



Versio 095  
2492008

## Kulttuurisampo – suomalainen kulttuuri semanttisessa webissä

### Muistiorganisaatioiden ja kansalaisten yhteisöllinen kansallinen julkaisujärjestelmä

#### Eero Hyvönen

Professori, tutkimusjohtaja  
Teknillinen korkeakoulu, viestintäteknikan laitos ja  
Helsingin yliopisto, tietojenkäsittelytieteen laitos  
Semanttisen laskennan tutkimusryhmä  
<http://www.seco.tkk.fi/>  
eero.hyvonen[at]tkk.fi

*Raportti esittelee vision, mallin ja sen toteutuksen kansallisesta museoiden, kirjastojen, arkistojen ja muiden organisaatioiden kulttuurisisältöjen yhteisöllisestä julkaisukanavasta Kulttuurisammosta, joka ei unohda myöskään kansalaisia sisällöntuottajina. Kulttuurisampo-järjestelmä koostuu kolmesta osasta: 1) Kulttuurisammon asiakasliittymä on semanttiseen webiin perustuva portaali, joka yhdistelee sisällöllisesti toisiinsa n. 20 eri museon, kirjaston ja arkiston sisältöjä sekä kansalaisten mm. Wikipediaan tuottamia verkkoaineistoja. Portaali tarjoaa loppukäyttäjille uudenlaisia älykkäitä käyttöliittymiä, temaattisia näkökulmia ja visualisointeja tutustua kokoelmiin ja aineistoihin. 2) Sisällöntuotannon rajapintojen kautta museot, kirjastot, arkistot ja muut kulttuurisisältöjen tuottajaorganisaatiot ja kansalaiset, voivat julkaista yhdessä sisältöjään modernilla tavalla semanttisessa webissä. 3) Järjestelmä perustuu W3C:n suositukseen ja kansalliseen semanttisen webin FinnONTO-sisältöinfrastruktuuriin, joka tarjoaa kulttuuriorganisaatioille sanasto- ja ontologiapalvelut ja työvälineitä kustannustehokkaaseen sisällöntuotantoon. Kulttuurisampo-järjestelmä on kehitetty ja toteutettu v. 2004-2008 Teknillisessä korkeakoulussa ja Helsingin yliopistossa laajan kansallisen konsortion tuella. Järjestelmä on semanttisessa rikkaudessaan ja yhteisöllisyydessään ensimmäinen laatuaan maailmassa.*

<b>1</b>	<b>Muistiorganisaatiot ja semanttinen web</b>	<b>3</b>
<b>2</b>	<b>Kohti muistiorganisaatioiden yhteisportaalia</b>	<b>5</b>
2.1	Suunnitelmia ja ensimmäiset toteutukset	5
2.2	Kulttuurisampo yhdistää aineistot ja toimijat semanttisessa webissä	5
2.3	Miten Kulttuurisampo taottiin	8
<b>3</b>	<b>Kulttuurisampo-järjestelmän visio, malli ja sen toteutus</b>	<b>11</b>
3.1	Yhteisen julkaisemisen haasteet	11
3.2	Kulttuurisammon ratkaisumalli	13
<b>4</b>	<b>Semanttinen sisällöntuotantoprosessi</b>	<b>17</b>
<b>5</b>	<b>Sisältöinfrastruktuuri: ontologiat ja ontologiapalvelut</b>	<b>20</b>
5.1	FinnONTO-ontologiatyön periaatteita	20
5.2	Ontologiat	21
5.3	Kansallinen ontologiakirjastopalvelu ONKI	26
<b>6</b>	<b>Kulttuurisampo portaali ihmisille ja koneille</b>	<b>32</b>
6.1	Asiakaskäyttöliittymän perustoiminnot: haku ja selailu	32
6.2	Temaattiset näkökulmat	33
6.3	Portaalin käyttö ulkopuolelta palveluina: leijukkeet	40
<b>7</b>	<b>Yhteenveto: miten tästä eteenpäin?</b>	<b>42</b>
7.1	Sisällöntuotannon mahdollisuudet ja haasteet	42
7.2	Kansallinen digitaalinen kirjasto ja Kulttuurisampo	45
7.3	Miten tästä eteenpäin?	46
	<b>Kiitokset</b>	<b>47</b>
	<b>Kirjallisuutta</b>	<b>48</b>

## 1 Muistiorganisaatiot ja semanttinen web

---

Web tarjoaa yhä enemmän käytetyn kanavan kulttuurisisältöihin tutustumista ja tiedonhankintaa varten. Museokäynnin fyysistä kokemusta näyttöpäätte ei korvaa, mutta mahdollistaa tätä täydentäviä tapoja kulttuurisisältöihin tutustumiselle:

- **Ajasta ja paikasta riippumaton tiedonvälitys.** Kokoelmiin tutustuminen ei edellytä museorakennukseen tulemistä tiettyyn aukioloaikaan. Web on auki aina ja joka paikassa. Mobiilisovellusten kautta sisältöjä voidaan hakea luontevasti sen hetkisen kontekstin perusteella, kuten lähellä olevia kohteita Applen iPhone-puhelimella.
- **Tietosisällön määrän lisääminen.** Laajat kokoelmat, joiden fyysinen näytteille asettaminen olisi mahdotonta, voidaan avata yleisölle. Sisältöjä voidaan kerätä eri lähteistä, organisaatioista ja yli maiden rajojen, kuten esimerkiksi MichaelPlus<sup>1</sup> tai Europeana-hankkeissa<sup>2</sup>.
- **Tietosisällön laadun parantaminen.** Tietotekniikka mahdollistaa tietojen haun, yhdistämisen ja esittämisen tavoilla, jotka eivät ole teknisesti tai taloudellisesti mahdollisia fyysisissä näyttelytiloissa. Tällainen on esimerkiksi kohteiden haku ja visualisointi kartoilla (Google Maps) tai aikajanalle, tai reitityspalvelut luonnossa oleviin kohteisiin.
- **Personointi.** Tietojärjestelmien vuorovaikutteisuus (interactivity) ja sopeutuvuus (adaptivity, learning) mahdollistavat aiempaa joustavampia ja henkilökohtaisempia tapoja tutustua aineistoihin (Wang et al., 2008).
- **Audiovisualisuuden hyödyntäminen.** Tietotekniikka mahdollistaa sisältöjen esittelemisen audiovisuaalisilla keinoilla, kuten kuva-, musiikki- ja videotallenteilla.
- **Uudet kokoelmakohteet.** Tietotekniikka mahdollistaa uudentyyppisten aineistojen kuten kulttuuristen prosessien ja tarinoiden dokumentointia, luettelointia ja tallentamista kokoelmiin osana kulttuuriperimäämme (Hyvönen et al., 2007; Junnila et al., 2007).
- **Sosiaalinen media.** Web 2.0 tekniikoilla myös asiakkaat voidaan ottaa mukaan sisällöntuotantoon, kuten esimerkiksi Steve Museumissa<sup>3</sup>.

Web-sovellusten kehittämisessä museoyleisölle on runsaasti tietoteknisiä haasteita. Avainkysymys on, miten web-sivustoista saadaan kävijöiden kannalta aidosti mielenkiintoisia, hyödyllisiä ja helppokäyttöisiä. Tähän ei pelkkä perinteinen kokoelmatietojen hakumahdollisuus tietokannasta välttämättä riitä – ainakaan silloin ei hyödynnetä kuin pieni osa niistä mahdollisuuksista, joita kokoelmien rikas sisältö voisi asiakkaille tarjota *semanttisina portaaleina* (semantic portal) (Hyvönen, 2008b). Sanalla

---

<sup>1</sup> <http://www.michael-culture.org>

<sup>2</sup> <http://www.europeana.eu/>

<sup>3</sup> <http://www.steve.museum/>

”semanttinen” viitataan tässä yhteydessä webin kehityssuuntaan, jossa verkkosisällöt esitetään paitsi ihmisten myös koneiden ymmärtämässä muodossa. Tällöin tulee mahdolliseksi erilaisten tietosisältöjen sisällöllinen yhdistely sekä aiempaa olennaisesti älykkäämpien haku- ja selailujärjestelmien toteuttaminen.

