

# MUSEOALAN ONTOLOGIAT SUOMESSA

## KANSALLISEN FINNONTO-HANKKEEN TULOSTEN HYÖDYNTÄMINEN

*"Intellectuals solve problems, geniuses prevent them"* (Albert Einstein)

25.4.2012

### **Eero Hyvönen**

Semanttisen laskennan tutkimusryhmä (SeC0)

Mediatekniikan laitos, Aalto-yliopisto ja

Tietojenkäsittelytieteen laitos, Helsingin yliopisto

<http://www.seco.tkk.fi/>

[eero.hyvonen@aalto.fi](mailto:eero.hyvonen@aalto.fi)

## **1 JOHDANTO**

Suomalaiset semanttisen webin ontologiat -tutkimushankesarja FinnONTO<sup>1</sup> alkoi v. 2003 ja päättyi syksyllä 2012. Työn tavoitteena on ollut kehittää perusta kansalliselle semanttisen webin ontologiainfrastruktuurille ja -palveluille sekä arvioida ja demonstroida niiden hyötykäyttöä pilottisovelluksissa. Tutkimustyö on tehty pääosin Aalto-yliopiston ja Helsingin yliopiston Semanttisen laskennan tutkimusryhmässä. Mukana hankekonsortiossa on ollut hankkeen eri vaiheissa rahoittajana puolensataa julkista organisaatiota ja yritystä Tekesin ollessa hankkeen päärahoittaja.

Mukana työssä on ollut alusta saakka lukuisia muistiorganisaatioita: museoita (Museovirasto, Valtion taidemuseo, Antikvaria-ryhmän museot, Espoon kaupunginmuseo, Suomen valokuvataiteen museo, Luonnontieteellinen keskusmuseo, Suomen Maatalousmuseo Sarka ym.), kirjastoja (Kansalliskirjasto, Helsingin kaupunginkirjasto, yleisten kirjastojen Kirjastot.fi, Viikin tiedekirjasto, Helsingin yliopiston kirjasto, Taideteollisen korkeakoulun kirjasto) sekä arkistoja (SKS, SLS, Kansallisarkisto ja Kotus). Kulttuurialan ontologiat ja sovellukset ovatkin olleet yksi hankkeen fokusalueista semminkin, kun tutkimusryhmällä on ollut menossa samaan sektoriin liittyvää muutakin tutkimusta mm. Suomen Kulttuurirahaston ja EU:n tuella.

Työn tuloksena on syntynyt tieteellisten artikkeleiden, väitöskirjojen ja muiden opinnäytetöiden ohella hyödynnettävissä olevia tuloksia, jotka ovat avoimesti ja maksutta saatavilla ja käytettävissä. Näitä ovat kymmenet eri ontologiat ja sanastot, kansallinen ontologiapalvelu ONKI<sup>2</sup> ja sovellusprototyypit, joita kulttuurialalla ovat MuseoSuomi<sup>3</sup> (verkossa vuodesta 2004), Kulttuurisampo<sup>4</sup> (vuodesta 2008), Kirjasampo<sup>5</sup> (vuodesta 2011) ja Matkailusampo<sup>6</sup> (tulossa 2012). Kotimaisten tunnustusten ohella hanke on saanut mm. kahdesti kansainvälisen semanttisen webin tutkijayhteisön Semantic Web

---

<sup>1</sup> <http://www.seco.tkk.fi/projects/finnonto>

<sup>2</sup> <http://onki.fi>

<sup>3</sup> <http://www.museosuomifi>

<sup>4</sup> <http://www.kulttuurisampo.fi>

<sup>5</sup> <http://www.kirjasampo.fi>

<sup>6</sup> <http://www.matkailusampo.fi>

Challenge -palkinnon kahdesti (2004 ja 2008) ja vuonna YK:n Word Summit Award (WSA) -palkinnon (Kulttuurisammon teknologiaa hyödyntävä SmartMuseum<sup>7</sup>) (2010).

Tässä artikkelissa tiivistän tämän tutkimustyön pohjalta käsityksiäni siitä, miten semanttisen webin<sup>8</sup> ja avoimen yhdistetyn tiedon (Linked Open Data) teknologiaa (Heath, Bizer, 2011) ja FinnONTO-hankkeen tuloksia kannattaisi ottaa käyttöön Suomessa ajatellen erityisesti museoita ja muistiorganisaatioita.

Seuraavassa esitetään ensin yhteenveto FinnONTO-hankkeen puitteissa kehitetyistä keskeisimmistä ontologioista. FinnONTO-vision mukaan **ontologiat muodostavat metatietomallien ohella museoalan luettelointi- ja julkaisujärjestelmien sisällöllisen selkärangan**, ”sisällöllisen liiman” (Hyvönen, 2008a, 2008b). Metatietomallit kertovat, mitä tietoja sisältökohteisiin liittyen museojärjestelmään tallennetaan, kuten esimerkiksi esineen tyyppi, valmistaja ja materiaali, tai valokuvan aihe ja paikka. Metatietokenttien arvot ovat joko 1) vapaata kuvailevaa tekstiä, 2) määrämuotoista tietoa kuten aikamääreitä, tai 3) arvot valitaan etukäteen sovitusta sanastoista, esimerkiksi Museoalan ontologiasta MAO tai paikkatietorekisteristä.

Sanastoja ja nimistöjä on kehitetty aiemminkin eri formaateissa. FinnONTO-hankkeen ja yleisemmin semanttisen webin ja yhdistetyn tiedon kehittäjien innovaationa on kuitenkin ollut siirtyä sanastojen käytöstä ”ontologoiden” käyttöön (Hyvönen, 2005). Ontologiat eroavat perinteisistä sanastoista siitä, että käsitteiden merkityksiä on täsmennetty koneellista tulkintaa (”semantiikka”) varten, mikä mahdollistaa niiden hyödyntämisen aiempaa paremmin sekä kokoelmien luettelointi- että julkaisujärjestelmissä. Sanaston ”ontologisointi” merkitsee sitä, että sen käsitteiden muodostamalle viittausten verkostolle on luotu systemaattinen semanttinen rakenne (erityisesti luokkahierarkia), käsitteet on yksilöity WWW-maailmassa käytettävillä ns. URI-tunnisteilla ja ontologiat esitetään W3C:n<sup>9</sup> semanttisen webin standardien mukaisessa harmonisoidussa muodossa (RDF(S), SKOS, OWL), jolle on määritelty logiikkaan perustuva semantiikka ja jollaisen tiedon käsittelyyn on saatavilla runsaasti valmiita open source -työkaluja ja ohjelmistoja. Esimerkiksi FinnONTO:ssa on käytetty ontologioiden toimittamiseen Stanfordin yliopiston ilmaista avoimen koodin Protege-editoria<sup>10</sup>, jolla on maailmanlaajuisesti n. 177 000 rekisteröitynyttä käyttäjää. Ontologiat mahdollistavat perinteisiä sanastoja paremmin mm. täsmällisemmän sisällönkuvailun, monikielisyyden, sisältöjen semanttisen rikastamisen loogisen päättelyn avulla, semanttisen haun ja suosittelun sekä visualisoinnin. Europeanan<sup>11</sup> seuraavissa versioissa FinnONTO:ssa jo pilotoitua ideaa ja semanttisen webin teknologiaa (MuseoSuomi 2004 ja Kulttuurisampo 2008) ollaan tietävästi ottamassa laajamittaisesti käyttöön Euroopan mittakaavassa, ja 3,5 miljoonan objektin koe-erä aineistoa on jo julkaistu avoimena yhdistettynä tietona (Linked Open Data). Aihepiiri on juuri nyt erityisen ajankohtainen Suomessa ajatellen vasta käynnistynyttä kansallista Museo 2015 -hanketta<sup>12</sup> (2011-2015), jonka päätavoitteita ovat ”museoiden kokoelmahallinnan prosessien yhtenäistäminen, museoiden kokoelmahallinnan kokonaisarkkitehtuurin ja kokonaisarkkitehtuurin hallintamallin laatiminen sekä edellytysten

<sup>7</sup> <http://www.smartmuseum.eu/>

<sup>8</sup> <http://www.w3.org/standards/semanticweb/>

<sup>9</sup> <http://www.w3.org/>

<sup>10</sup> <http://protege.stanford.edu/>

<sup>11</sup> <http://www.europeana.eu/>

<sup>12</sup> [http://www.nba.fi/fi/museoalan\\_kehittaminen/museo\\_2015](http://www.nba.fi/fi/museoalan_kehittaminen/museo_2015)

luominen yhteisen kokoelmahallintajärjestelmän hankintaan ja käyttöönottoon”. Museoalan sanastojen uudistamistarvetta on arvioitu Museoviraston selvityksessä (Lång, Valanto, 2008), kokoelmanhallintajärjestelmistä on laadittu CSC:n toimesta selvitys (CSC, 2011) ja asiakokonaisuus liittyy Kansallinen digitaalinen kirjasto -hankkeen kokonaisarkkitehtuurisuunnitelmaan (KDK, 2011).

Seuraavassa esitellään ensin jaottelu museoalalla tarvittavista ontologisista sanastoista sekä FinnONTO:ssa saatuja tuloksia tämän jäsentelyn mukaisesti. Tämän jälkeen esitellään ontologioiden käyttöönoton kannalta keskeinen järjestelmä, Kansallinen ontologiakirjastopalvelu ONKI. Tehdyn työn perusteella esitetään lopuksi konkreettisia ehdotuksia siitä, millainen museoalan ontologiainfrastruktuuri maahamme oman käsitykseni mukaan tarvittaisiin ja miten työtä kannattaisi jatkaa FinnONTO:n perinnön pohjalta.

## 2 MUSEOALAN KÄSITEONTOLOGIAT – FINNONTO:N PERINTÖ

Termit ”ontologia” ja ”sanasto” ovat merkitykseltään moniselitteisiä. Tietotekniikan piirissä niillä voidaan tarkoittaa semanttisesti esitettyä metatietomallia, kuten Dublin Core<sup>13</sup> (DC), CIDOC CRM<sup>14</sup> tai Europeana Data Model (EDM)<sup>15</sup>, tai tesauruksen kaltaista käsitejärjestelmää tai tietorekisteriä, kuten Getty säätiön Art and Architecture Thesaurus AAT, Universal List of Artis Names (ULAN) tai Thesaurus for Geographical Names (TGN)<sup>16</sup>. Tässä artikkelissa käsitellään metatietomallien elementtien arvojen täyttämässä käytettäviä tesaurus- ja rekisterityyppisiä ontologioita, joista voitaneen käyttää selvennyksen vuoksi termiä ”käsiteontologia” korostamaan eroa CIDOC CRM:n kaltaisiin ontologisiin metadatatamalleihin. *Käsiteontologiat ovat välttämätön lisäkomponentti DC, CIDOC CRM tai EDM -tyyppisille metadatatamalleilla/ontologioille, joissa ei oteta kantaa esimerkiksi siihen, mitä eri esinetyyppejä, asiasanoja, henkilöitä tai paikkoja sisällönkuvailussa pitäisi yhteentoimivuuden ja älykkäiden toiminnallisuuksien takia käyttää.*

Kansallisessa ontologiainfrastruktuurissa tarvittavat keskeiset käsiteontologiat voidaan ryhmitellä seuraavasti:

- 1) **Yleiskäsiteontologiat.** Nämä ontologiat vastaavat käsitteistöltään karkeasti nykyisiä asiasanastoja, kuten YSA ja MASA, ilman vapaita indeksoinnin termejä. Yleiskäsitteiden määrä on kohtuullinen pieni ja lasketaan yleensä kymmenissä tuhansissa. Museoalan tiettävästi laajimmassa yleiskäsitesanastossa AAT on n. 131 000 termiä, joskin mukana on hyvinkin yksityiskohtaisia käsitteitä, esimerkiksi erilaisia tuolityyppejä.
- 2) **Toimijaontologiat.** Toimijaontologiat ovat henkilö- ja organisaatiorekistereitä, ja muistuttavat kirjastoissa käytettyjä ns. auktoriteettitietokantoja. Toimijaontologian avulla samannimiset toimijat voidaan yksilöidä ja erottaa toisistaan eri tunnisteilla ja lisätietojen avulla, esimerkiksi henkilökaimat syntymävuoden ja -paikan avulla. Toimijaontologioissa voi olla erittäin paljon yksilöitä, mutta luokkarakenne on yksinkertainen (esimerkiksi kansallisuudet ja ammatit) ja perustuu yleiskäsiteontologioihin. Esimerkiksi taiteilijoihin keskittyvässä ULAN-sanastossa on

<sup>13</sup> <http://dublincore.org/>

<sup>14</sup> <http://www.cidoc-crm.org/>

<sup>15</sup> <http://pro.europeana.eu/edm-documentation>

<sup>16</sup> <http://www.getty.edu/research/tools/vocabularies/>

n. 120 000 yksilötoimijaa (kuten Akseli Gallen-Kallela), joilla on n. 293,000 erilaista nimeä, ja toimijoita on luokiteltu kansallisuuden ja ammatin mukaan, joiden määrä lasketaan sadoissa.

