjoaa myös rajapinnat sanastojen ja ontologioiden hyödyntämiseen muissa ohjelmistoissa.²⁵

formaatti (1)

Tiedostoformaatti (-muoto) ilmaisee tiedoston rakenteen. Monista tiedostomuodoista on olemassa julkinen määritelmä (standardi), jossa kuvataan miten kyseisessä formaatissa oleva informaatio on tallennettava digitaaliseen muotoon. Esimerkiksi Portable Document Format (pdf), Tagged Image File Format (tiff).

formaatti (2)

Metadataformaatti ilmaisee metadatan rakenteen. Monista metadataformaateista on olemassa julkinen määritelmä (standardi), jossa kuvataan miten kyseisessä formaatissa oleva metadata on tallennettava digitaaliseen muotoon. Esimerkiksi MARC 21, Dublin Core.

Metadataformaatille voidaan määritellä vaihtomuoto, joka kuvaa miten metadatatietueet voidaan tallentaa peräkkäistiedostoksi. Esimerkiksi MARC 21 –formaatin ISO 2709 –standardissa määritelty vaihtomuoto.

formaattikonversio

Metadatatietueen tai tiedoston muuntaminen toiseen tiedostoformaattiin. Muunnokset voivat olla häviöllisiä (tietosisältö tai ulkoasu muuttuu) tai häviöttömiä. Formaattikonversiot ovat migraatioita.

formaattimuunnin

Yleisesti sovellus, jonka avulla metatieto on muunnettavissa formaatista toiseen; esimerkiksi Usemarcon-ohjelmisto joka mahdollistaa mm. muunnoksen FINMARCista MARC21:een. .

ISNI

International Standard Name Identifier (ISO 27729).

ISNI on ISO:n standardoima pysyvä tunnus, jota käytetään tunnistamaan (identifioimaan) toimijoiden julkisia identiteettejä, riippumatta siitä ovatko nämä toimijat henkilöitä (esim. tutkijat, kuvataiteilijat, kirjailijat) tai yhteisöjä (esim. julkishallinnon organisaatiot, yritykset).

julkinen identiteetti

Nimi, jolla toimija tai fiktiivinen hahmo tunnetaan tai tunnettiin julkisesti (ISO 27729)

²⁴ Core Vocabularies-esite: http://ec.europa.eu/isa/actions/documents/eudig12a-1401-i01-core-vocabularies-lr_en.pdf

²⁵ Finto: <http://finto.fi/fi/>.

kontrolloitu nimimuoto (nimiauktoriteetti)

Ks. auktoriteettitietokanta

Legal Entity

Ks. oikeushenkilö

metatieto

Aineiston kontekstia, sisältöä ja rakennetta sekä niiden hallintaa ja käsittelyä koko elinkaaren ajan kuvaava tieto. Metatietoa voidaan käyttää muun muassa aineiston hakuun, paikallistamiseen, pitkäaikaissäilyttämiseen ja tunnistamiseen.²⁶ Metatietoa on kolmenlaista: kuvailevaa, hallinnollista ja rakenteista.

migraatio

Tapahtuma, jossa pitkäaikaissäilytyksessä olevan säilytyspaketin tietosisältöä tai säilytystietoa muutetaan. Esimerkiksi säilytettävän tekstidokumentin ASCII-koodien muuttaminen UNICODEksi.

nimi

Merkkijono, jolla toimijaan viitataan tai on viitattu yleisesti.

oikeushenkilö

Oikeushenkilö eli *juridinen henkilö* on henkilöiden tai pääomien yhteenliittymä, jolla on omat oikeutensa ja velvollisuutensa. Oikeushenkilöitä tarvitaan taloudellisen, aatteellisen tai yhteiskunnallisen toiminnan järjestämistä varten. Vain lainsäädännön tuntemat oikeushenkilöt voivat saada oikeussubjektin aseman.²⁷

ontologia

Formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä²⁸.