Kulttuurisisältöjen verkkojulkaisemisen lähtökohtana pitäisi olla, että webin kautta saatava museokokemus syntyy - kuten fyysisissä näyttelyissäkin - kokoelmaesineiden ohella näiden välisistä kiinnostavista suhteista ja tietojen avulla luoduista temaattisista kokonaisuuksista. Webin linkitysmekanismi tarjoaa tähän periaatteellisen mahdollisuuden. Esimerkiksi Akseli Gallen-Kallelan taulua ”Aino-triptyyksi” (kuva 1.1) esittelevältä web-sivulta voidaan luoda linkkejä eri kokoelmien ja muistiorganisaatioiden aineistoihin, kuten Kalevalan siihen tekstikohtaan, jota taulu kuvaa, eri taidemuseoiden muihin samaa teemaa kuvaaviin tauluihin, kulttuurihistoriallisten museoiden kokoelmissa oleviin esineisiin, joita taulussa on kuvattuna, Kansallisbiografiassa olevaan Akseli Gallen-Kallelan elämäkertaan, Wikipediassa oleviin sivuihin Aino-tarusta ja taiteilijasta, kirjastojen arkistoissa oleviin tutkimuksiin Kalevalan Aino- ja Väinämöinen-hahmoista, muihin Pariin maailmannäyttelyssä samaan aikaan palkittuihin teoksiin jne.



**Kuva 1.1.** Akseli Gallen-Kallelan Aino triptyyksi, Pariisi, 1891 (Ateneumin taidemuseo).

Temaattisia virtuaalinäyttelyitä on tehty käsityönä sekä CD-levytykseen että webiin. Tämä onnistuu nykyisin pienille aineistoille kohtuullisilla kustannuksilla. Kokoelmien verkkojulkaisemisessa on kuitenkin kysymys kokoelmatietokantoihin perustuvista laajoista haku- ja selailujärjestelmistä, jossa voi olla kymmeniä tuhansia tai jopa miljoonia erityyppisiä kohteita eri organisaatioiden kokoelmatietokannoissa olevia kohteita. Esimerkiksi suomalaisissa museoissa, joita on nykyisin yli 1000, on tällä hetkellä tallennettuna runsaasti kulttuuri- ja luonnonperintöä<sup>4</sup>: esineitä on yli 4 miljoonaa, taideteoksia lähes 300 000, kuvakokoelmien laajuus on 19 miljoonaa ja luonnontieteellisiä näytteiden laajuus 19 miljoonaa kohdetta.

Lisäksi kokoelmat karttavat vuosi vuodelta. Käsintehdä kiinteä teemoitus tai linkitys ei aineistojen laajuuden, monimuotoisuuden ja muuttumisen vuoksi ole käytännössä mahdollista vaan tähän tarvitaan tietokoneen apua. Tietokone on hyvä väline myös silloin, kun samasta aineistosta halutaan automaattisesti tuottaa erilaisia näyttelyitä.

<sup>4</sup> <http://www.museot.fi/museo-kokoelmat>

## 2 Kohti muistiorganisaatioiden yhteisportaalia

---

### 2.1 Suunnitelmia ja ensimmäiset toteutukset

Suomessa muistiorganisaatioiden digitointiyhteistyö on rakentunut kirjastojen, arkistojen ja museoiden yhteistyölle, mm. eKam-hankkeessa rakennetulle pohjalle<sup>5</sup>: ”Opetusministeriö asetti 11.1.2000 Kulttuuriperinnön digitointityöryhmän (KULDI) osana valtioneuvoston Sisältötuotanto-hanketta. Työryhmässä ovat olleet edustettuina museot, arkistot ja tieteelliset kirjastot. Työryhmän tehtävänä on ollut edistää kulttuuriperintöaineistojen digitointia, digitoitujen aineistojen saatavuutta tietoverkon kautta ja tätä aineistoa hyödyntävän sisältötuotannon syntymistä.” Kirjastojen, arkistojen ja museoiden tietovarantojen yhteiskäyttömahdollisuuksia selvittämään perustettiin v. 2003 erillinen Kamut-2-ryhmä, joka julkisti loppuraportin ”Kamut 2 Muistiorganisaatioiden tietovarannot yhteiskäyttöön”<sup>6</sup> (Vakkari et al., 2004). Samaan aikaan v. 2002-2004 kehitettiin Tietotekniikan tutkimuslaitos HIIT:ssä ja Helsingin yliopistossa maailmankin mittakaavassa ensimmäinen semanttiseen webiin perustuva, eri organisaatioiden sisältöjä yhdistelevä MuseoSuomi – Suomen museot semanttisessa webissä<sup>7</sup> (Hyvönen et al., 2005) (sisällöntuottajina Espoon kaupunginmuseo, Museovirasto ja Lahden kaupunginmuseo / Antikvaria-ryhmä) ja vuosien 2003-2004 vaiheessa Museoviraston omana hankkeena Suomen museot online –portaali<sup>8</sup> (SMOL). Molemmat järjestelmät julkistettiin webissä samanaikaisesti Kansallismuseossa pidetyssä julkistustilaisuudessa 8.3.2004<sup>9</sup>.

Semanttisen webin mahdollisuuksien arviointi on vahvasti esillä Kamut-2-raportissa. Ontologiatekniikkaa on otettu käyttöön museoalan maailmanjärjestössä International Council of Museums (ICOM), sen sisältöjen dokumentointiin keskittyvää CIDOC-komiteassa. Komitean piirissä kehitetty CIDOC Conceptual Reference Model (CRM) on standardoitu vuonna 2006 ISO-standardiksi-ISO 21127. Semanttisen webin tutkimuksen puolella kulttuurisisällöistä on muodostunut yksi keskeinen sovellusalue. Aihepiirin sovelluksia mm. on palkittu useana vuotena alan kansainvälisellä Semantic Web Challenge –palkinnolla: MuseoSuomi 2004 (Hyvönen et al., 2005); MutimedialN Demonstrator 2006 (Schreiber et al., 2006); CHIP Demonstrator 2007 (Wang et al., 2008). Yleiseurooppalaisesta European Digital Library –portaalista, jonka nimi muutettiin sisältöneutraalimmaksi Europeanaksi museoiden, arkistoiden ja muiden sisällöntuottajien toivomuksesta, kehitetään tiettävästi jatkossa sekä ”semanttinen” että ”perinteinen” versio.

### 2.2 Kulttuurisampo yhdistää aineistot ja toimijat semanttisessa webissä

Suomen museot online –järjestelmään on vuoden 2004 jälkeen liittynyt uusia museoita. Palvelun web-sivujen mukaan systeemissä on tällä hetkellä n. 15 000 esineen ja kuvan metatiedot. Aineisto on peräisin Museoviraston omista museoista, Tampereen museoista

<sup>5</sup> [http://www.nba.fi/fi/eKAM\\_yhteistyö](http://www.nba.fi/fi/eKAM_yhteistyö)

<sup>6</sup> [http://www.nba.fi/fi/kamut2\\_tiedote](http://www.nba.fi/fi/kamut2_tiedote)

<sup>7</sup> <http://www.museosuomi.fi/>

<sup>8</sup> <http://www.suomenmuseotonline.fi/>

<sup>9</sup> <http://www.seco.tkk.fi/events/2004/2004-03-08-kansallismuseo/>

sekä eräistä muista museoista ja uusia sopimuksia ollaan tekemässä. Itse järjestelmään ei ole tullut merkittäviä teknisiä muutoksia.

Samaan aikaan on jatkunut MuseoSuomeen ja semanttisen webin lähestymistapaan perustuva kehitystyö Tekesin ja 38 muun suomalaisen organisaation rahoittamassa kansallisessa Suomalaiset semanttisen webin ontologiat –hankkeessa FinnONTO 2003-2008, 2009-2010. MuseoSuomi toimii edelleen palveluna verkossa ja sitä on käytetty vilkkaasti, parhaimmillaan n. 10 000 uniikkia eri käyttäjää kuukaudessa. Järjestelmää ei ole v. 2004 julkistuksen jälkeen kehitetty teknisesti eikä portaaliin ole tuotu uusia aineistoja. Sen sijaan tutkimus- ja kehitystyö suunnattiin MuseoSuomen seuraajaan, tässä artikkelissa kuvattavaan Kulttuurisampo-järjestelmään (Hyvönen et al., 2008), joka laajentaa MuseoSuomen ideoita uusille tasoille.