- 3) **Paikkaontologiat.** Paikkaontologiat vastaavat kansallisten maanmittauslaitosten ylläpitämiä paikkarekistereitä. Niiden avulla voidaan yksilöidä paikat, sijoittaa ne koordinaatistoon ja tallentaa paikkoihin liittyviä lisätietoja kuten paikkatyypit (kylä, järvi, asema jne). Paikkaontologioissa voi olla erittäin paljon yksilöitä (miljoonia), mutta luokkarakenne on toimijaontologioiden tapaan yksinkertainen ja perustuu yleiskäsitteisiin. Esimerkiksi TGN-sanastossa on n. 912 000 historiallista paikkaa ja Geonames-sanastossa<sup>17</sup> yli 8 miljoonaa nykypaikkaa.
- 4) **Aikaontologiat.** Museoalalla keskeistä on ajan esittäminen. Kulttuurisällöissä viitataan lineaarisen kalenteriajan ohella päivän ja vuodenvaihteen aikoihin (kevät, aamu) sekä nimettyihin aika- ja tyylikausiin (esim. rauta-aika, valistuksen aika, art deco). Aikakäsitteitä ja -jaksoja ei ole kovin paljon, puhuttaneen sadoista. Ajan esittäminen on suurempi haaste metatietomalleille. Haasteena on mm. epätäsmällisen (historiallisen) aikatiedon esittäminen ja tulkinta sekä epäsäännöllisten syklisten aikojen esittäminen, kuten esimerkiksi museoiden aukiolo.
- 5) **Tapahtumaontologiat.** Tapahtumat liittävät toisiinsa henkilöitä, paikkoja, aikoja ja toisia tapahtumia, ja mahdollistavat kulttuuristen sekä provenienssitietoon liittyvien ilmiöiden kuvaamisen semanttisesti yhteentoimivalla tavalla. Tästä johtuen mm. CIDOC-CRM-järjestelmän perustaksi on valittu tiedon esittäminen tapahtumina, eikä esimerkiksi Dublin Coren dokumenttiperustaista tietomallia. Jotta kaksi sisältökohdetta voidaan yhdistää toisiinsa yhteisen tapahtuman kautta, pitää olla käytössä keino viitata tapahtumiin yksikäsitteisesti samalla tavalla kuin henkilöihin tai paikkoihin. Tapahtumaontologiaan tallennetaan tähän tarvittava viitetieto historiallisista tapahtumista, kuten ”toinen maailmansota” tai ”Napoleonin kruunajaiset”. CIDOC CRM ei metatietomallina eikä ota kantaa yksittäisiin tapahtumiin, vaan ainoastaan tapahtumien yleiseen rakenteeseen, esimerkiksi että tapahtumiin liittyy aina joku aika ja paikka.
- 6) **Nimistöontologiat.** Monilla tieteenaloilla on käytössä laajoja nimistöjä, joita voidaan liittää yleisontologioihin. Tällaisia ovat esimerkiksi biologian eliöiden lajilistat, geologian mineraalit, lääketieteen taudit ja lääkkeaineet, kielitieteen kielet yms. Koska tällaisia yksityiskohtaisia käsitteitä saattaa olla käytössä suuria määriä ja niiden käyttö fokusoituu erikoisaloille, on tarkoituksenmukaista toteuttaa nimistöontologioita erillisinä ontologioina, jotka on kuitenkin sillattu (ripustettu) yleisempään käsitteistöön. Esimerkiksi FinnONTO:n YSO-ontologian käsite ”linnut” laajenee kattavaksi, yli 10 000 maailman lintulajin taksonomiaksi AVIO-nimistöontologian kautta. Nimistöjen termit ovat usein luonteeltaan samanaikaisesti sekä yleiskäsitteitä tai luokkia ja yksilöitä. Esimerkiksi ”sinijalkasuula” on biologisen taksonomian taksoni, yksilö luokasta ”lajit”, mutta samalla kaikkien yksittäisen sinijalkasuulayksilöiden yleinen luokka.

Museoala kattaa kulttuurihistorian ja taiteet (tangible heritage), luonnonhistorian ja -ympäristön (natural heritage) sekä ei-esineellisen kulttuurin (intangible heritage), kuten perinteet, eli koko ihmiselämän ja luonnon kirjon. Alalla tarvitaan siksi potentiaalisesti laajinta mahdollista kokoelmaa ontologioita.

---

<sup>17</sup> <http://www.geonames.org/>

Seuraavassa luodaan lyhyt yhteenveto FinnONTO:n piirissä kehitetyistä ja julkaistuista tärkeimmistä ontologioista ja meneillään olevasta työstä yllä olevien ontologiatyyppien mukaisesti jäsennettyinä. Kokemuksista ontologioiden käytöstä semanttisissa kulttuuriportaaleissa on raportoitu tarkemmin artikkeleissa kuten (Hyvönen et al., 2005; Mäkelä et al., 2011, 2012).

## 2.1 Yleiskäsiteontologiat

FinnONTO-hankkeen keskeinen tulos on KOKO-yleiskäsiteontologia. KOKO on joukko suomalaisista asiasanastoista tuotettuja ontologioita, jotka on liitetty (sillattu) toisiinsa Linked Data -tyyppiseksi yhtenäiseksi ontologiapilveksi. KOKO:n ontologioiden yhdistäminen perustuu erityisesti Yleisestä suomalaisesta asiasanastosta johdettuun Yleiseen suomalaiseen ontologiaan YSO, joka muodostaa KOKO:n yläontologian (upper ontology). Alla olevassa taulukossa on lueteltu KOKO:on sisältyviä, YSO:n sillattuja ontologioita, joista moni on peräisin muistiorganisaatioista.

	Name	Ontology domain	Underlying thesaurus	Size	Maintaining Organization
1	YSO	General domain	General Finnish Thesaurus, YSA, Allärs	23700	National Library, Åbo Academy
2	MUSO	Music	Thesaurus of Music, MUSA/CILLA	1000	National Library
3	MAO	Museum domain	Thesaurus of Museum Domain, MASA	6800	National Board of Antiquities
4	AFO	Agriculture, forestry	Agriforest Thesaurus	5500	Viikki Science Library
5	TAO	Applied arts	Thesaurus of Applied Arts	2600	University of Eastern Finland and Library of Aalto University
6	VALO	Photography	Thesaurus of Photography Literature, Thesaurus of Photography Technology	1900	Finnish Museum of Photography
7	MERO	Seafaring, shipping	Thesaurus of Seafaring	1400	Finnish Transport Agency
8	KAUNO	Literature subjects	Thesaurus of Literature, Bella	4900	Finnish Public Libraries, Kirjastot.fi
9	JUHO	Public government	Thesaurus of Finnish Government, VNAS	6400	Ministry of Finance
10	TERO	Health promotion	YSA, TESA, MeSH, Stameta	22000	Various organizations
11	KITO	Literature research	Thesaurus of Literature Research	900	Finnish Literature Society
12	KULO	Culture research	Thesaurus for Folk Culture Studies	1600	Finnish Literature Society
13	KTO	Linguistics	Thesaurus of Linguistics	1000	Research Institute for the Languages in
14	PUHO	Defense	Thesaurus of Defence Administration	2000	Finnish Defence Forces
15	POIO	Points of interest	TGN, Geonames, LDG, SUO	4600	Various organizations
	<b>TOTAL</b>			<b>86300</b>	

**Taulukko 2.1.** Suomalaisista asiasanastoista tuotettuja FinnONTO-ontologioita, jotka on sillattu toisiinsa yläontologian YSO kautta KOKO-ontologiapilveksi.

Alla lyhyt yhteenveto museoalan kannalta keskeisistä KOKO:n osaontologioista. KOKO, YSO tai MAO on kytketty ONKI-rajapinnan kautta useisiin maassamme käytettyihin museoiden luettelointijärjestelmiin, kuten E-Kuva, Siiri ja Kauko. Tätä kautta yli 30:n suomalaisen museon luettelointijärjestelmä on kytketty ONKI:iin. (Lisäksi tulevat MUUSA-järjestelmän kautta ONKI:iin kytkeytyneet niin ikään yli 30 taidemuseota).

### 2.1.1 Yleinen suomalainen ontologia: YSO

KOKO-ontologian ylempi käsitteistö YSO sisältää YSA:n käsitteet ilman paikannimiä. YSO:on on otettu lisäksi mukaan ontologisoinnin, esimerkiksi merkitysten erottelun yhteydessä väistämättä syntyviä uusia käsitteitä. Esimerkiksi YSA:n ”lapset” käsite on jakautunut merkitykseltään lapsiin ikäluokkana (hyvin nuori henkilö) ja perhesuhteena (vanhuskin on vanhempiensa lapsi). Pyrkimyksenä on ollut sisällyttää YSA mahdollisimman kattavasti YSO:oon. Uutta käsitteistöä ei ole aktiivisesti lisätty FinnONTO:n taholta. FinnONTO:n puitteissa myös YSA:n ruotsinkielinen versio Allso on ontologioitu (Åbo Akademin ja FinnONTO:n yhteistyönä) ja liitetty YSO:oon. Lisäksi YSA/YSO:n useimmat käsitteet käännettiin englanniksi (Kirjastot.fi:n ja FinnONTO:n yhteistyönä), mikä mahdollistaa mm. ensimmäistä kertaa englanninkieliset haut YSA/YSO:lla indeksoituihin tietokantoihin. Englanninnoksen kautta YSO

ja sitä kautta KOKO voidaan myös sillata kansainvälisiin ontologioihin. Monikielisyys toimii rajoitetusti esimerkiksi KOKO-ontologiaan perustuvassa Kulttuurisammossa, jossa aineistot ovat lähes kokonaan suomenkielisiä, mutta niitä voidaan hakea myös ruotsiksi ja englanniksi.

Kaikki muut KOKO-ontologiat ovat YSO:a tarkentavia ontologioita; esimerkiksi museoalan MAO sisältää runsaasti YSO:sta puuttuvia historiallisia esinetyyppejä ja näiden valmistuksessa käytettyjä materiaaleja.

### **2.1.2 Museoalan ontologia: MAO**

MuseoSuomi-järjestelmän ideana on yhdistää eri museoiden kokoelmantietoja yhteisten ontologioiden (sanastojen) avulla. Koska suomalaisia ontologioita ei v. 2002 vielä ollut, sellainen tehtiin MASA:n pohjalta MuseoSuomea varten. Ontologisoinnin yhteydessä sanastoon liitettiin n. 1000 uutta käsitettä. Tuloksena syntyi Museoalan ontologia MAO, joka on julkaistu ONKI-palvelussa (Kettula, 2009). MAO on sillattu Yleiseen suomalaiseen ontologiaan YSO ja Taideteollisuusalan ontologiaan TAO. KOKO-ontologiapilvessä on myös valokuvausalan ontologia VALO, jonka juuret ovat Suomen valokuvataiteen museon kahdessa asiasanastossa. Taidemuseoiden MUUSA-järjestelmässä käytetty hollantilainen Iconclass-sanasto<sup>18</sup> on sillattu suomenkielisten asiasanojen käännösten kautta YSO: on ja KOKO: on ja on julkaistu ONKI:ssa. MUUSA kytketty ONKI-palvelun rajapintaan ja sitä kautta Suomen keskeisimpien 33 taidemuseon luettelointijärjestelmät on kytketty semanttisen webin ontologioihin. Valtion taidemuseon yli 33 000 taideteosta ovat myös mukana Kulttuurisampo-palvelun semanttisessa verkossa ja tämän rajapinnan kautta saavutettavissa esimerkiksi yleisten Kirjastot.fi:n Kirjasampo-palvelussa.