RDA–Resource Description and Access

Resource Description and Access (RDA) pohjautuu anglo-amerikkalaisiin luettelointisääntöihin (Anglo-American Cataloguing Rules, AACR2). RDA:n viitekehyksenä ovat [Functional Requirements for Bibliographic Records](#) (FRBR) ja sen johdannaiset [Functional Requirements for Authority Records](#) (FRAD) ja [Functional Requirements for Subject Authority Data](#) (FRSAD). Nämä käsitteelliset määrittelevät tiedonhaun kohteet, entiteetit ja näitä kuvailevat tiedot eli attribuutit ja entiteettien keskinäiset suhteet. KDK-kokonaisarkkitehtuurissa RDA on määritelty toimijan kuvailun yhteisiksi säännöiksi.

semanttinen yhteismitallisuus

²⁶ Käsitteen määritelmä on kopioitu KDK-kokonaisarkkitehtuurin (versio 3.0, 26.10.2015) sanastosta.

²⁷ <https://fi.wikipedia.org/wiki/Oikeussubjekti>

²⁸ Käsitteen määritelmä on kopioitu KDK-kokonaisarkkitehtuurin (versio 3.0, 26.10.2015) sanastosta.

Merkityksien yhteismitallisuus. Kuvailussa semanttinen yhteismitallisuus tarkoittaa metatietojen merkitysten yhdenmukaisuutta.

Nimitietopalvelussa semanttinen yhteismitallisuus tarkoittaa niin RDA-kuvailusääntöihin perustuvien kuvailutietojen yhdenmukaisuutta kuin nimien merkitysten yhteismitallisuutta. Semanttinen yhteismitallisuus toteutuu yhteisten kuvailusääntöjen (esim. RDA) ja yhteisten standardien (esim. ISNI) ja yhteisten sanastojen (esim. EU Core Vocabulary) kautta.

tietoelementti

Metadatatassa tietoelementti on datan pienin, jakamaton tiedon yksikkö jolla on sekä täsmällinen merkitys että määritelmä (ks. ydinelementti)²⁹.

tietomalli

Tiedon rakenteen määrittelevä datan organisoinnin kuvaus³⁰.

tietovaranto

Tietovaranto on tietokokonaisuus tai erilaisten tietokokonaisuuksien koontipiste. Tietovaranto voi joko olla kaikkien yhteisesti ja vapaasti käytettävissä tai sen käyttö voi olla rajoitettua. Yhteisesti käytettäviä vapaita tietovarantoja ovat esimerkiksi kaikille avoimet yliopistokurssit verkossa (Massive Online Open Courses, MOOC) ja pian Suomessa käyttöön tulevan Kansallisen palveluväylän kautta saatutavat palvelut ja tietovarannot. Rajoitettuja tietovarantoja ovat esimerkiksi erikoisalojen tietokannat, joihin pääsy edellyttää lisenssin ostamista tai vuosimaksua.

toimija

Luonnollinen henkilö, oikeushenkilö tai luonnollisten henkilöiden tai oikeushenkilöiden muodostama ryhmä (ISO 27729)

toimijaentiteetti

Ks. entiteetti

toimijatietokanta

Toimijatietovarannon tekninen toteutus, joka perustuu selvityksen yhteydessä työstettyyn toimijatietokannan tietomalliin

toimijatietovaranto

Toimijoiden/julkisten identiteettien metatietojen muodostama kokonaisuus

tunnus

²⁹ In [metadata](#), the term data element is an atomic unit of data that has precise meaning or precise semantics (Wikipedia)

³⁰ description of the organization of *data* in a manner that reflects an *information* structure (ISO 5127)

Merkkijono tai linkki joka identifioi tunnistettavan objektin, henkilön tai organisaation joko yksinään tai yhdessä muiden tietoelementtien kanssa.³¹

tunnusjärjestelmä

Standardiin tai muuhun yleisesti tunnettuun dokumenttiin perustuva järjestelmä, jonka avulla voidaan antaa pysyviä tunnuksia yleensä jollakin tavoin rajatulle joukolle objekteja, henkilöitä tai organisaatioita. Nimitietopalvelun kannalta merkittäviä tunnusjärjestelmiä ovat henkilön tai organisaation julkisen identiteetin ISNI-tunnus, yksittäisen tutkijan ORCID tai opiskelijan oppijanumero, joiden avulla identifioituun kohteeseen on mahdollista viitata yksilöidysti.

ydinelementti

RDA-kuvailusääntöjen mukaisen hyvän kuvailutason edellyttämät tietoelementit (ks. tietoelementti).

³¹ *data string* or pointer that establishes the identity of an *item*, institution or person alone or in combination with other *elements* (ISO 5127)