Kulttuurisampo liittyy myös Suomen museot online –järjestelmään (SMOL) siten, että SMOL:n sisältöjä voidaan siirtää Kulttuurisampoon puoliautomaattisella muuntimella. Tällöin SMOL:ssa määritellyn XML-skeeman<sup>10</sup> muotoinen tieto muutetaan semanttisen webin metatietomallin RDF<sup>11</sup> mukaiseksi ja metatiedoissa olevien kenttien merkkijonoarvoiset arvot (litteraalit) yksilöidään merkityksiltään yksikäsitteiksi URI-viittauksiksi Kulttuurisammon ontologioihin (ns. aineistojen ripustus ontologioihin). Vastaavanlainen muunnin on kehitetty suomalaisissa taidemuseoissa käytettyyn Muusajärjestelmään, jossa käytetään CIDOC CRM –standardia - Valtion taidemuseossa kehitetty Muusa on ollut ensimmäisiä CIDOC CRM:n soveltajia maailmassa. Koska SMOL ja Muusan CIDOC CRM eivät ainakaan vielä käytä ontologioita sisältöjen kuvailussa, ei muunnos voi olla täysin automaattinen, vaan käsityötä tarvitaan. Esimerkiksi merkkijonosta ”Nokia” ei kone voi välttämättä tietää, viittaako se yritykseen vain kaupunkiin, tai viittaako ”villa” materiaaliin vai rakennukseen. Termien ontologiseen ripustamiseen liittyen on valmistumassa opinnäytetyö aineistona MuseoSuomen tekstiilit (Kettula, 2008).

MuseoSuomi ja Kulttuurisampo-hankkeen yksi keskeisimpiä opetuksia on ollut: luettelointikäytäntöjä kannattaisi harmonisoida siten, että niissä käytettyjen ilmausten merkitykset saadaan yksikäsitteiksi tietokoneiden, ei vain ihmisluettelointisijoiden kannalta. Tarve asiasanoituksen harmonisointiin on toki tiedostettu informaatiotutkimuksen piirissä jo pitkään (Aitchison et al., 2000), mutta semanttinen web nostaa ongelmatiikan uudelle tasolle vaatiessaan sisällönkuvailulta konetulkittavuutta (Hyvönen, 2005). Esimerkiksi Kulttuurisampotyössä turhauttavan suuri määrä voimavaroja on kulunut aineistojen muokkaamiseen ja ripustamiseen ontologioihin; koneymmärrettävän metatiedon tuottaminen on semanttisen webin ydinkysymyksiä käytännön tasolla. Semanttisen webin suositeltava ratkaisumalli on käyttää sisällönkuvailussa ontologioita (Staab, Studer, 2008). Tällöin aineistojen julkaiseminen Kulttuurisammon kaltaisessa ”julkaisukoneessa” voisi onnistu ”nappia painamalla” ilman työlästä ja kallista käsityötä.

<sup>10</sup> <http://www.nba.fi/tiedostot/dd709cdb.pdf>

<sup>11</sup> <http://www.w3.org/RDF/>

Luettelointikäytäntöjen harmonisointi *semanttisella* tasolla kannattaisi ottaa jatkossa selkeäksi kansalliseksi tavoitteeksi. Tarvitaan yhteiset pelisäännöt sille, millaisia metadataskeemoja eri aineistotyypeille pitäisi käyttää ja mitä sisällönkuvailun sanastoja ja ontologioita käytetään kenttien arvoina. SMOL:n XML-skeema on tärkeä askel tähän suuntaan, mutta askel ei ole riittävän pitkä yltääkseen semanttiseen webiin saakka. Skeema määrittelee vain yhden mallin, joka ei kata kaikkien eri aineistotyyppien tarpeita, eikä määritä metatietoelementtien arvoalueita riittävän täsmällisesti, jotta tiedot voitaisiin ripustaa automaattisesti ontologioihin tai validoida semanttisesti.

Kulttuurisampo-hankkeen tärkeänä tavoitteena on ollut tutkia, miten muodoltaan ja luonteeltaan hyvin erilaiset sisällöt, joita on tuotettu hajautetusti museoissa, kirjastoissa, arkistoissa ja muissa kulttuurialan organisaatioissa, voidaan yhdistää kansallisella tasolla toimivaksi älykkääksi haku- ja suosittelupalveluksi semanttisessa webissä. Järjestelmä perustuu FinnONTO-hankkeessa niin ikään valmistuneeseen prototyyppiin Kansallisesta semanttisen webin infrastruktuurista (Hyvönen, 2008ac; Hyvönen et al, 2008a), joka koostuu kymmenistä tuhansista toisiinsa liitetystä käsitteistä, ja näiden hyödyntämisen välineistä, kuten Kansallisesta ontologiapalvelusta ONKI<sup>12</sup>. Tämä ontologia- ja ohjelmistokokonaisuus on kehitetty open source –lisenssillä eikä sen käyttäminen edellytä lisenssimaksuja. Kulttuurisampo on toteutettu hyödyntäen perinteisiä hakukoneita semanttisessa haussa, joten järjestelmä skaalautuu paljon laajempien aineistojen samanaikaiseen käsittelyyn kuin MuseoSuomessa käytetty muistinvarainen arkkitehtuuri.

Kulttuurisammon ensimmäinen julkinen pilottiversio sisältää n. 20 suomalaisen museon, kirjaston, arkiston ja muun organisaation kokoelmia sekä muita aineistoja. Suomalaisia kokoelmakohteita järjestelmässä on yli 50 000 aineistotyyppinä mm. esineet, valokuvat, kartat, maalaukset, runot, kirjat, kansansävelet, videot, taitokuvaukset sekä tarinat (Kalevala). Lisäksi järjestelmä sisältää mm. Getty-säätiön n. 120 000 taiteilijan rekisterin ja siihen on yhdistetty verkkoaineistoja Agricola - Suomen historiaverkosta, artikkeleita Wikipediasta ja valokuvia Panoramio-palvelusta<sup>13</sup>.

Pilottiversion kotimaisina sisällöntuottajina ovat olleet (aakkosjärjestyksessä): Antikvaria-ryhmä (Lahden kaupungin museo, Kuopion kulttuurihistoriallinen museo, Pohjois-Karjalan museo), Espoon kaupunginmuseo, Geologian tutkimuskeskus GTK, Helsingin kaupunginkirjasto, Kansalliskirjasto, Kansallismuseo ja muita Museoviraston museoita, Kirjastot.fi, Maanmittauslaitos, Suomalaisen Kirjallisuuden Seura SKS, Suomen maatalousmuseo Sarka, Suomen valokuvataiteen museo, Taideteollisen korkeakoulun kirjasto, Valtion taidemuseo ja Veljekset Karhumäki Oy. Järjestelmä on kehitetty Teknillisen korkeakoulun ja Helsingin yliopiston Semanttisen laskennan tutkimusryhmässä. Lista osoittaa, että kulttuuriportaalityössä tarvitaan mukana myös kulttuurialan ulkopuolisia toimijoita, kuten paikkatietojen ja karttojen osalta Maanmittauslaitosta ja Geologian tutkimuskeskusta ja yrityksiä, joiden arkistoissa on paljon hyviä aineistoja.

<sup>12</sup> <http://www.yso.fi/>

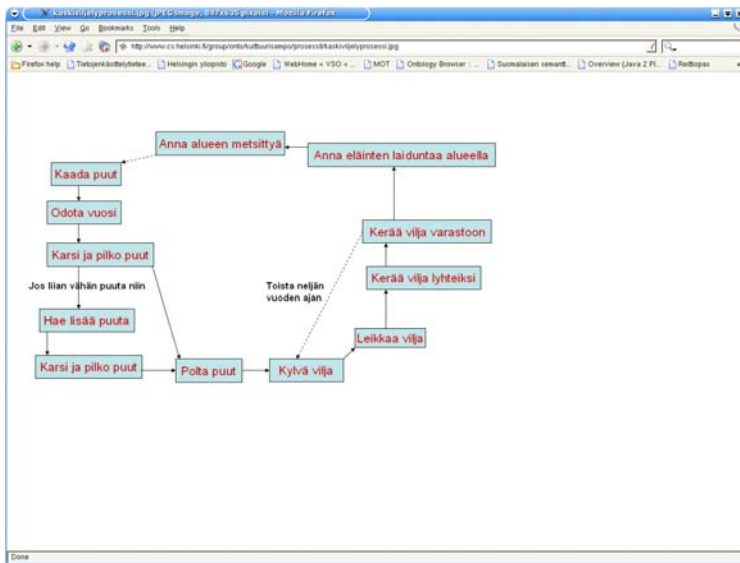
<sup>13</sup> <http://www.panoramio.com/>

Seuraavassa esitellään ensin Kulttuurisampo-hankkeen taustaa, visiota ja historiaa. Tämän jälkeen on vuorossa järjestelmän eri komponenttien, sisällöntuotantoprosessin, siinä käytettävän ontologiainfrastruktuurin ja itse portaaliratkaisun (asiakasliittymän) esittely. Hankkeen rahoitus on varmistunut osana FinnONTO:a ainakin vuoden 2009 loppuun. Lisäksi hanketta rahoittaa erityisesti sisältötyön osalta Suomen Kulttuurirahasto ja siinä käytettävää teknologiaa kehitetään Teknillisessä korkeakoulussa myös EU:n rahoittaman SmartMuseum hankkeen<sup>14</sup> 2008-2010 yhteydessä.