### **2.1.3 Taideteollisuusalan ontologia: TAO**

Taideteollisuusalan tarpeisiin kehitettiin Taideteollisuusalan ontologia TAO pohjana Kuopion yliopiston Taideteollisuusalan asiasanasto. TAO, MAO ja YSO on liitetty toisiinsa kolmikkona; muut KOKO-ontologiat on sillattu vain pareittain YSO:oon.

### **2.1.4 Valokuvauksen ontologia: VALO**

Suomen valokuvataiteen museon kaksi asiasanastoa liitettiin yhteen ja ontologisoitiin Valokuvausalan ontologiaksi VALO.

### **2.1.5 Muut kulttuurialan yleisontologiat**

KOKO-pilveen kuuluu museoalan lisäksi muidenkin muistorganisaatioiden sanastoja, kuten Suomalaisen kirjallisuuden seuran asiasanastoihin perustuvat Kulttuurien tutkimuksen ontologia KULO ja Kirjallisuuden tutkimuksen ontologia KITO.

Useita muitakin kulttuurialan yleisontologioita on kehitetty ja julkaistu, kuten Helsingin kaupunginkirjaston luokitusjärjestelmän (HKL) ja YSA/YSO:n yhdistelmä Kysy kirjastohoitajalta -palveluun (Kirjastot.fi), Yalen yliopiston Outline of Cultural Materials -luokituksen (OCM) muunnos, Iconclass Valtion taidemuseon laatimilla suomalaisilla käännöksillä ja Kongressin kirjaston massiivinen Library of Congress Subject Headings (LCSH). Suuri osa muunnoksista, mukaan lukien moni koti- ja ulkomainen asiasanasto SKOS-muodossa ilman varsinaista ontologointia, on julkaistu avoimesti ONKI-palvelun kautta. Getty-säätiön sanastojen osalta tämä ei kuitenkaan esimerkiksi ole mahdollista lisenssiehdoista johtuen.

---

<sup>18</sup> <http://www.iconclass.org/>

## 2.2 Toimijaontologiat

Keskeinen resurssi sisällönkuvailussa ovat toimijat, ts. henkilöt, organisaatiot ja näiden ryhmät. Kirjastomaailmassa puhutaan auktoriteettitietueista. Kansainvälinen tunnettu referenssiontologia/sanasto on esimerkiksi Getty-säätiön ULAN, jota käytetään mm. Kulttuurisammossa. Saksan, Ranskan ja Ruotsin kansalliskirjastot ovat julkaisseet omat auktoriteettitietueensa avoimessa Linked Data -muodossa, mutta erilaista metadatamallia käyttäen. FinnONTO-hankkeessa on kehitetty useitakin prototyyppejä toimijaontologioista, ja esimerkiksi Kulttuurisammossa (ja siihen perustuvassa Kirjasammossa) on kymmeniä tuhansia eri toimijoita.

FinnONTO:n toimijaontologioita ei ole julkaistu yleisessä verkossa.

## 2.3 Paikkaontologiat

FinnONTO:ssa on kehitetty paikkaontologioita sekä nykyisille että historiallisille paikoille, joita molempia tarvitaan museosalalla.

1. **Nykyiset paikat.** Maanmittauslaitoksen paikannimirekisteriin (n. 800 000 paikkaa) ja ulkomaisiin paikkatietolähteisiin (USA:n GEOnet Names Service) perustuen kehitettiin ontologia ja sille ONKI Geo -palvelu, jonka demonstraattori julkaistiin verkossa (Lindroos, 2008), mutta mm. paikka-aineiston suuren määrän (useita miljoonia paikkoja) ja puutteellisen hierarkkisen rakenteen takia sitä ei ole käytetty FinnONTO:n semanttisissa portaaleissa. Suomen kunnista on muodostettu erillinen Kunnat 2011 -ontologia Tilastokeskuksen aineistoon perustuen.
2. **Historialliset paikat.** Suomen kunnille ja niiden muutoshistorialle kehitettiin Suomen ajallinen paikkaontologia SAPO (Hyvönen et al., 2011), joka on saatavilla ONKI-palvelusta. Tätä on täydennetty vanhemmilla aineistoilla Venäjän ja Ruotsin vallan ajalta (mm. linnaläänit) sekä luovutetun Karjalan alueen tarkemmilla paikoilla (tuhansia paikkoja). Tältä pohjalta työtä ollaan jatkamassa tavoitteena laaja kansallinen historiallisten paikkojen ontologia. Kehitysvälineenä on SAHA-HAKO-työkalu<sup>19</sup>. SAPO on käytössä Kulttuurisammossa.

Lisäksi kulttuurimatkailun tarpeisiin on kehitetty point-of-interest-ontologiaa POI, jossa on luonnon ja rakennetun ympäristön paikkatyyppejä. Maanmittauslaitoksen paikannimirekisterin n. 50 paikkatyyppin luokitusta on täydennetty erityisesti kansainvälisen Geonames-järjestelmän luokilla sekä suomalaisilla paikkakäsitteillä. Kerättynä on n. 4600 käsitettä, mutta niiden järjestäminen ja organisointi on vielä kesken. POI-ontologia on luonteeltaan paikkatietoon liittyvä yleisontologia, ei varsinainen paikka(yksilö)ontologia.

## 2.4 Aikaontologiat

Ajan esittämiseen haasteita ovat mm. aikamääreiden epätäsmällisyyden esittäminen (esimerkiksi arkeologiassa) ja aikakausien riippuvuus paikasta ja kulttuurista (esimerkiksi pronssikausi Suomessa oli eri aikaan kuin Välimeren maissa). FinnONTO:n YSO-ontologiasta löytyy aikakausia ja aikakäsitteitä. Epätäsmällisen ajan esittämiseen on käytetty vastaavan tyyppistä metatietomallia kuin CIDOC-CRM-järjestelmässä ja aikamäärien esittämiseen W3C:n standardia. FinnONTO:n aikaontologiaa, josta on käytetty työnimeä AIKO, ei tätä kirjoitettaessa ole julkaistu.

<sup>19</sup> <http://www.seco.tkk.fi/services/saha/>

## 2.5 Tapahtumaontologiat

Toimija- ja paikkaontologioiden ohella kolmas, jatkossa yhä tärkeämpi yksilöontologiatyyppi on tapahtumaontologiat. Tapahtumakeskeinen luettelointi on esimerkiksi CIDOC-CRM:n ydinajatus, mutta sanastoja/ontologioita itse tapahtumille ei ole olemassa samaan tapaan kuin toimijoille ja paikoille. Asiasanastoihin voidaan ottaa ja on otettukin mukaan tapahtumia. Esimerkiksi ”Pahkinäsaaren rauha” ja ”toinen maailmansota” löytyvät YSO:sta. Tapahtumat eivät kuitenkaan ole yleiskäsitteitä vaan yksilökäsitteitä. Niillä on mutkikas ontologinen rakenne, joten erillinen tapahtumaontologia on tarpeen. Tapahtumiin esimerkiksi liittyy paikkatietoa, henkilöitä ja toisia tapahtumia, joiden hyödyntäminen ja visualisointi kulttuurialan verkkosovelluksissa on hyödyllistä. Tätä on demonstroitu Kulttuurisammossa aineettoman kulttuurin, erityisesti historian ja taitojen tallentamisessa.

FinnONTO:ssa visiona on ollut luoda kansallisen historian tapahtumaontologia HISTO. Ontologia on kehitetty yhteistyössä Suomen historian tutkijoiden Agricola-yhteisön kanssa, jonka aikajana ”Suomen historian kronologia”<sup>20</sup> muodostaa HISTO:n selkärangan. HISTO:a kehitetään muiden FinnONTO:n yksilöontologioiden tapaan SAHA-HAKO-järjestelmän avulla.

”Tapahtumalla” voidaan museotalalla viitata paitsi historiallisiin tapahtumiin, myös museoiden tapahtumiin, kuten näyttelyihin, konsertteihin, työpajoihin jne. FinnONTO:ssa on kehitetty ontologinen malli myös yleisötapahtumille, ja esimerkiksi Kulttuurisammosta löytyvät Museoliiton näyttelykalenterin perusteella muodostetut tapahtumat. Museoiden sisäisiä tapahtumia, kuten esineiden konservointi- tai näyttelyhistoriaa, ei FinnONTO:n piirissä ole tutkittu.

## 2.6 Nimistöontologiat

Edellä kuvattujen yleiskäsite- ja yksilöontologioiden ohella museoissa tarvitaan ajoittain muitakin laajoja sanastoja, joita kutsun tässä *nimistöontologioiksi*. Tällaisiin nimistöt tai nimilistat kuuluvat yleensä ns. vapaisiin asiasanastoihin (free keyword), joita ei niiden monilukuisuuden vuoksi voida luetella asiasanastoissa, mutta joita kuitenkin tarvitaan. Tällaisia ovat esimerkiksi kasvien, mineraalien tai eläinten lajit. Myös yksilöontologioiden käsitteitä, kuten paikkojen ja henkilöiden nimiä käytetään vapaina asiasanastoina. Museossa olevan kilpikonnan kuoresta tehdyn korun yhteydessä saatetaan haluta kertoa, että korun materiaali on kreikankilpikonnan kuorta.

FinnONTO:ssa on luotu ontologiat kaikista Luonnontieteelliseen keskusmuseon ns. nimilistoista. Lisäksi Bird Life -organisaation maailman linnut -nimilistasta on tuotettu kolmikielinen AVIO ontologia (n. 10 000 taksonia) ja nisäkkäistä MAMO ontologia (n. 6 000 taksonia), mukaan lukien nisäkästoimikunnan uusimmat suomenkieliset käännökset. Lisäksi putkilokasveista on oma laaja sanastonsa, joka perustuu Suomen Biologian Seura Vanamon KASSU-sanastoon. ONKI-palvelussa on n. 88 000 eläin- ja kasvimaailman nimikettä, jonka voidaan ottaa käyttöön rajapintojen kautta.

ONKI-leijukkeen valitsimeen voidaan ottaa mukaan valikko, jonka avulla haku asiasanoituksessa voidaan kohdistaa eri sanastoihin. Tällöin automaattitäydennyksen tarjoamat vaihtoehdot fokusoituvat tarkemmin kulloinkin valitulle alueelle.

## 2.7 Yhteenveto

FinnONTO:ssa tehty tutkimus- ja kehitystyö on kattanut kaikkia keskeisiä käsiteontologiatyyppejä, joita museoalalla tarvitaan. Joillain alueilla ollaan pidemmällä kuin toisilla, ja kaikkialla tarvitaan lisätyötä. FinnONTO-hankkeen tavoitteena ei varsinaisesti ole edes ollut ontologioiden kehittäminen, vaan

<sup>20</sup> <http://agricola.utu.fi/hist/kronologia/index.php>



lähestymistavan tutkiminen, työvälineiden ja menetelmien kehittäminen sekä mahdollisuuksien arvioiminen ja demonstroitu pilottisovelluksilla. Ontologisen lähestymistavan etuja on testattu useissa maailmankin mittakaavassa edistyksellisissä verkkopalveluissa. Tavoitteena on ollut, että hankkeen tuloksena saataisiin maassamme käyntiin systemaattinen, jatkuva ontologiatyö eri alojen välisenä yhteistyönä.

### 3 ONKI-ONTOLOGIAPALVELUT MUSEOILLE

Kehitetyt ja julkaistut FinnONTO-ontologiat ovat sekä ihmiskäyttäjän selailtavissa perinteiseen tapaan, että myös suoraan kytkettävissä museojärjestelmiin rajapintojen kautta. Molemmat toiminnallisuudet on toteutettu FinnONTO:n Kansallisen ontologiakirjastopalvelun ONKI avulla (Tuominen et al., 2009; Viljanen et al., 2009; Tuominen et al., 2012). Kantavana ajatuksena on, että keskitetyt ONKI-rajapintapalvelut toteuttavat yhteiset sanastopalvelun toiminnallisuudet, kuten termien haun ja URI-tunnisteiden kopioinnin luettelointijärjestelmään. Etuna on, että *jaettuja keskitettyjä toiminnallisuuksia ei tarvitse toteuttaa erikseen eri asiakasjärjestelmissä*. Näin museoilta säästyy vaivaa ja rahaa, kustannuksia voidaan jakaa ja sanastot ovat automaattisesti ajan tasalla kaikkien käytettävissä. Tärkeintä kokonaisuuden kannalta on kuitenkin se, että samalla eri muistiorganisaatioiden sisällöntuotantoa saadaan harmonisoitua. Näin sisällöntuotannon laatu paranee ja ennen kaikkea eri organisaatioiden kokoelmatiedot saadaan automaattisesti yhteentoimiviksi. Semanttinen yhteentoimivuus on välttämätöntä MuseoSuomen, Kulttuurisammon, Kirjasammon ja Matkailusammon kaltaisissa semanttisissa ”älykkäissä” sovelluksissa. Yhteentoimivuus on tärkeää myös KDK:n ja nykyisen European kaltaisissa perinteisissä järjestelmissä.

Seuraavassa yhteenveto ONKI:n ihmiskäyttöliittymästä ja rajapinnoista.

#### 3.1 Ihmiskäyttöliittymä

Ihmiskäyttäjälle Kansallisesta ontologiakirjastopalvelusta ONKI on kehitetty useampiakin versioita. Alla yhteenvetona joitain toiminnallisuuksia:

- Sanastojen etsimiseen on käytettävissä oma sanastojen luokituksiin perustuva hakukone.
- Käsitteiden haku tapahtuu automaattitäydennyksellä tai käsitepuita ja linkkejä selaamalla.
- Käyttöliittymä on monikielinen ja monikielisten sanastojen kieli voidaan valita erikseen.
- Haku voidaan suorittaa monesta eri sanastosta samanaikaisesti ja jopa monesta eri ontologia- tai tietopalvelusta (esimerkiksi SAHA tai Kulttuurisampo), jotka toteuttavat ONKI:n rajapinnan.
- ONKI:in voidaan yhdistää myös haku RDF-tietokannasta, joita voidaan tuottaa esimerkiksi SAHA-työkalulla.
- Palveluun voi ehdottaa ja ladata uusia sanastoja käyttäjien toimesta. Ylläpitäjien harkinnan mukaan nämä voidaan julkaista välittömästi, jos ne ovat standardien mukaisessa eheässä muodossa.

#### 3.2 ONKI-rajapinnat

ONKIn kansainvälisenä innovaationa on ollut ajatus sanastojen hyödyntämisestä myös rajapintojen kautta valmiina toiminnallisuuksina. Näitä ovat:

1. **ONKI-leijuke.** ONKI-leijukkeen (widget) avulla esimerkiksi luettelointijärjestelmään voidaan tuoda automaattitäydennyksellä varustettu monikielinen haku vain 1-2 rivillä HTML/Javascript-koodia. Leijukekoodi voidaan luoda ja testata interaktiivisesti ONKI:ssa<sup>21</sup>.
2. **ONKI REST-rajapinta.** Yksinkertainen REST-henkinen rajapinta.
3. **ONKI-verkkopalvelurajapinta** (Web Service). Monipuolisempi Web Services -teknologiaan perustuva rajapinta.
4. **ONKI SPARQL -rajapinta.** Kehitteillä oleva uusi W3C:n semanttisen webin SPARQL-kyselykielistandardiin perustuva rajapinta. Tällä on jo ensimmäinen koekäyttäjä.

ONKI:in sanastoja voidaan ladata ja päivittää automaattisesti ulkoisista julkaisuista. Tällainen mekanismi on toteutettu mm. Kansalliskirjaston YSA:lle, jonka viimeisin versio ladataan palvelimelle Kansalliskirjastosta päivittäin.

MUUSA-museot (34)	E-Kuva-museot (24)	Memoron- ja Renki-museot (19 museota)	SIIRI-Museot (7 museota)	Muut (1)
Valtion taidemuseo	Forssan museo	Designmuseo	Vapriikki	Espoon kaupunginmuseo
Amos Andersonin taidemuseo	Helsingin yliopistomuseo	Sotamuseo	Suomen jääkiekkomuseo	
Hyvinkään taidemuseo	Hyvinkään kaupungin museo	Suomen ilmailumuseo	Hiekan taidemuseo	
Hämeenlinnan taidemuseo	Ilomantsin Museosäätiö	Keski-Suomen ilmailumuseo	Tampereen suojeluskunta- ja lottamuseo	
Joensuun taidemuseo	ITE-museo	Ilmatorjuntamuseo	Hämeenkyrön kunta	
Järvenpään taidemuseo	Kansan Arkisto	Suomen maatalousmuseo Sarka	Nokian kaupunki	
K.H.Renlundin museo	Keravan museon	15 maatalousteemaista museota ja kokoelmaa	Pirkanmaan sairaanhoitopiiri	
Kouvolan taidemuseo	Kultamuseo			
Kuopion taidemuseo	Lapin metsämuseo			
Lahden kaupunginmuseo	Mäntsälän museotoimi			
Lahden taidemuseo ja Julistemuseo	Nurmeksen museo			
Lappeenrannan taidemuseo	Nurmijärven museo			
Lapuan taidemuseo	Pielisen museo			
Lönnströmin taidemuseo	Polisimuseo			
Nelimarkka-museo	Päivälehdin museo			
Oulun taidemuseo	Rikosmuseo			
Pohjanmaan museo	Suomen Metsämuseo Lusto			
Porin taidemuseo	Suomen metsästysmuseo			
Päijät-Hämeen sosiaali- ja terveystyöryhmä	Suomen siirtolaisuusmuseo			
Rauman taidemuseo	Sähkötalo Elektra			
Riihimäen taidemuseo	Tekniikan museo			
Salon taidemuseo	Työväen Arkisto			
Sara Hildénin taidemuseo	Työväenmuseo Werstas			
Serlachius-museot	Verlan tehdasmuseo			
Suomen Pankki				
Taidesäätiö Merita				
Tampereen taidemuseo				
Tikanojan taidekoti				
Turun museokeskus				
Turun taidemuseo				
Tuusulan taidemuseo				
Valtion taideteostoimikunta				
Vantaan taidemuseo				
Villa Gyllenberg taidemuseo				

**Taulukko 3.1** ONKI-järjestelmään kytketyt suomalaiset museot 24.4.2012.

ONKI-palvelu on käytettävissä rajapintojen kautta tätä kirjoitettaessa ainakin 84 suomalaisen museon luettelointijärjestelmässä taulukon 3.1 mukaisesti.

ONKI:n ihmisille tarkoitettulla käyttöliittymällä on n. 8 700 yksilöityä käyttäjää kuukaudessa. Keskimääräinen palvelun käyttöaika on hieman yli 9 minuuttia; n. 2% sivuille tulleista poistuu niiltä

<sup>21</sup> <http://onki.fi/widget/selector/?l=en>

välittömästi. Rajapintapalvelun käyttäjäksi on rekisteröitynyt n. 296 eri tahoja, 367 domain-nimeä ja asiakasrekisterissä on 397 sähköpostiosoitetta.

ONKI-palvelu on mukana tärkeänä elementtinä valtiovarainministeriön (VM) kansallisessa tietoarkkitehtuurisuunnitelmassa. Kansalliskirjasto ja FinnONTO-hanke ovat laatineet VM:lle ja opetus- ja kulttuuriministeriölle (OKM) rahoitushakemuksen ONKI-palvelun ja siihen liittyvän ontologiatyön koordinoinnin vakinaistamisesta kansallisena infrastruktuuripalveluna, mutta päätöstä ministeriöissä ei ole vielä tehty.

### 3.3 Yhteenveto

ONKI:n käyttöliittymä ja alla olevat mekanismit on kehitetty tutkimuksen lähtökohdista. Prototyyppi toimii ja on nykyisin varsin laajassa käytössä, mutta on ehkä liiankin monipuolinen ja teknisesti monimutkainen puhtaasti tuotannollisen käytön näkökulmasta. Haittana on järjestelmän ajoittainen hitaus ja silloin tällöin tapahtuva järjestelmän kaatuminen. Palvelun vakinaistamisen yhteydessä, josta parhaillaan neuvotellaan valtiovarainministeriön, opetusministeriön, Kansalliskirjaston ja FinnONTO-hankkeen kesken tarjoutuu mahdollisuus ohjelmiston yksinkertaistamiseen ja viimeistelyyn samalla, kun uudet vastuuorganisaatiot ryhtyvät palvelua ylläpitämään ja kehittämään. Ehdotuksen mukaan ONKI-palveluun kuuluu paitsi palvelimien tekninen ylläpito ja kehittäminen myös ontologioiden ja sanastojen sisällöllinen koordinoitu kehittäminen.

## 4 EHDOTUS MUSEOALAN ONTOLOGIAINFRASTRUKTUURIKSI

Tässä luvussa esitetään ehdotus siitä, miten museoalan sanasto- ja ontologiatyö kannattaisi organisoida ja niihin liittyviä palveluita jatkossa kehittää. Esitän näkemyksen omasta puolestani, mutta siinä uskossa että käsityksillä on myös laajempaa kannatusta FinnONTO:n kymmenien osallistujaorganisaatioiden sekä tutkijoiden taholla. Tavoitteena on edesauttaa FinnONTO:n eri tulosten arviointia ja käyttöönottoa, herättää keskustelua eri ratkaisumalleista ja antaa suuntaviivoja tuleville tutkimus- ja kehityshankkeille.

Ehdotus perustuu FinnONTO- ja Kulttuurisampo-työn taustalla olevaan visioon kansallisiin asia- ja muihin sanastoihin perustuvasta, yhteisöllisestä sisältö/tietoinfrastruktuurista. Sen pohjana ovat kansainväliset W3C:n semanttisen webin standardit ja Linked Open Data -ajattelu. Vuosien kuluessa käsitykseni siitä, että semanttisen webin teknologiat ovatärkevin perusta kulttuurialan sisältöjen esittämiseen, yhdistämiseen ja julkaisemiseen verkossa on vain vahvistunut, ja viimeaikainen teknologian käyttöönotto esimerkiksi Europeanassa ja kirjastoalan uusissa standardeissa vahvistaa käsitystä edelleen. Vision kulttuurisisältöjen semanttisesta julkaisutuotannosta olen esittänyt tarkemmin Kulttuurisammon julkistustilaisuudessa 2008 esitetyssä artikkelissa ”Kulttuurisampo - suomalainen kulttuuri semanttisessa webissä. Muistiorganisaatioiden ja kansalaisten yhteisöllinen kansallinen julkaisujärjestelmä” (Hyvönen, 2008b), ja vastaavasti ontologiainfrastruktuurin osalta ONKI-palvelun julkistamistilaisuudessa artikkelissa ”FinnONTO-malli kansallisen semanttisen webin sisältöinfrastruktuurin perustaksi - visio ja sen toteus” (Hyvönen, 2008a). SeCo oli mukana W3C:n ”Library Linked Data Incubator Group” -työryhmässä, jonka loppuraportit kirjastoalalle on julkaistu W3C:n sivuilla<sup>22</sup>, samoin Stanfordin yliopiston työryhmässä, jonka loppuraportti ”Linked Data for

<sup>22</sup> <http://www.w3.org/2005/Incubator/lld/>

Libraries, Museums, and Archives: Survey and Workshop Report” on niin ikään verkosta saatavilla<sup>23</sup>. Europeanan sivuilta löytyy lisätietoa Europeana Linked Data -työstä<sup>24</sup>.