## 2.3 Miten Kulttuurisampo taottiin

### 2.3.1 Ensimmäiset demonstraatiot 2005-2007

Kulttuurisampo-hanke käynnistyi vuonna 2004 lähtökohtana MuseoSuomi-järjestelmä, jonka ajatuksena oli 1) yhdistää semanttisen webin ontologioiden avulla kulttuurihistoriallisten museoiden heterogeenisiä kokoelmia yhteiseksi älykkääksi kokoelmapalveluksi asiakkaille ja 2) luoda museoille yhteinen julkaisukanava semanttiseen webiin. Järjestelmästä valmistui semanttisen portaalin prototyyppi, joka havaittiin toimivaksi ja on edelleenkin verkossa aktiivisessa käytössä. Kulttuurisammossa MuseoSuomen malli haluttiin yleistää erilaisten muistiorganisaatioiden, museoiden, kirjastojen, arkistojen, media-alan toimijoiden ja verkossa olevien yhteisöllisten aineistojen yhteiseksi julkaisukanavaksi. Yhteistyöverkosta Kulttuurisammossa on laajempi kuin eKamut-yhteistyön perinteiset kirjastot, arkistot ja museot. Hanke liitettiin osaksi Tekesin pääosin rahoittamaa FinnONTO-hanketta, jonka tavoitteena oli kansallisen semanttisen webin ontologiainfrastruktuurin kehittäminen maahan. Kulttuurisampo on yksi hankkeen yksi pääsovellus, mikä on taannut järjestelmän kehittämiselle teknologian edistämiseen tähtäävää rahoitusta.



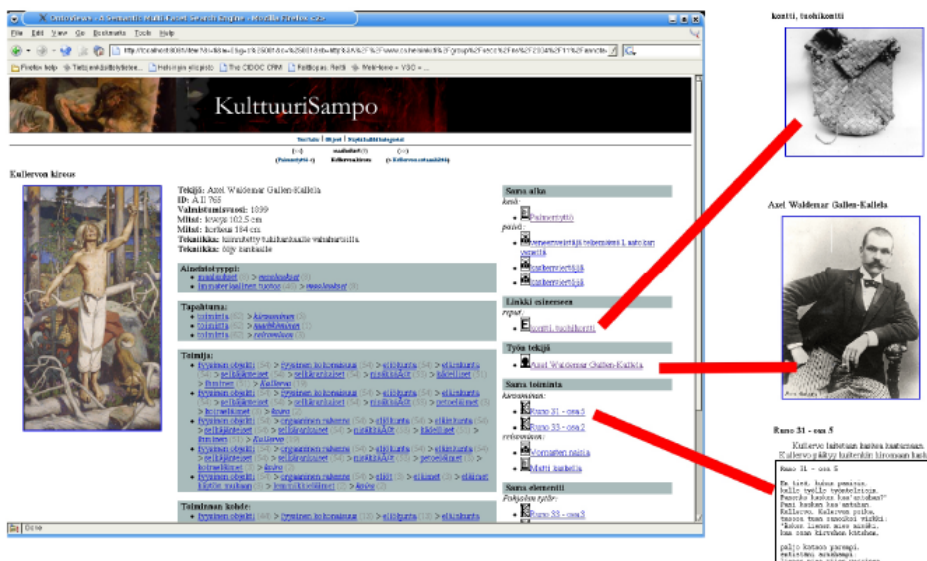
**Kuva 2.2.** Kaskiviljely-prosessin semanttisen kuvauksen visualisointi Kulttuurisampo-portaalin ensimmäisessä demonstraattorissa v. 2005.

<sup>14</sup> <http://smartmuseum.org/>



Hankkeen ensimmäisenä kehitystä ohjannut tutkimushypoteesi oli, että tietämyksen esittämisessä ja tekoälytutkimuksessa paljon käytetty tapahtumaperustainen lähestymistapa (Sowa, 2000), joka on omaksuttu mm. CIDOC CRM –mallin yleiseksi lähtökohdaksi, voisi tarjota hyödyllisen tavan integroida sisältöjä toisiinsa asiakkaan kannalta mielekkäiksi kulttuurisiksi kokonaisuuksiksi. MuseoSuomen arkkitehtuurille toteutettiin v. 2005 ensimmäinen prototyyppi Kulttuurisammosta, jossa aineistona oli MuseoSuomen esineiden lisäksi pieni määrä maalauksia, henkilötietoja, videoita ja ennen kaikkea muutama sisältöjä temaattisesti sitova, tapahtumanperustainen, narratiivinen kuvaus kulttuurisista prosesseista. Näitä olivat kaskiviljely, kalastus ja tarinat kahdesta Kalevalan Kullervoä käsittelevästä runosta. Kuvassa 2.1 näkyvässä selainikkunassa on esimerkkinä kaskiviljelyn prosessin kuvaus Kulttuurisammossa (Junnila et al., 2007).

Tapahtumarakenteiden ja kohteiden metadatan avulla voitiin luoda esimerkiksi linkki Valtion taidemuseossa olevasta Kullervon kirous –maalauksesta Kansallismuseossa olevaan tuohikonttiin, biografiseen tietoon Akseli Gallen-Kallelasta ja maalauksen aihetta vastaavaan kohtaan Kalevalan Kullervo-runossa. Nämä yhteydet voitiin tarjota loppukäyttäjälle koneen löytäminä semanttisina suosittelevina linkkeinä kuvan 2.2 mukaisesti. Kulttuurisammossa ensimmäinen versio 2005 on dokumentoitu Miikka Junnilan (2006) ja Mirva Salmisen (2006) pro gradu –töinä ja artikkeleissa (Hyvönen et al., 2008; Junnila et al., 2008).



**Kuva 2.2.** Kulttuurisammossa prototyyppi 2005. Valtion taidemuseon maalaus linkittyy semanttisesti muunlaisiin kulttuurisisältöihin eri lähteistä.

Seuraavana Kulttuurisampo-hankkeen tutkimusongelmana oli skeemaintegraatio, ts. kysymys siitä, miten metatietorakenteeltaan erilaiset sisällöt (näiden metatiedot), kuten tarinat, esineet, rakennukset ja musiikki voitaisiin yhdistää toisiinsa automaattisesti semanttisella tasolla ja käyttää yhdistelmää sisältöperustaisen älykkään haun ja suosittelevien perustana. Työhypoteesina ollut tapahtumaperustaista sisältöjen kuvailua

testattiin myös tähän ongelmaan. Tuloksena syntyi tapahtumaperustainen menetelmä aineistojen yhdistämiseen ja suositteluun (Ruotsalo, Hyvönen 2007ab) ja Kulttuurisammon toinen inkarnaatio vuonna 2006 diplomityönä (Häggström, 2007). Uusi prototyyppi sisälsi n. 10 000 kohdetta useista eri kokoelmista ja hyödynsi ensimmäistä kertaa FinnONTO:ssa kehitteillä olleen Yleisen suomalaisen ontologian YSO ensimmäisiä versioita. Järjestelmä perustui MuseoSuomen muistinvaraisen Ontogator-hakukoneen (Saarela, 2005; Mäkelä et al., 2006) pohjalta kehitettyyn uuteen ohjelmistoarkkitehtuuriin ja perinteiseen hakukoneeseen (Apache Lucene), joka indeksoitiin tekemään semanttisia hakuja.

Uuden prototyypin muina keskeisinä kehityskohteina olivat ontologiat ja käyttöliittymät. Ontologioiden saralla valmistui mm. menetelmä ajallisten muutosten esittämisestä historialliselle alueille (Kauppinen, Hyvönen, 2007) sekä ensimmäinen ontologia maamme kuntahistoriasta 1865-2007 (Väättäinen, 2007; Kauppinen et al., 2008), joka on käytössä Kulttuurisammon nykyisessä versiossa. Aineistojen visualisointi karttojen avulla nousi yhdeksi kehitysteemaksi Google Maps -palvelun myötä. Kulttuurisammon digitointiin Maanmittauslaitoksen vanhat Karjalan kartat ja vanha geologinen kartta Espoosta, ja kerrostettiin ne modernien Google Maps -karttojen päälle visualisoimaan historiallisia muutoksia alueilla (Kauppinen et al., 2007a) ja liittämään karttoihin erilaisia palveluita. Kulttuurisammon uudessa käyttöliittymässä hyödynnettiin Web 2.0 mash-up tekniikkaa, jolloin esineitä ja muita kohteita voitiin näyttää ja hakea Google Maps -kartoilla ja Simile-aikajanalla. Kehitystyön kohteena oli myös MuseoSuomessa käytetyn ns. näkymähaun edelleen kehittäminen (Mäkelä et al., 2007a) ja ”näyttelyiden” automaattinen tuottaminen semanttisen sisällön perusteella (Mäkelä et al., 2007c).