Kulttuurialan semanttisessa ontologiainfrastruktuurissa tarvitaan mielestäni kolme pääosaa:

- 1) **Ontologiat.** Joukko toisiinsa sillattuja, kansalliseen terminologiaan perustuvia ontologioita, joilla on siltauksia myös kansainvälisiin sanastoihin ja tietosisältöihin. Näin suomalaiset kulttuurisisällöt saadaan kuvattua semanttisesti ja kytkettyä sekä toisiinsa että kansainvälisiin aineistoihin ja palveluihin, kuten Linked Open Data -pilveen ja Europeanaan.
- 2) **Ontologiapalvelut.** Ontologiapalveluiden kautta ajantasaiset ontologiat voidaan ottaa kustannustehokkaasti käyttöön museoissa ja muissa muistiorganisaatioissa. Tällainen palvelu kannattaa toteuttaa keskitetysti perusteluina kustannussäästöt, ajantasaisuus ja yhteentoimivuus. Keskitettyjä ONKI-palveluita on pilotoitu useita vuosia ONKI Living lab -ympäristössä ja sen seurauksena kymmenien museoiden luettelointijärjestelmät on jo kytketty ONKI:iin. Kokemusten perusteella innovaatio toimii ja palvelu pitäisi nyt vakinaistaa.
- 3) **Ontologiajärjestelmän ylläpitoprosessi.** Systemaattinen ja koordinoitu kansallinen organisaatio ja mekanismi, joka kantaa vastuun museoissa tarvittavien sanastojen yhteisöllisesti kehittämisestä jatkossa. Tässä tarvitaan keskitettyä koordinaointia, mutta ontologiatyö kannattaa hajauttaa tarpeen mukaan sinne, missä siihen on paras asiantuntemus ja resursseja käytettävissä.

Seuraavassa ehdotus näiden kolmen pääkomponentin kehittämisestä jatkossa.

## 4.1 Ontologiat

Seuraavassa esitellään käsitykseni mukaan tärkeimmät museoalalla tarvittavat ontologiat, joiden ylläpito pitäisi tavalla tai toisella järjestää.

### 4.1.1 Yleiskäsitteet KOKO-MAO

Museoalalla tarvittavista ontologioista kannattaa monesta vaihtoehdosta valita kaikki, eli KOKO-ontologia, jonka perustana ovat eri alojen asiasanasto-ontologiat. Yleiskäsitteiden osalta museoalan ylläpidettäväksi ontologiaksi kannattaa ottaa yhdistelmä MAO+TAO+VALO, jota kutsutaan tässä artikkelissa jatkossa nimellä KOKO-MAO. Tämä siksi, että nämä ontologiat ovat käsitteistöltään ja sovellusaloiltaan voimakkaasti museoalaan ja toisiinsa linkittyneitä, jolloin niiden kehittäminen toisistaan erillään on haasteellista. Lähtökohtaisesti KOKO-työssä kannattaa pyrkiä mieluummin vähentämään kuin lisäämään ontologioiden määrää, koska yhteistyö toiminee joustavammin yhden yhteisen ontologian kehittäjien kesken kuin monen eri ryhmän välillä. Nykyinen KOKO:n ontologiajako perustuu suoraan olemassa oleviin asiasanastoihin ja historiallisiin syihin, eikä välttämättä ole semanttisesti hyvin perusteltu. Vaikka museoala keskittyisi ontologian kehitystyössä KOKO-MAO:oon, kannattaa luetteloinnissa ja sovelluksissa kuitenkin käyttää koko yhteisön KOKO-ontologiaa, johon kuuluu monia muitakin käsiteontologioita.

KOKO-MAO-ontologian ylläpitoa varten pitäisi perustaa museoalan työryhmä, jonka keskeisenä tehtävänä olisi 1) sanaston ja sen struktuurin editointi kentän tarpeiden mukaan, 2) eri sanastoista

<sup>23</sup> <http://www.clir.org/pubs/abstract/reports/pub152>

<sup>24</sup> <http://pro.europeana.eu/linked-open-data>

syntyneiden turhien päällekkäisyyksien karsiminen ja 3) siltausten ylläpitäminen kansallisen KOKO-ontologiapilven muihin ontologioihin, erityisesti YSO:on. Työlistalla pitäisi myös olla 4) siltaukset kansainvälisiin sanastoihin, kuten Getty-säätiön AAT, jotta suomalaiset sisällöt voitaisiin yhdistää esimerkiksi European aineistoihin. 5) Taidemuseoissa ja MUUSA:ssa käytetty Iconclass ja sen KOKO-siltaukset vaativat myös kehittämistä ja ylläpitoa.

Yleiskäsitteiden ohella tarvitaan myös yksilöontologioita, joissa luetellaan yleiskäsitteiden (class, type) nimettyjä yksilöitä (individual, instance) sopivan metatietomallin mukaisesti kuvattuna. Esimerkiksi "kaupunki" on yleiskäsite ja yksittäiset kaupungit kuten "Espoo" löytyvät paikkayksilöontologiasta (gazetteer, authority file). Monissa perinteisissä sanastoissa kuten YSA:ssa on sekä yleis- että valittuja yksilökäsitteitä, kuten paikkoja. Yksilöontologiat on kuitenkin syytä erottaa yleiskäsiteontologioista seuraavista syistä: 1) Yksilöiden ontologinen rakenne hyvin erilainen kuin käsitteiden. Esimerkiksi henkilöillä on syntymäaika ja paikka, ammatti, eri nimiä eri aikoina jne., paikoilla taas erikielisiä nimiä, koordinaattitietoa, paikkatyyppi ja topologiaa suhteita toisiin paikkoihin. 2) Yksilöiden määrä on erittäin suuri, esimerkiksi Maanmittauslaitoksen paikannimirekisterissä on n. 800 000 paikkaa Suomesta. 3) Yksilöiden tunnistaminen ja hakeminen pelkän nimen perusteella on usein haasteellista, sillä eri yksilöillä on usein sama nimi, esimerkiksi Suomessa on 49 Pyhäjärvi-nimistä paikkaa, jotka voivat vielä olla tyyppiltään erilaisia, kuten vakavesiä tai kyliä. Lisäksi yksilöiden nimillä on usein erilaisia kirjoitusasuja eri kielissä ja translitterointijärjestelmissä. Näiden tallentaminen ja hyödyntäminen tietojärjestelmissä on erittäin tärkeää käytettävyyden kannalta.

Yksilökäsitteiden osalta kriittisimmät ontologiat ovat toimija- ja paikkaontologiat. Näiden kehittämistä vastuu olisi syytä lisätä museoalan KOKO-MAO-työryhmän toimintaan. Tietääkseni museoalalla ei Suomessa tällaista työtä ole toistaiseksi juurikaan tehty, joten kyseessä olisi uusi avaus. Jokaisessa luettelointijärjestelmässä toki on oma ratkaisunsa paikkojen ja toimijoiden esittämistä varten, mutta nyt haasteeseen pitäisi tarttua kansallisella ja kansainvälisellä tasolla, kuten tietoa halutaan yhdistellä yhtä organisaatiota tai luettelointijärjestelmää laajemmissa konteksteissa.

#### **4.1.2 Toimijaontologia TOIMO**

Tavoitteena pitää olla yhteinen malli ja rekisteri toimijaontologiaksi TOIMO, joka sisältää (historiallisia) henkilöitä, organisaatioita ja ryhmiä. Esikuvana tästä kannattaa käyttää Getty säätiön ULAN rekisteriä, jota on hyödynnetty mm. Kulttuurisammossa. Henkilötietojen julkaisemiseen liittyy tietosuojakysymyksiä, mutta esimerkiksi Ranskan, Saksan ja Ruotsin kansalliskirjastojen auktoriteettitietojen julkaiseminen avoimena yhdistettynä tietona osoittaa, nämä haasteet eivät voi olla ylipääsemättömiä. Kansallista museoalan toimijaontologiaa TOIMO ei edes tarvitse julkaista avoimesti, kunhan sellainen saadaan museoiden sisäiseen yhteiskäyttöön ONKI-tyyppisen palvelun kautta.

Toimijarekisteri kannattaa sillata Wikipediassa/DBPediassa käytettyihin henkilötunnisteina toimiviin verkko-osoitteisiin (URI), jota kautta avautuvat yhteydet LOD-pilven runsaisiin sisältöihin, vaikkapa tietoihin henkilöitä käsittelevistä teoksista tai säveltäjän levytettyyn tuotantoon. Samaten tarvitaan siltauksia kotimaisiin henkilöaineistoihin, kuten Kulttuuri/Kirjasampoon, Suomalaisen Kirjallisuuden Seuran kansallisbiografiaan (n. 6000 elämäkertaa) ja Kansalliskirjaston auktoriteettitietoihin.

FinnONTO-hankkeessa on kehitetty toimijaontologioita SAHA-HAKO-työkalun avulla, ja koottuna on kymmeniä tuhansia toimijoita. Käytettävissä on myös Getty Säätiön ULAN-rekisteri semanttisen webin RDF-formaatissa. TOIMO-ontologia on tutkimus- ja kehitystyön kohteena FinnONTO:n jatkohankkeessa Linked Data Finland (LDF)<sup>25</sup>, jossa yhtenä case-tutkimuksena toteutetaan prototyyppi "Semanttisesta

<sup>25</sup> <http://www.seco.tkk.fi/projects/ldf/>

kansallisbiografiasta” nykyisen SKS:n Kansallisbiografian pohjalta. TOIMO-ontologian kehitystyötä kannattaisi jatkaa tältä pohjalta. Tähän kannattaisi perustaa laajempi erillinen kehityshanke, jossa on mukana paitsi museoita myös muiden muistiorganisaatioiden edustajia erityisesti auktoriteettitiedostoja tuottavista kirjastoista ja arkistoista. Lisäksi tarvitaan teknistä osaamista uudesta teknologiasta.

### **4.1.3 Paikkaontologiat SAPO ja SUO**

Paikkojen osalta museosisältöjen kuvailussa tarvitaan sekä 1) ontologiaviittauksia (URI) että 2) koordinaattitietoa. Ontologisen viittauksen avulla luettelointitieto voidaan yhdistää paikkaontologian kautta muihin paikkoihin (esimerkiksi ”Espoo” ”pääkaupunkiseutuun” tai ”Etelä-Suomen lääniin”), käsitteisiin (esimerkiksi Espoo on ”kaupunki”), ja sisältökuvauksiin (esimerkiksi Espoossa valmistettu esine). Koordinaattitieto taas tarvitaan karttanäyttöjä varten. Paikkayksilöiden kautta on teoriassa mahdollista löytää koordinaatteja (suoraan tai ns. geokoodauksen avulla) ja koordinaattien kautta voi löytää paikkaontologian paikkoja, mutta käytännössä tämä on monasti haasteellista: 1) Paikkatieto paikkaontologioissa on monasti puutteellista tai epätäsmällistä. Erityisesti historiallisten paikkojen ja alueiden sijainnista ei useinkaan ole tarkkaa tietoa. 2) Sama alue kuuluu tyypillisesti samanaikaisesti moneen eri paikkaan, jolloin pelkkä koordinaattitieto ei aina yksilöi mihin paikkaan kuvattu sisältö liittyy, ja tällainen yksilöinti voi olla sovelluksen kannalta tärkeää. Esimerkiksi arkeologisen löytöpaikan halutaan luultavasti liittyvän tiettyyn historialliseen alueeseen, vaikkapa linnalääniin, eikä ehkä samalla alueella nykyisin olevaan hallinnolliseen alueeseen.

Vaikka koordinaattitietoa paikasta olisi käytettävissä, halutaan joskus tehdä viittaus paikkaan ilman, että halutaan ottaa kantaa koordinaatteihin. Esimerkiksi paikoista on usein käytettävissä vain tieto niiden keskipisteestä. Jos esineen tiedetään valmistetun ylipäätään Espoossa, ei Espoo-viittauksella luetteloinnissa haluta väittää tai sovelluksessa välttämättä visualisoida esineen olevan valmistetun juuri Espoon keskipisteessä. Ongelma ratkeaa, jos käytettävissä on mahdollisuus antaa paikkatieto tarvittaessa sekä ontologisella paikkaviittauksella että koordinaattien avulla toisistaan riippumatta.

Ontologisten paikkaviittausten osalta museoalalla tarvitaan keinoja viitata sekä historiallisiin että nykyisiin paikkoihin. FinnONTO:ssa historiallisia paikkoja on kerätty Suomen ajalliseen paikkaontologiaan SAPO (Hyvönen et al., 2011). Modernien suomalaisen paikkojen ontologiasta, joka perustuu Maanmittauslaitoksen paikannimirekisteriin, käytetään nimeä SUO. Koordinaattitiedon esittämisen osalta tarvitaan pisteitä (esimerkiksi Topeliuksen syntymäkodin pistemäinen sijainti), reittejä (esimerkiksi Kuninkaantien tai opastetun kulttuurireitin kuvaus) ja alueita (esimerkiksi historiallinen Porvoo). Koordinaattien esittämiseen kannattaa käyttää modernia WGS84-järjestelmää, joka on eniten käytetty kansainvälinen järjestelmä verkossa ja yhteentoimiva mm. mobiililaitteiden GPS-järjestelmien kanssa.

FinnONTO:ssa kehitetty ja käytetty työkalu paikkaontologioiden kehittämiseen on SAHA-HAKO, johon on toteutettu mm. mahdollisuus pisteiden, reittien ja alueiden interaktiiviseen kuvailuun kartoilla. Työn jatkuu LDF-hankkeessa tavoitteena mm. julkaista SAPO ja SUO avoimena ONKI-tyyppisenä palveluna, jossa on mahdollisuus uusien paikkojen reaaliaikaiseen ja hajautettuun lisäämiseen ja toimittamiseen verkossa.

Museoalalle pitäisi perustaa työryhmä ylläpitämään historiallisten paikkojen ontologiaa monestakin syystä. 1) Paikkatieto on erittäin keskeistä ja sitä liittyy suurimpaan osaan museoissa luetteloitavia kohteita. 2) Karttanäyttöjen ja GPS-paikkannuksen myötä paikkatiedon hyödyntämismahdollisuudet sovelluksissa ovat voimakkaassa kasvussa.

Museoalan ontologioiden ja ONKI-palvelun laajentaminen yleiskäsitteistä toimijoihin ja paikkoihin olisi ensiarvoisen kriittistä aineistojen yhteentoimivuuden kannalta, koska kokoelmatiedoissa on erityisen runsaasti viittauksia paikkoihin ja toimijoihin.

Muiden edellä lueteltujen ontologiatyyppien tilanne ei ole aivan yhtä kriittinen, mutta on tärkeää jatkaa työtä niidenkin osalta. Seuraavassa toimenpide-ehdotuksia:

#### **4.1.4 Aikaontologia AIKO**

Museoalalla olisi tärkeää luoda ja/tai ottaa käyttöön yhteinen ajan esittämisen metatietomalli ja viiteontologia keskeisistä aikakäsitteistä ja jaksoista. Pelkästään määrämuotoisilla vuosiluvuilla ja päiväyksillä päästään kuitenkin jo melko pitkälle. Ajan esittämisessä voidaan harkita siirtymistä aikaintervalleista semanttisesti rikkaampien, CIDOC-CRM-standardin mukaisten sumeiden ajankuvausten käyttöön, milloin tällaiseen on tarvetta<sup>26</sup>. Yksinkertaiset intervallit ovat rikkaamman mallin erikoistapaus ja yhteentoimivia sen kanssa.

#### **4.1.5 Tapahtumaontologia HISTO**

FinnONTO:ssa visiona on ollut luoda Suomen historian tapahtumaontologia HISTO. Ontologia on kehitetty yhteistyössä Suomen historian tutkijoiden Agricola-yhteisön kanssa, jonka aikajana muodostaa HISTO:n selkärangan. HISTO:a kehitetään muiden FinnONTO:n yksilöontologioiden tapaan SAHA-HAKO-järjestelmän avulla. HISTO:n varhainen koeversio on käytössä Kulttuurisammossa, ja uuden koko Agricolan kronologian (aikajanan) kattava aineisto jo olemassa SAHA-HAKO:ssa, mutta ontologian muokkaaminen, esimerkiksi toimijoiden eristäminen systemaattisesti TOIMO-ontologiaan, on vielä kesken ja jatkuu LDF-hankkeen puitteissa sekä Kulttuurirahaston Kulttuurisampo-projektin tuella.

#### **4.1.6 Alakohtaiset nimistöontologiat**

Erytisnimistöjä voidaan ottaa käyttöön ONKI:n kautta sitä mukaan kun sellaisia julkaistaan. Nimistöt ovat erityisen tärkeitä erityisalojen museoille, esimerkiksi eliöontologiat luonnontieteellisille museoille, mineraalit kivimuseoille, tekniikan termit tekniikan museoille jne.

## **4.2 Ontologiajärjestelmän ylläpito**

Ontologiajärjestelmän prototyypin kehittämisestä ja ylläpidosta on toistaiseksi kantanut vastuun FinnONTO-hanke yhteistyössä asiasanastojen ja ontologioiden kehittäjien kanssa. FinnONTO:n piirissä on laadittu suunnitelma erityisen ONKI-konsortion perustamisesta, joka ottaisi vastuulleen sekä ONKI-palvelun ylläpidon että ontologioiden ylläpitämisen. Työn koordinoinnista ja ONKI-palveluista vastaisi Kansalliskirjasto ja ontologiastruktuurin ylläpidosta myös Sanastokeskus TSK. KOKO-ontologiatyön jatkon varmistamiseksi on pidetty työpaja, jossa päätettiin alustavasti pilotoida ontologiatyötä museo- ja terveysalojen edustajien kanssa MAO+TAO+VALO ontologioiden ja TERO-ontologioiden osalta, mikäli kentältä löytyy riittävästi halukkuutta. Varsinainen sanastotyö tehtäisiin alakohtaisten asiantuntijaryhmien toimesta, jollainen perustettaisiin museoalalle, terveysalalle ja vastaaville tarpeen mukaan. FinnONTO tarjoaisi tähän teknistä tukea ja ontologiaosaamista ja työ voisi jatkua ONKI-konsortion puitteissa, jos ONKI-palvelu vakinaistetaan ja saa rahoitusta.

KOKO-ontologian YSO, YSA:n ontologisoitu muoto, on ontologiainfrastruktuurin kannalta tärkein ja suurin yleissanasto. Oman käsitykseni mukaan YSA-asiasanastoa kehittävän ryhmän kannattaisi

<sup>26</sup> [http://www.cidoc-crm.org/docs/How\\_to%20implement%20CRM\\_Time\\_in%20RDF.pdf](http://www.cidoc-crm.org/docs/How_to%20implement%20CRM_Time_in%20RDF.pdf)

jatkossa ryhtyä kehittämään YSO-ontologiaa. Kahden päällekkäisen sanastostruktuurin kehittäminen on voimavarojen tuhlaamista. Työtä riittäisi kehitystyön ohella esimerkiksi YSO:n alustavien englanninnosten tarkistamisessa, joissa on vielä paljon epätarkkuuksia – kansallisista ydinsanastoista tarvitaan suomen ja ruotsin ohella ehdottomasti englanninkieliset käännökset. YSA:aa tarvitaan jatkossakin eräissä Kansalliskirjaston palveluissa, joihin sanasto on integroitu, mutta tällainen esitysmuoto voidaan tuottaa ohjelmallisesti semanttisesti tarkemmasta YSO-ontologiasta. Jatkossa olisi mahdollisuus siirtyä käyttämään myös huomattavasti laajempaa KOKO-ontologiaa samalla tavalla. Struktuurin ja sen osien viimeistelyssä tarvitaan vielä työtä: FinnONTO:n tavoitteena on ollut kehittää KOKO:sta innovaatiokonsepti, ensimmäinen prototyyppi ja tarvittavaa ydinteknologiaa, muttei varsinaisesti ryhtyä laajempaan ontologiatyöhön.

FinnONTO:n yleiskäsitteiden ontologiat ovat muodoltaan kevyitä RDFS/OWL-standardien mukaisia ontologioita. RDF(S) ja OWL ovat W3C:n semanttisen webin esitysstandardeja. Tämä muoto voidaan mekaanisesti muuntaa W3C:n SKOS-sanastostandardin mukaiseen muotoon FinnONTO:ssa kehitetyllä Skosify-työkalulla. Muunnos ja rinnakkaisjulkaisu SKOS-muotoon voidaan tehdä automaattisesti ONKI-palvelussa ontologian päivityksen yhteydessä. Koska SKOS on nimenomaan sanastoille ja kevyille ontologioille suunniteltu standardi ja koska sen avulla voidaan nykyisin esittää olennainen KOKO-ontologioiden semantiikka, on SKOS-muodon käyttö suositeltavaa jatkossa. Formaattilla ei ole vaikutusta luetteloinnin kannalta, sillä molemmissa tapauksissa voidaan käyttää samoja URI-tunnisteita koneita varten, käsitteiden nimikkeitä ihmiskäyttöliittymissä ja käsitteiden rakenteet on hyvin pitkälle samanmuotoisia.

#### **4.2.1 Uusien käsitteiden hallinta ja virheiden korjaaminen**

Ontologiat tai asiasanastot eivät ole täydellisiä, vaan niistä löytyy tyypillisesti jatkuvasti puutteita ja virheitä. Lisäksi maailma muuttuu jatkuvasti ja synnyttää uusia käsitteitä, samalla kun toiset käsitteet saattavat vanhentua ja poistua käytöstä. Erityisen nopeita ovat muutokset esimerkiksi toimija- ja tapahtumaontologioissa uusien henkilöiden ja organisaatioiden syntymisen myötä ja nykytapahtumien muuttuessa osaksi historian aikajanaa. Seuraavassa tarkastellaan muutosten hallintaa ja virheiden korjaamista ontologityössä.

Keskeinen ontologian hallinnan ongelma on tilanne, jossa sopivaa käsitettä ontologiasta ei löydy, ja uusi käsite pitäisi saada käytettäväksi. Ongelman ratkaisulle voidaan esittää kaksi reunaehto:

- 1) Reaaliaikainen käyttöönotto omassa museossa. Tieto uudesta käsitteestä pitäisi voida ottaa *heti* käyttöön puutteeseen törmänneessä museossa, sillä muuten luettelointia ei voida tehdä halutulla tavalla.
- 2) Reaaliaikainen käyttöönotto muissa museossa. Tieto uuden käsitteen käyttöönotosta pitäisi saada levitettyä reaaliajassa myös muille museoille, sillä muuten sama käsite, esimerkiksi uusi henkilö, ”keksitään” uudelleen moneen kertaan ja samalle käsitteelle syntyy useita erilaisia tunnisteita ja esitysmuotoja, mikä johtaa semanttiseen sekaannukseen ja estää tietojen yhteentoimivuutta.

Ratkaisumalli uuden käsitteen ongelmaan on antaa museon luetteloijalle valtuudet luoda uusi käsite ehdolle kansalliseen ontologiajärjestelmään ja ONKI-palveluun. Uudelle käsitteelle vaaditaan määriteltäväksi ehdokasvaiheessa vain minimitiedot: KOKO-viitekehyksessä tarvitaan vähintään käsitteen nimike ja yläluokka (tai useampia) KOKO-ontologiassa, jotta käsite voidaan ripustaa osaksi kokonaisuutta. Tämän jälkeen käsitteelle voidaan luoda automaattisesti yksilöivä ja muuttumaton URI-tunniste ja käsite voidaan ottaa välittömästi mukaan KOKO-ontologiaan ”ehdotus”-statuksella. Tällöin se näkyy kaikille ontologian käyttäjille automaattitäydennyksen kautta ja uudella käsitteellä (URI-tunnisteella) annotointi ONKI-rajapintojen kautta on mahdollista. Käsitettä ei kuitenkaan



automaattisesti ole syytä ottaa mukaan viralliseen KOKO-struktuuriin, ennenkuin ontologiasta vastaava toimituskunta on tarkistanut ehdokkaan ja sen sijainnin ontologiassa, täydentänyt käsitteen kaikki tarvittavat semanttiset suhteen, ja määritellyt käsitteen nimikkeet eri kielillä. Tarkastukseen ja hyväksymiseen kuluva aika ei estä käsitteen käyttöä luetteloinnissa, sillä ontologioita käytettäessä luettelointi perustuu käsitteen URI-tunnisteisiin, joita tarkistusprosessissa ei muuteta. Vaikka käsitteen nimike muuttuisi prosessin tuloksena, ei virhettä luettelointitietoihin synny, sillä URI-tunniste ei muutu ja uudet nimikkeet saadaan aina käyttöön ontologiasta. Ei haittaa, jos esimerkiksi ehdotettaisiin käyttöön uutta materiaalia "petäjä", jolla käsitteellä sitten annotoitaisiin jonkin aikaa eri museoissa, ja nimikkeeksi valittaisiinkin lopulta "mänty (materiaali)". Ottamalla myös ehdotettu nimike mukaan sanastoon, mutta vaihtoehtoisena nimikkeenä, hakujärjestelmät löytävät jatkossa mäntyesineet myös hakusanalla "petäjä".