### **2.3.2 Kulttuurisammon prototyyppi 2008 valmistuu**

Kulttuurisammon varhaiset prototyypit olivat tutkimuksellisia demonstraatioita eikä niitä ole julkaistu MuseoSuomen tapaan yleisöpalveluna verkossa. Tämä toteutuu vasta Kulttuurisammon Kullervon päivänä 25.9.2008 julkistettavassa prototyypissä. Järjestelmä avataan osoitteessa

<http://www.kulttuurisampo.fi>

jossa oli aiemmin 2000-luvulla opetusministeriön kulttuurialan tietoportaaali, jonka toiminta lopetettiin. Tavoitteena on, että ettei Kulttuurisammon uusi inkarnaatio kokisi samaa portaalille niin yleistä kohtaloa. Siksi hankkeessa on alusta asti pyritty sellaisiin ratkaisuihin, jotka voisivat johtaa pysyvän semanttiseen julkaisukanavan syntymiseen maassamme, ei ainoastaan ainutkertaiseen demonstraatioon.

Seuraavassa esitellään Kulttuurisampo-työn taustalla ollut visio, ratkaisumalli ja sen toteuttamisessa tarvittavat erilaiset komponentit.

### 3 Kulttuurisampo-järjestelmän visio, malli ja sen toteutus

---

#### 3.1 Yhteisen julkaisemisen haasteet

Keskeinen haaste muistiorganisaatioiden yhteisportaalien rakentamisessa on aineistojen monimuotoisuus ja luonne:

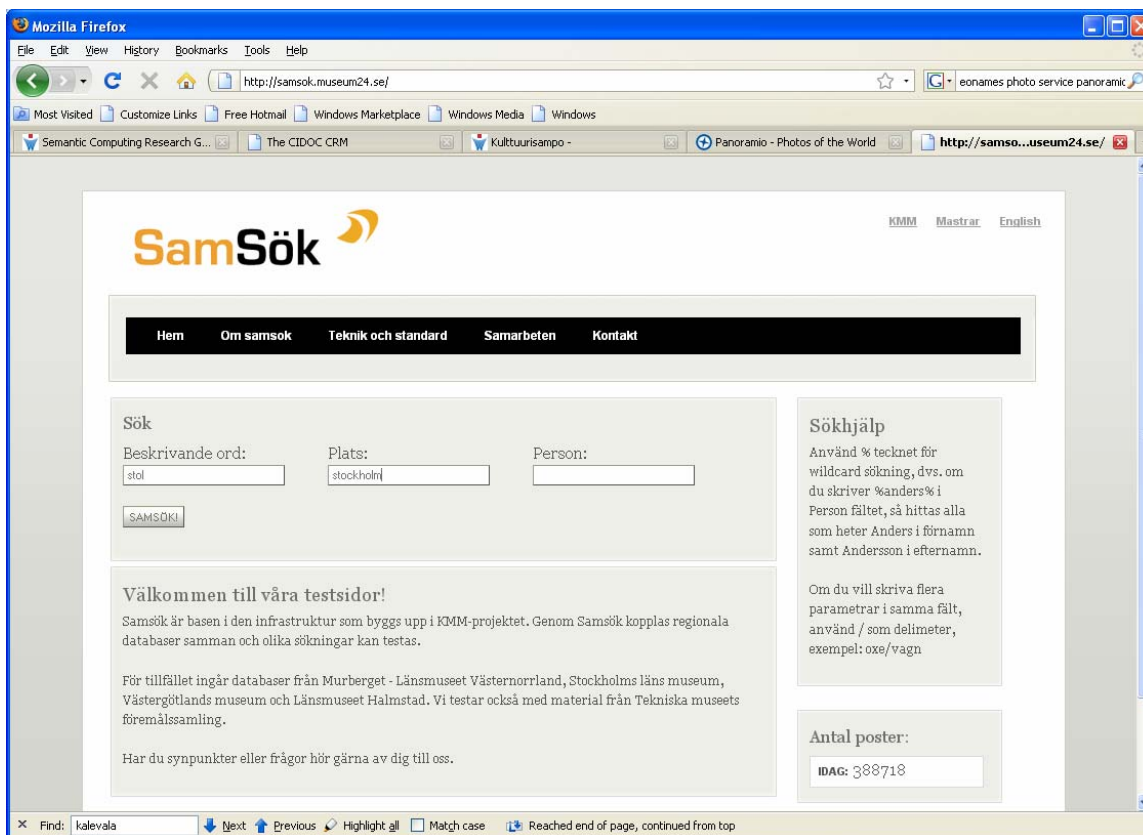
**Luettelointitietojen heterogeenisyys ja laatu.** MuseoSuomi-hankkeen aikana kävi ilmeiseksi, että tietokannoissa käytetyt luettelointitavat ja käytännöt, esimerkiksi sanastot ja tietojen kirjaamisessa käytetyt ilmaukset, poikkeavat toisistaan merkittävästi. Lisäksi maassamme on käytössä useita erilaisia museojärjestelmiä ja tietokantaratkaisuja. Kahden samanlaisen esineen kuvaus eri museoissa voi siksi olla hyvin erilainen. Lisäksi tietoja on kirjattu tietokantoihin monesti lähinnä organisaatioiden omaa käyttöä varten eikä suurelle yleisölle. Esimerkiksi monet MuseoSuomessa olevat esinekuvat ovat tunnistekuvia museohenkilöstölle sisäiseen käyttöön. Sisältöjen julkaiseminen yhteisportaalissa nostaa luettelointikäytäntöjen harmonisointitarpeen ja laatuvaatimukset uuteen arvoonsa, kun eri museoissa tehty työ tulee julkiseksi ja väistämättä vertautuu toisiin museoihin portaalissa.

**Erilaiset aineistotyypit: skeemaintegration haaste.** Kulttuurisisältöjen metatiedot ovat rakenteeltaan monimuotoisia ja jakautuen erilaisiin sisältötyyppeihin, kuten esineisiin, maalauksiin, valokuviin, kirjallisiin teoksiin, musiikkiin, rakennuksiin ja aineettomiin sisältöihin, kuten historiallisiin tapahtumiin tai kulttuurisiin prosesseihin, esimerkiksi maanviljelyyn, keramiikan valmistamiseen tai vaikkapa kansantansseihin. Eri aineistotyyppeihin liittyvät metatiedot voivat poiketa toisistaan huomattavasti, jolloin tarvitaan ratkaisumalli skeemojen integrointiin sisältöjä yhteen sovitettaessa (Hyvönen, 2008). CIDOC CRM –työ tähtää tämän ongelman ratkaisemiseen eri tietoja harmonisoivan CRM-mallin kautta. CRM on 84 käsitteestä ja näiden 141 ominaisuudesta muodostuva ontologinen malli (Crofts et al., 2006), jolle kulttuurisisältöjä voidaan muuntaa erilaisten aineistojen sisällöllisen yhteentoimivuuden saavuttamiseksi. Kirjastoalan kattojärjestön International Federation of Library Associations (IFLA) kehittämää bibliografisen tiedon kuvailun FRBR-järjestelmää ollaan liittämässä CIDOC CRM -malliin. Näiden mallien haasteeksi on muodostunut järjestelmien monimutkaisuus. Oma käsitykseni on, että näitä järjestelmiä on lähdetty kehittämään pikemminkin standardoinnin ja tietointegraation kuin käytännön soveltamisen näkökulmasta. Malliin perustuvia älykkäitä semanttisia portaalreja ei tiettävästi ole ainakaan vielä olemassa, hakujärjestelmiä kylläkin. Malli ollaan kuitenkin otettu perustaksi monissa hankkeissa, esimerkiksi Ruotsin kansallisen Knowledge Management in Museums –projektin SamSök-hakukoneessa<sup>15</sup> ja Saksan kansallismuseossa Nürnbergissä. Kuvassa 3.1. haetaan esimerkkinä Tukholmalaisia tuoleja ruotsalaisten muistiorganisaatioiden yhteisellä SamSök-järjestelmällä, jota ollaan yhdistämässä yleiseurooppalaiseen muistiorganisaatioiden Europeana-järjestelmään<sup>16</sup>.

---

<sup>15</sup> <http://samsok.museum24.se/>

<sup>16</sup> <http://www.europeana.eu/>



**Kuva 3.1.** Ruotsalaisten muistiorganisaatioiden yhteinen hakujärjestelmä, jota ollaan yhdistämässä Europeanaa.

**Arvojen ja sanastojen harmonisointi.** Metatietoskeemojen kehittämisen ja keskinäisen integroinnin ohella toinen ongelma on, mitä arvoja eri metatietoelementeille eri tapauksissa pitäisi antaa. Tähän sanasto-ongelmaan CIDOC CRM ei ota kantaa. FinnONTO-mallissa ratkaisuksi on kehitetty kansallinen ontologioiden ja sanastojen muodostama kokonaisuus KOKO, joka voidaan ottaa käyttöön luettelointi ja hakujärjestelmissä Kansallisen ontologiakirjastopalvelun ONKI kautta (Hyvönen, 2008a; Hyvönen et al., 2008a).

**Sisällöllinen rikkaus ja linkittyvyys.** Aineistot ovat sisällöllisesti usein voimakkaasti toisiinsa liittyviä. Aineistojen semanttista yhdistelyä varten tarvitaan kytkentä metatiedoista ontologioihin ja näihin perustuvaa päättelyä, jos tietoa halutaan käyttää hyväksi syvällisemmin kuin vain perinteisen tekstihaun kautta. Kone ei automaattisesti esimerkiksi ymmärrä, että ruginlapoja käytetään kehräämiseen tai että onki ja verkko ovat kalastamisessa käytettyjä työvälineinä (sen lisäksi että ne ovat tässä artikkelissa semanttisen webin käsitteitä).

**Monikielisyys ja -kulttuurisuus.** Suomalaiset aineistot on sisällönkuvailtu yleensä suomeksi ja suomalaisilla sanastoilla, joista ei ole käytettävissä esimerkiksi englanninkielisiä käännöksiä. Tämä on haaste liitettäessä kokoelmatietoja esimerkiksi

Europeanan kaltaisiin monikielisiin järjestelmiin. Sanastomme ja käsitteistömme on lisäksi monilta osin omaleimaista, kuten vaikkapa poronhoitoon liittyvät käsitteet. Sanastojen ja ontologioiden kääntämiseen olisikin välittömästi kohdennettava resursseja.

Myös kulttuurisisältöjen nykyiseen tuotantomalliin liittyy isoja haasteita:

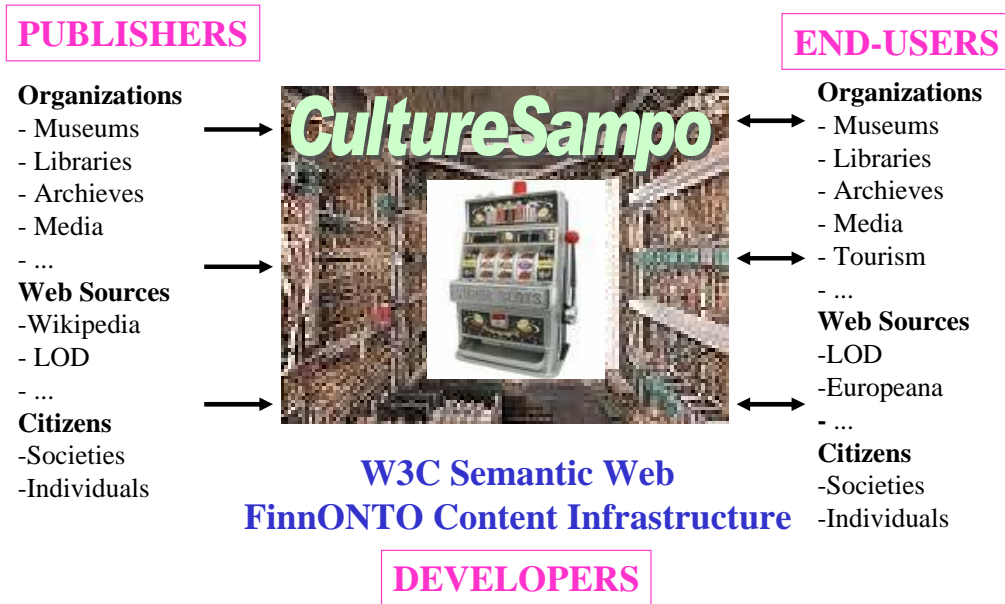
**Hajautettu sisällöntuotanto.** Sisällön tuotanto tapahtuu fyysisesti eri muistiorganisaatioissa ja erilaisilla työvälineillä. Perinteisesti kulttuuritiedon tallentamisesta ovat vastanneet museot, kirjastot ja arkistot, mutta kulttuurista tietoa luovat ja tallentavat myös mm. media-alan organisaatiot, erilaiset perinnejärjestöt, keräilijät ja yksityiset kansalaiset. Webin myötä merkittävään asemaan ovat nousemassa verkkoyhteisöjen tuottamat aineistot, kuten yhteisöllisesti tuotettu Wikipedia. Sisältötyöhön ja sen organisointiin tarvitaan sekä uusia työvälineitä että sosiaalista innovointia.

**Itsenäiset toimijat ja palkitseminen.** Kulttuurisisältöjä tuottavat itsenäiset, toisistaan paljolti riippumattomat toimijat. Tämä johtaa helposti epäyhtenäisiin ja organisaatiokohtaisiin käytäntöihin ja ratkaisuihin. Näiden harmonisointi yhteistä päämäärään varten edellyttää voimakasta yhteistä tahtotilaa, eikä ole vain tekninen vaan erityisesti ”poliittinen” haaste. Yhteisportaaliin liittymisen pitää olla niin palkitsevaa yksittäisille sisällöntuottajille, että järjestelmästä väistämättä aiheutuva lisärasite jää sitä selvästi pienemmäksi ja muutosvastarinta voitetaan. Kulttuurisampo-konseptissa yksittäiset sisällöntuottajat tuodaan siksi voimakkaasti esiin omien ”kotisivujen” avulla, ja portaalin yhteiset sisällöt voidaan ottaa käyttöön semanttisesti rikastuneina sisällöntuottajien omilla verkkosivuilla erityisinä ”leijukkeina”.

### 3.2 Kulttuurisammon ratkaisumalli

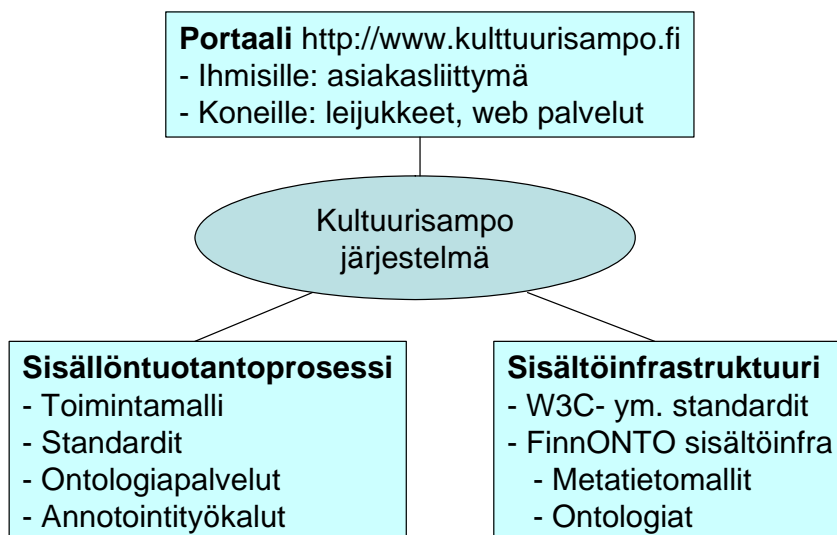
Kulttuurisampo-hankeen tavoitteena on ollut kehittää kansallinen ratkaisumalli joka vastaa edellä kuvattuihin haasteisiin. Mallin idea on tiivistetty kuvassa 3.2. Järjestelmä määrittelee avoimet metatietoskeemat julkaisijoille (kuvassa vasemmalla), joiden mukaista tietosisältöä voidaan ”nappia painamalla” julkaista verkossa. Järjestelmän asiakkaina ovat periaatteessa samat toimijat kuin sisällön tuottajinakin, nyt vain eri roolissa (kuvassa oikealla). Järjestelmää ylläpidetään ja kehitetään W3C:n semanttisen webin standardien ja FinnONTO-sisäätöinfrastruktuurin luomalla perustalla.