Jotta reaaliaikainen uusien käsitteiden luominen ja jakaminen kentällä olisi mahdollista, tarvitaan myös mekanismit virheiden korjaamiseen. Ensinnäkin jokin käsite tai sen määritelmän osa voi osoittautua huonoksi tai virheelliseksi, jolloin tarvitaan korjaustoimenpiteitä. Toinen virhetapaus on, kun luetteloija ehdottaa uutta käsitettä, ottaa sen itsenäisesti käyttöön, eikä huomaa sinänsä kelvollisen käsitteen käsitteen jo olevan olemassa mutta eri nimisenä. Tällaista ehdokasta ei voida hyväksyä ontologiaan. Virhe voidaan korjata myöhemmin poistamalla tuplakäsite ja 1) ylläpitämällä erillistä siltaustiedostoa samoista käsitteistä ja/tai 2) korjaamalla luettelointitiedot museoissa jälkikäteen. Synonyymisiä ilmauksia voidaan ottaa käyttöön ontologioissa valinnaisina nimikkeinä.

Edellä esitettiin ratkaisumalli virheiden korjaamiseen tilanteessa, jossa yksittäisen luetteloijan ehdokasta täsmennetään tai se osoittautuu suorastaan virheelliseksi myöhemmin. Myös ontologiasta vastaava työryhmä tekee virheitä ja käsitykset ontologian rakenteesta voivat muutenkin muuttua ajan kuluessa. Voidaan esimerkiksi huomata, että jonkin käsitteen nimeen pitäisi lisätä tarkenne, että pitäisi tehdä siitä ohjaustermi tai että käsitteen semanttisia suhteita pitäisi täydentää tai korjata. Esimerkiksi YSA:an tehdään vuosittain jopa tuhansia muutoksia, joista tiedotetaan neljännesvuosittain sähköpostitse korjauslistoilla. Tässä tapauksessa kyse on ontologian tai sanaston ylläpitotyöstä.

Keskeisiä kysymyksiä sanaston ylläpitotyössä ovat mm:

- 1) Miten tehdyt muutokset käsitteiden nimissä vaikuttavat museossa aikaisemmalla nimiversiolla kuvailtuihin tietoihin? Pitäisikö vanhat luettelointitiedot esimerkiksi korjata? Hyvä uutinen ontologioiden käyttöönotossa on, että jos luetteloinnissa käytetään pysyviä URI-tunnisteita, ei niitä tarvitse muuttaa ja ajantasaiset nimikkeet eri kielillä ovat aina automaattisesti käytettävissä ontologioiden kautta.
- 2) Virheiden korjaamiseen kuluva aika lasketaan yleensä kuukausissa eikä prosessi ole välttämättä selkeä asiakkaille: kenelle havainnot virheistä eri osaontologioiden osalta toimitetaan, onko samasta virheestä jo raportoitu ja miten virheilmoituksia sitten käsitellään? Tältä osin edellä hahmoteltu mekanismi, jolla luetteloijat voivat ehdottaa ja ottaa käyttöön yhteisöllisesti uusia käsite-ehdokkaita, jotka työryhmä sitten tarkistaa, luo pohjan avoimelle sanaston kehittämisprosessille, joka toimii toisaalta reaaliajassa ehdokkaita keräten ja julkaisten, mutta johon sisältyy myös toimituksellinen komponentti laadun varmistamista varten.
- 3) Muutoksen tekeminen yhteen KOKO:n osaan voi aiheuttaa muutospainetta toisissa osa-ontologioissa. Jos esimerkiksi YSO:ssa olevan käsitteen nimi ja semanttisia suhteita muutetaan, ja sama käsite on käytössä MAO:ssa ja merkitty siellä samaksi kuin YSO:ssa, pitäisi MAO:n kehittäjän epäilemättä tarkistaa muutoksen merkitys MAO:n osalta. FinnONTO:n KOKO-viitekehyksessä ratkaisumallia sanastojen välisten riippuvuuksien hallintaan ollaan lähestymässä kehittämällä

ongelmien hallintaan kahta työkalua: 1) MUTU-työkalulla voidaan tunnistaa ja analysoida muutopaineita, joita muuttunut yläontologia YSO aiheuttaa sitä tarkentavilla alakohtaisille ontologioille. 2) KOAN-työkalun avulla voidaan tunnistaa ontologioiden välisiä yhteisiä leikkausalueita, joita vähentämällä ontologiatyötä voitaisiin jakaa mielekkäämpiin osiin ja samojen käsitteiden erilaisia tulkintoja naapuriontologioissa selkeyttää, esimerkiksi harmonisoida saman käsitteen erilaisia jakoja alaluokkiin eri ontologioissa.

#### 4.2.2 Yhteenveto

FinnONTO:ssa saatujen kokemusten mukaan 30-70% eri alojen kotimaisesta asiasanastoista ovat päällekkäisiä YSA:n kanssa. Tämän lisäksi sanastoilla on runsaasti muita keskinäisiä yhteisiä käsitteitä. Sanastotyötä ei ole koordinoitua maassamme kuin eräillä osa-alueilla, työ on ollut projektiluonteista eikä jatkuvuutta ole. KOKO-ontologiatyön kautta museo- ja muille aloille tarjoutuu mahdollisuus koota voimia yhteen ja mahdollisuus uudenlaiseen työnjakoon. Vähäiset voimavarat voidaan kohdistaa nimenomaan oman käsitteistön kehittämiseen, jos naapurikäsitteet saadaan "ilmaiseksi" toisen alan sanaston kautta tietoja yhdistämällä. FinnONTO:n kehittämää mallia on kokeiltu hankkeen sisällä siten, että hanke on kouluttanut asiasanastojen kehittäjiä ontologiatyöhön. Nämä ovat sitten paljolti itsenäisesti kehittäneet oman ontologiamuunnoksensa tai tarjonneet asiantuntija-apua FinnONTO-hankkeen tutkijoille, milloin ontologisointi on tehty projektin henkilöstön toimesta. Tuloksena on syntynyt prototyyppi KOKO-ontologiapilvestä. Tavoitteena olevaa itsenäistä, projektin ulkoista ontologiaprosessia ei vielä ole.

## 5 SEURAAVAT ASKELEET

FinnONTO:n tulokset ovat vapaasti ja maksutta hyödynnettävissä. Artikkelin ehdotukset jatkotoimenpiteistä museotalalla voidaan kiteyttää seuraaviksi pääsuosituksiksi ajatellen ontologioita ja ontologiapalveluita:

- **ONKI-konsortion perustaminen.** FinnONTO:n piirissä on laadittu suunnitelmaehdotus "ONKI-konsortion" perustamisesta, joka ottaisi tehtäväkseen 1) ONKI-palvelun vakinaistamisen (kehittämisen ja ylläpidon) ja 2) vastuun siinä julkaistavien ontologioiden ja sanastojen kehitystyön koordinoinnista. Konsortion jäseninä olisivat sanastoja käyttävät organisaatiot. Ehdotus on konkretisoitunut siten, että Kansalliskirjaston rahoitushakemus tästä on käsiteltävänä VM:ssä ja OKM:ssä:
  - **ONKI-palveluiden vakinaistaminen.** Palvelun teknisestä toteutuksesta ja ylläpidosta vastaisi Kansalliskirjasto.
  - **Ontologioiden ja sanastojen kehitystyön koordinointi.** Mielestäni sopivin ja monialaisin taho tähän tehtävään olisi ONKI-konsortioon kuuluva Sanastokeskus TSK, joka on osallistunut FinnONTO:ssa tavalla tai toisella kaikkien eri ontologioiden kehitystyöhön.
- **KOKO-ontologiatyöryhmien perustaminen** ONKI-konsortion ontologiatyö pitäisi organisoida joukoksi alakohtaisia työryhmiä, joilla on yhteinen intressi oman alansa sanastojen kehittämiseen. Tästä on FinnONTO:n piirissä pidetty yksi laajempi järeistämistyöpaja. Ensi vaiheessa olisi hyvä käynnistää ainakin seuraavat ryhmät: 1) Kirjastostoalan ryhmä, joka kehittäisi mm. Kansalliskirjaston ontologioita (YSO, MUSO, ALLSO) ja yleisten kirjastojen ontologioita (KAUNO). Tähän olisi olemassa verkosto ja valmiudet Kansalliskirjaston nykyisen

YSA-työryhmän ja yleisten kirjastojen Kirjasto.fi-yhteisön pohjalta. Sopivin taho vetovastuuseen olisi Kansalliskirjasto. 2) Museoalan työryhmä, joka kehittäisi museo-, taideteollisuus- ja valokuvausalojen erityisontologioita, kuten MAO, TAO ja VALO. Tämä työ olisi luontevaa tehdä FinnONTO:ssa pitkään vaikuttaneen Museoviraston johdolla ja käynnistää Museo 2015-hankkeen yhteydessä. 3) Terveysala, jonka vastuulla olisivat TERO ontologian eri osaontologiat. Tällä sektorilla FinnONTO:n kokoonpanoon vuosikausia kuulunut Terveystieteiden ja hyvinvoinnin laitos THL olisi paras taho vastuuta kantamaan osana THL:n vastuulla nykyään jo olevia ”koodistopalveluita”. 4) Luonnontieteellisiä nimistöjä ylläpitäisi Luonnontieteellinen keskusmuseo yhteistyössä mm. alan järjestöjen kanssa (kuten BirdLife ja Vanamo). Muita alueita ja ontologioita tulisi mukaan kiinnostuksen ja resurssien mukaan – ovet ONKI-konsortiossa olisivat kaikille halukkaille aina auki.