Kulttuurisampo ei käytä museoiden omia kokoelma- tai julkaisujärjestelmä, vaan määrittelee ainoastaan semanttiset esittämismuodot erilaisille tietosisällöille. Nämä voidaan toimittaa järjestelmään joko verkon kautta tai offline tiedostoina. Organisaatioiden ohella tietosisältöä tuodaan verkossa olevista yhteisöllisistä palveluista, kuten artikkeleita Wikipediasta ja eri paikkoihin liittyviä valokuvia Geonames.org-palvelusta. Kulttuurisampoon ollaan kehittämässä myös omia Web 2.0 sisällöntuotantokanavia järjestöille ja yksityisille kansalaisille, joiden avulla he voivat julkista omia kokoelmiaan tai muuta sisältöä Kulttuurisammossa.



**Kuva 3.2.** Kulttuurisampo-järjestelmä ja siihen liittyvät toimijat: sisältöjen tuottajat, käyttäjät ja teknologian kehittäjät.

Kulttuurisammon tavoitteena on ollut rakentaa perusta yhteistyöverkostolle ja tekniselle infrastruktuurille, jonka kautta kulttuurisisältöjen tuotantoon ja julkaisemiseen liittyviä haasteita voidaan ratkaista uuden web-teknologian mahdollisuuksia hyödyntäen. Kyse ei siis ole vain yksittäisen portaalidemon teknisestä kehittämisestä vaan paljon laajemmasta kokonaisuudesta ja prosessista, jolla on toteutuessaan vaikutuksia kulttuurisisältöjen koko tuotantojärjestelmään maassamme.



**Kuva 3.2.** Kulttuurisampo-järjestelmän kolme pääkomponenttia

Ratkaisumalli voidaan jakaa kolmeen osa-alueeseen kuva 3.2. mukaisesti:

**Sisällöntuotantoprosessi.** Järjestelmä määrittelee sisällölliset rajapinnat erilaisille aineistotyypeille ja tarjoaa prosessin ja tukea, jolla tiedot voidaan tuottaa ja kerätä sisällöntuottajilta.

**Sisältöinfrastruktuuri: palvelut ja työkalut.** Järjestelmä perustuu FinnONTO-hankkeessa kehitettyyn kansalliseen ontologiajärjestelmään, jossa eri alojen sanastoja on toimitettu tietokoneen paremmin ”ymmärtäviksi” ontologioiksi. Infrastruktuurin yleisenä perustana ovat erityisesti webin infrastruktuuria koordinoivan W3C-järjestön semanttisen webin standardit ja hyvät käytännöt, kuten RDS(S), Web Ontology Language OWL<sup>17</sup>, Simple Knowledge Organization System SKOS<sup>18</sup>, ja SPARQL<sup>19</sup>. Ontologiat on yhdistetty toisiinsa eri alojen väliseksi KOKO-ontologiaksi ja eräiksi muiksi järjestelmiksi (esimerkiksi hollantilainen Iconclass-järjestelmä ja sen yhdistäminen Yleiseen suomalaiseen ontologiaan YSO). Sanasto/ontologiapalvelut ovat käytettävissä keskitetysti Kansallisen ontologikirjastopalvelun ONKI kautta. ONKI:n avulla voidaan paitsi julkaista sanastot myös liittää niihin liittyvät indeksointi ja hakupalvelut osaksi luettelointi- ja muita sisällöntuotantoprosesseja erittäin kustannustehokkaalla tavalla (AJAX). Menetelmä muistuttaa Google Maps palveluiden liitämistä sovellukseen.

Sanastopalveluiden lisäksi FinnONTO tarjoaa käyttöön erilaisia välineitä metatietojen tuotantoon, kuten metatietotoimitin SAHA (Valkeapää et al., 2008; Valkeapää 2006), jonka avulla metatietojen sisältöjen tuotanto voidaan mahdollistaa verkossa, sekä sisältöjen automaattiseen kuvailuun ja validointiin tarkoitettuja työvälineitä, kuten POKA (Alm, 2007) ja VERA.

**Kulttuurisampo-portaali ihmisille ja koneille.** Itse portaali yhdistelee eri tiedontuottajien toimittamat aineistot ontologioiden avulla laajaksi semanttiseksi verkoksi. Tämän tietämyskannan varaan on toteutettu *käyttöliittymät sekä ihmisasiakkaille että verkkosovelluksille*. Asiakkaalle luodun käyttöliittymän ideana on tarjota erilaisia näkökulmia suomalaiseen kulttuuriin, joista perinteiset hakutoiminnot ovat vain yksi. Aineistoja voi esimerkiksi tutkia kartoilla ja aikajanalla, selvittää henkilöiden välisiä sosiaalisia yhteyksiä tai tarkastella sisältöjä jonkin tekstin kuten Kalevalan ja Kansallisbiografian elämäkertojen viitekehystä käsin. Järjestelmän tavoitteet ovat asiakasliittymän osalta erilaiset kuin esimerkiksi kuvan 3.1 ruotsalaisessa SamSök-järjestelmässä tai kotimaisessa Suomen museot online -sovelluksessa,

Ihmisten ohella Kulttuurisampo voivat käyttää toiset verkossa olevat verkkosovellukset. Kulttuurisampo ei ole musta-aukko, johon sisällöntuottajat vievät aineistojaan, vaan ajatuksena on tarjota semanttisesti eri toimijoiden aineistoista rikastetut sisällöt verkkopalveluina takaisin sisällöntuottajille ja muille toimijoille Web 2.0 –tekniikan avulla semanttisina ”leijukkeina” (widget). Esimerkiksi museon omalle, yhtä taideteosta esittelevälle verkkosivulle voidaan yhdellä rivillä Javascript-koodia lisätä linkit saman taiteilijan muihin teoksiin omassa tai muissa Kulttuurisampo-museoissa, linkit tietohin taiteilijan elämästä kansallisbiografiassa, Wikipediassa tai muissa lähteissä jne. Palvelua

<sup>17</sup> <http://www.w3.org/RDF/>

<sup>18</sup> <http://www.w3.org/2004/02/skos/>

<sup>19</sup> [www.w3.org/TR/rdf-sparql-query/](http://www.w3.org/TR/rdf-sparql-query/)

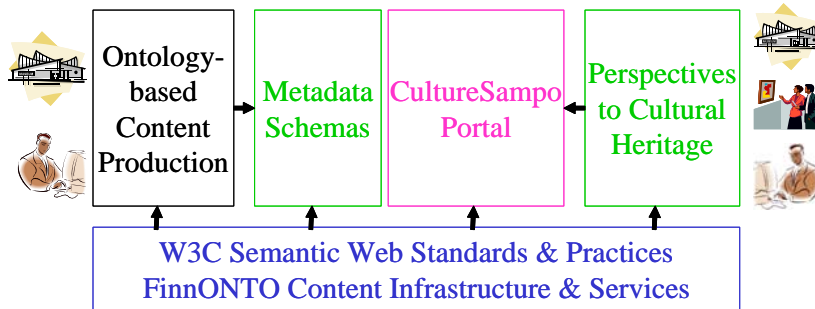
voidaan hyödyntää myös mm. turistiportaaleissa esiteltäessä vaikkapa tietyn kaupungin omaa kulttuuritarjontaa tai paikkaan liittyvää kulttuuriperimää eri lähteistä. Tämä tarjoaa uusia mahdollisuuksia kulttuurisisältöjen liiketoiminnalliseen hyödyntämiseen.

Seuraavassa esitellään ratkaisumallin osat tarkemmin.



## 4 Semanttinen sisällöntuotantoprosessi

Kuva 4.1. esittää Kulttuurisammon julkaisuprosessin keskeisimmät vaiheet: sisältöjä tuotetaan metatietoina portaalin, josta ne ovat asiakkaiden haettavissa ja selattavissa temaattisten näkökulmien kautta. Kuvassa 4.1 on prosessin tarkempi kuva. Julkaisijan kannalta keskeisiä metatietoskeemoja on hyvin monia johtuen sisältötyyppien erilaisuudesta. Esimerkiksi rakennuksen, kuten Espoossa sijaitsevan Akseli Gallen-Kallelan Tarvaspää-ateljeen metatiedot ovat hyvin erilaiset kuin vaikkapa hänen maalaamansa ”Kullervon kirous” –maalauksen, joka löytyy Ateneumin kokoelmista, tai Suomalaisen Kirjallisuuden Seuran Kansallisbiografiassa olevan elämäkerran.