- **Museoalan paikkaontologioiden kehittäminen.** Historiallisten paikkojen kerääminen ontologiaksi on syytä käynnistää FinnONTO:n SAPO:n pohjalta. Tämä kannattaa tehdä saman museoalalle perustettavan ontologiatyöryhmän toimesta. Nykyisten paikkojen ontologian osalta voidaan käyttää Maanmittauslaitoksen toukokuussa avoimena tietona julkaistavia paikkarekistereitä ja vastaavia ulkomaisia lähteitä. Siltaus LOD-pilven paikkaontologiaan Geonames olisi tärkeä muodostaa. FinnONTO:ssa on kehitetty prototyyppi laajasta paikkaontologiasta ja -palvelusta ”ONKI Geo” (Lindroos, 2008), joka sisältää miljoonia nykypaikkoja koordinaattitietoinen, mukaan lukien nyt avattava Maanmittauslaitoksen Paikannimirekisteri.
- **Museoalan toimijaontologian TOIMO kehittäminen.** Historiallisten toimijoiden kerääminen yhteisesti jaetuksi ontologiaksi on syytä käynnistää Gettyn säätien ULAN-rekisterin tapaan. Toimijarekisteri yhdistyy mm. historiallisiin paikkoihin ja tapahtumiin. Linked Data Finland –hankkeessa toimijoita varten luodaan erillinen SAHA-HAKO-palvelu, jota voidaan ylläpitää yhteisöllisesti vastaavalla tavalla kuin Kirjasammon sisältöä luotaessa. Toimijaontologiasta ja -palvelusta on FinnONTO:ssa tehty ensimmäinen prototyyppi ”ONKI People” (Kurki, Hyvönen, 2009; Kurki, 2011) aineistona osa Getty säätien ULAN-rekisteriä. SAHA:n sisällöt voidaan ottaa käyttöön ONKI-rajapinnan kautta samaan tapaan kuin nykyisessä ONKI-palvelussa. Vielä ei museoiden luettelointijärjestelmiä ole SAHA:an kytketty. Toimijaontologian kehittäminen on haaste, mutta liikelle kannattaa lähteä heti pienin askelin olemassa olevia aineistoja hyödyntäen. Museoalan historialliset toimijat kannattaa pidemmällä tähtäimellä sillata mm. Kansalliskirjaston auktoriteettitietokantaan. DBPedian (Wikipedian RDF-muunnos) henkilöiden kautta rakentuu yhteys LOD-aineistoihin. Saksan, Ranskan ja Ruotsin kansalliskirjastot ovat jo julkaisseet auktoriteettitietojaan avoimena yhdistettynä tietona. Aineistoja riittää; ongelmana on pikemminkin niiden runsaus kuin niukkuus.
- **Museoalan aikaontologian AIKO kehittäminen.** Aikaontologia on tärkeä kehityskohde, muttei niin kriittinen, iso ja suuritöinen kuin paikka- ja toimijaontologiat.
- **Museoalan aikaontologian HISTO kehittäminen.** Tapahtumaontologiat ovat tutkimuksellisesti kiinnostava kehityskohde, josta FinnONTO:ssa on jo saatu lupaavia tuloksia. Työ jatkuu SeCo-ryhmän osalta Linked Data Finland –hankkeessa ja erillisessä sotahistoriaan keskittyvässä kansainvälisessä tutkimushankkeessa.
- **Tekninen tutkimus- ja kehitystyö** Tutkimusta tarvitaan tukemaan käytännön työtä, sillä web-teknologiat kehittyvät nopeasti. Yhdistämällä työhön liiketoiminnallinen näkökulma, esimerkiksi matkailu, voidaan kulttuurialalle saada merkittäviä määriä teknologian kehittämisen rahoitusta, kuten FinnONTO on osoittanut. Kiinnostavia tutkimusalueita ovat mm.

Linked Data Finland -jatkohankkeessa tutkittava automaattinen annotointi, avoimen tiedon laadunvarmistus, metatietomallit, semanttinen suosittelu ja tietämyksen louhinta (knowledge extraction and discovery) sekä tiedon visualisointi. Uusia semanttisen laskennan menetelmiä voidaan soveltaa Digital Humanities -alueella humanististen tutkimusongelmien ratkaisemiseen.

Museoissa ja kirjastoissa ollaan siirtymässä uuden semanttisen web-tekniikan käyttöön. Ei ole epäilystä siitä, etteikö tulevaisuus olisi ontologioiden. Semantiikaltaan epätasaisemmat tai yksinkertaisemmat tietämysrakenteet (knowledge organization system, KOS), kuten luokitukset ja asiasanastot, voidaan ottaa käyttöön vaivattomasti yksinkertaisina ontologioina W3C:n semanttisen webin RDF-tietomalliin perustuvan SKOS (Simple KOS) standardin avulla<sup>27</sup>. Ontologisoinnin kautta rakenteita voidaan rikastaa vähitellen, kuten on tehty FinnONTO:ssa suomalaisille asiasanastoille. Tarvittavan infrastruktuurin rakentaminen, tekniikan omaksuminen, uusien mahdollisuuksien ymmärtäminen, sisällöntuotannon käytäntöjen muuttaminen ja uusien työkalujen käyttöönotto vaatii kuitenkin aikansa. Muutoksessa luettelointijärjestelmät ovat keskeisessä asemassa ja Museo2015-hanke on tässä mielessä kriittisen tärkeässä asemassa jatkoa ajatellen.

FinnONTO:n keskeinen sanoma museoalalle on ollut vuodesta 2003 saakka (Hyvönen, 2010):

*Museoiden sisällöntuotanto voidaan saada toiminaan yhteentoimivasti yhteisten ontologioiden avulla. Yhteisen ontologiaperustaisen tietoinfrastruktuurin avulla monet julkaisujärjestelmien ongelmat ratkeavat itsestään, ja kulttuurialan verkkosovellusten kehittäminen on huomattavasti helpompaa ja kustannustehokkaampaa. Tuloksena syntyy aiempaa älykkäämpiä ja käytettävämpiä monikielisiä museojärjestelmiä ei vain yksittäisille museoille, vaan MuseoSuomen, Kulttuurisammon ja KDK:n kaltaisille kansallisille järjestelmille. Semanttisen webin (Linked Data) teknologia mahdollistaa tämän Suomessa pilotoidun julkaisukonseptin laajentamisen kieli- ja kulttuurirajojen yli myös kansainvälisesti Euroopan kaltaisiin hankkeisiin.*

Kokemustemme mukaan keskeisin haaste tässä on yhteentoimivan semanttisen metatiedon tuottaminen kustannustehokkaasti **jo tiedon tuotantovaiheessa museoissa**, missä ontologiaperustaiset luettelointijärjestelmät ovat ratkaisevassa asemassa. Ongelmien ratkominen jälkikäteen sovelluksia kehitettäessä on paljon hankalampaa ja kalliimpi ratkaisu. Albert Einsteinin ajatus pätee tässäkin asiassa:

“Intellectuals solve problems, geniuses prevent them.”

## KIITOKSET

Tässä artikkelissa esitelty työ on osa FinnONTO-hanketta ja perustuu kymmenien Semanttisen laskennan tutkimusryhmässä työtä eri aikoina tehneiden tutkijoiden tuloksiin. Hanketta ovat rahoittaneet ja ohjanneet Tekes ja n. puolensadan suomalaisen yrityksen ja julkisen sektorin organisaation konsortio. Kulttuurisammon kehittämiseen on lisäksi saatu Suomen Kulttuurirahaston tukea.

<sup>27</sup> <http://www.w3.org/2004/02/skos/>

# VIITTEET

CSC: Museoiden kokoelmanhallintajärjestelmät. Nykytila ja vaihtoehtoiset kehitysmallit. Selvitystyön loppuraportti, CSC Oy, 16.2.2011.

<http://www.nba.fi/fi/File/1127/museoiden-kokoelmahallintajarjestelmat.pdf>

Heath T., Bizer C.: Linked Data: Evolving the Web into a Global Data Space. Claypool & Morgan, CA, 2011.

Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila and Suvi Kettula: MuseumFinland - Finnish Museums on the Semantic Web. Journal of Web Semantics, vol. 3, no. 2, pp. 25, 2005.

<http://www.seco.tkk.fi/publications/2005/hyvonen-makela-et-al-museumfinland-finnish-2005.pdf>

Eero Hyvönen: Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita?

<http://www.seco.tkk.fi/publications/2005/hyvonen-miksi-asiasanastot-eivat-riita-2005.pdf>

Eero Hyvönen: FinnONTO-malli kansallisen semanttisen webin sisältöinfrastruktuurin perustaksi - visio ja sen toteutus. Teknillinen korkeakoulu, Espoo, 12.9.2008a.

<http://www.seco.tkk.fi/publications/2008/hyvonen-ONKI-yleisesitys-2008.pdf>

Eero Hyvönen: Kulttuurisampo - suomalainen kulttuuri semanttisessa webissä. Muistiorganisaatioiden ja kansalaisten yhteisöllinen kansallinen julkaisujärjestelmä. Teknillinen korkeakoulu, Espoo, 25.9.2008b.

<http://www.seco.tkk.fi/publications/2008/hyvonen-Kulttuurisampo-2008.pdf>

Eero Hyvönen: Preventing Interoperability Problems Instead of Solving Them. Semantic Web Journal, IOS Press, no. 2-3, December, 2010. <http://www.semantic-web-journal.net/content/new-submission-preventing-interoperability-problems-instead-solving-them>

Eero Hyvönen, Kim Viljanen, Jouni Tuominen and Katri Seppälä: Building a National Semantic Web Ontology and Ontology Service Infrastructure--The FinnONTO Approach. Proceedings of the European Semantic Web Conference ESWC 2008, Springer-Verlag, 2008.

<http://www.seco.tkk.fi/publications/2008/hyvonen-et-al-building-2008.pdf>

Eero Hyvönen, Jouni Tuominen, Tomi Kauppinen, Jari Väätäinen: Representing and Utilizing Changing Historical Places as an Ontology Time Series. Geospatial Semantics and Semantic Web: Foundations, Algorithms, and Applications (Naveen Ashish and Amit Sheth (eds.)), Springer-Verlag, 2011.

<http://www.seco.tkk.fi/publications/2011/hyvonen-et-al-sapo-2011.pdf>

KDK: Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri, v 1.0. Opetus- ja kulttuuriministeriö, 29.8.2011. [http://www.kdk.fi/images/stories/KDK\\_kokonaisarkkitehtuuri.pdf](http://www.kdk.fi/images/stories/KDK_kokonaisarkkitehtuuri.pdf)

KDK: Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri. Liite B: Standardisalkku. Opetus- ja kulttuuriministeriö, 29.8.2011.

<http://www.kdk.fi/images/stories/tiedostot/kdk%20standardisalkku%202011-08-29.pdf>

Helka Ketonen (toim.) Taideteosten luettelointiohje, Museotyöntekijän käsikirja 6. Valtion taidemuseo, 2009.

Suvi Kettula: Semanttisen webin ontologisen tekstiilikäsitteistön kehittäminen ja liittäminen museon luettelointijärjestelmiin. Väitöskirja, Helsingin yliopisto, käsityötieteen laitos, 2009.

<http://www.doria.fi/handle/10024/50313>

Jussi Kurki: Toimijaontologia ja niiden käyttö semanttisessa webissä. Pro Gradu, Helsingin yliopisto, 2011. <http://www.seco.tkk.fi/publications/2011/kurki-gradu-2011.pdf>

Jussi Kurki and Eero Hyvönen: Authority Control of People and Organizations on the Semantic Web. Proceedings of the International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD2009), September, 2009. Trento, Italy.

<http://www.seco.tkk.fi/publications/2009/kurki-hyvonen-onki-people-2009.pdf>

Robin Lindroos: Paikkatiedon ontologiapalvelu. Diplomityö, Teknillinen korkeakoulu, 2008.

<http://www.seco.tkk.fi/publications/2008/onkipaikka-diplomityo-2008.pdf>

Reija Lång, Sirkka Valanto: Museoalan asiasanastot uudistaminen. Esiselvitys. Museoalan asiasanastn (MASA) uudistamishankkeen ensimmäisen vaiheen loppuraportti 31.12.2008. Museovirasto, 2008.

<http://www.nba.fi/fi/File/775/masa-uudistaminen.pdf>

Eetu Mäkelä, Eero Hyvönen and Tuukka Ruotsalo: How to deal with massively heterogeneous cultural heritage data – lessons learned in CultureSampo. Semantic Web Journal, IOS Press, 2012.

<http://www.semantic-web-journal.net/content/how-deal-massively-heterogeneous-cultural-heritage-data-%E2%80%93-lessons-learned-culturesampo>

Eetu Mäkelä, Kaisa Hypén and Eero Hyvönen: BookSampo--Lessons Learned in Creating a Semantic Portal for Fiction Literature. Proceedings of ISWC-2011, Bonn, Germany, Springer-Verlag, 2011.

<http://www.seco.tkk.fi/publications/2012/tuominen-et-al-ontologiapalvelut-2012.pdf>

Jouni Tuominen, Matias Frosterus, Kim Viljanen and Eero Hyvönen: ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services. Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Heraklion, Greece, May 31 - June 4, 2009. Springer-Verlag.

<http://www.seco.tkk.fi/publications/2009/tuominen-et-al-onki-skos-2009.pdf>

Jouni Tuominen, Kim Viljanen and Eero Hyvönen: Ontologiapalvelut semanttisessa webissä, Tietojenkäsittelytiede, Tietojenkäsittelytieteen Seura ry, 2012.

<http://www.seco.tkk.fi/publications/2012/tuominen-et-al-ontologiapalvelut-2012.pdf>

Kim Viljanen, Jouni Tuominen and Eero Hyvönen: Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI. Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Heraklion, Greece, May 31 - June 4, 2009. Springer-Verlag.

<http://www.seco.tkk.fi/publications/2009/viljanen-et-al-www.yso.fi-2009.pdf>