**Kuva 4.1.** Kulttuurisammon julkaisuprosessi.

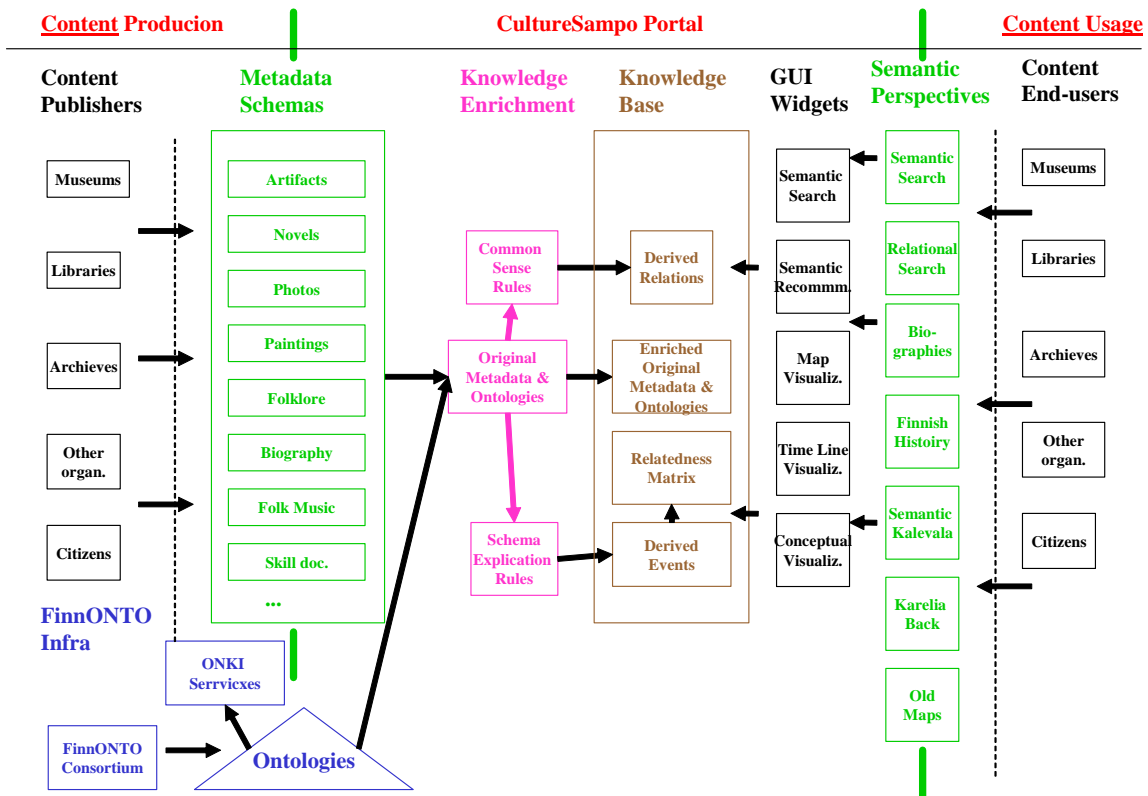
Metatietoskeema	Aineistotyytit	Aineistojen toimittajat
1 museoesineet	museoesineet	Espoon kaupunginmuseo Lahden kaupunginmuseo Kansallismuseo Pohjois-Karjalan museo Suomen maatalousmuseo Sarka
2 taide	maalaukset veistokset piirustukset esittävä ja abstrakti taide	Valtion taidemuseo
3 kaunokirjallisuus	romaanit tai novellit sarjakuvat	Helsingin kaupunginkirjasto
4 WWW-tietosivut	web sivut	Wikipedia
5 runot	loitsut runot tarina	Suomalaisen Kirjallisuuden Seura Suomen kansan vanhat runot
6 fiktiiviset kohteet	Kalevalan paikat Kalevalan henkilöt	Suomalaisen Kirjallisuuden Seura
7 kansansävelmät	hengelliset sävelmät kansantanssit kantele- ja jouhikkosävelmät laulusävelmät runosävelmät	Jyväskylän yliopisto, musiikin laitos Suomalaisen kirjallisuuden seura
8 valokuvat	valokuvat	Kuopion kulttuurihistoriallinen museo Kansallismuseo
9 ilmavalokuvat	ilmavalokuvat	Veljekset Karhumäki Oy
10 toimijat	henkilöt ja organisaatiot	Getty-säätiö, Union List of Artist Name
11 biografiat	biografiat	Suomalaisen Kirjallisuuden Seura SKS Kansallisbiografia
12 historialliset tapahtumat	historialliset tapahtumat	Agricola – Suomen historiaverkko
13 taidot	taitojen prosessikuvaukset	Taideteollisen korkeakoulun kirjasto Suomen maatalousmuseo Sarka
14 videot	prosessien dokumentoinnit	Espoon kaupunginmuseo
15 rakennetut kohteet	rakennetut kohteet luonnossa	Espoon kaupunginmuseo
16 muinaismuistokohteet	muinaismuistokohteet	Museovirasto

**Kuva 4.1.** Portaalin metatietomallit, aineistotyytit ja aineistojen toimittajat.

Taulukko 4.1 luettelee Kulttuurisammossa käytössä olevat metatietomallit ja niillä esitettävät sisältötyypit. Joissain tapauksessa samalla mallilla esitetään erilaisia aineistoja (esim. maalaukset ja veistokset), mutta yleensä jokaisella aineistotyypillä on ainakin jossain määrin toisista poikkeava metatietomalli. Tämän perusteella määräytyy mm. kohteen esitystapa portaalin sivuilla.

Kun metatieto on saatu portaaliin, aiheistoa voidaan rikastaa semanttisesti eri tavoilla (kuvassa keskellä). 1) Ensinnäkin voidaan tehdä ontologioiden yleisen RDFS-semantiikan perusteella päätelmiä, kuten että jos posliini on keramiikkaa ja kokoelmassa on posliininen kahvikuppi, niin kuppi on myös keramiikkaa. 2) Erityisten kulttuurista tietoa mallintavien ontologisten yhteyksien ja päättelysääntöjen avulla voidaan päätellä lisää tietoa kohteista (common sense) vastaavaan tapaan kuin MuseoSuomessa (Hyvönen et al., 2005). 3) Metatietoskeemojen rakenteen perusteella voidaan eksplikoida niihin sisältyvää implisiittistä tietoa (Ruotsalo, Hyvönen, 2008). Esimerkiksi siitä, että Aino-triptyykin (kuva 1.1) luoja (dc.subject) on Akseli Gallen-Kallela seuraa, että hän maalasi kyseisen taulun Pariisissa 1891.

### CultureSampo: Finnish Cultural Heritage on the Semantic Web



Kuva 4.3. Kulttuurisammon yleisarkkitehtuuri.

Kulttuurisammon sisällöntuotantomalli perustuu MuseoSuomen yhteydessä kehitettyyn systeemiin (Hyvönen et al., 2004). Siinä sisällöntuotanto jakautuu seuraaviin osiin:

**Ontologioiden kehittäminen.** Järjestelmän taustalla olevia ontologioita on tarkoitus kehittää osana kansallista FinnONTO-infrastruktuuria, joka esitellään tarkemmin tuonnempana.

**Sanastojen kehittäminen.** Sanastolla tarkoitetaan tässä yhteydessä muistiorganisaatioiden luetteloinnissa ja sisällönkuvailussa käyttämiä ilmaisuja, jotka on tavalla tai toisella peilattava ontologisille käsitteille. Eri organisaatioissa samoilla termeillä voi olla erilaisia merkityksiä, samoin eri aikoina sanojen merkitys voi vaihdella. Tästä johtuen sanastojen peilaukset on voitava tehdä organisaatio- ja aineistokohtaisesti.

**Annotointi (sisältöjen kuvailu).** Annotointi vastaa perinteistä luettelointia, ts. metatietojen tuottamista. Aineistojen verkkojulkaiseminen (semanttisessa) webissä edellyttää kuitenkin aiempaa täsmällisempää ja kurinalaisempaa luettelointia, jotta myös koneet voisivat yksikäsitteisesti ”ymmärtää” metatietoja. Kulttuurisammon mallissa organisaatioita kannustetaan ontologiaperustaiseen sisällönkuvailuun, joka perustuu käsitteiden kieliriippumattomiin URI-tunnisteisiin (tai muihin yksikäsitteisiin koodeihin tai merkintöihin), eikä vain ihmisille tarkoitettuihin kielellisiin ilmaisiin. Prosessia tuetaan Kansallisen ontologiapalvelun ONKI avulla.

**Suosittelut.** Aineistojen semanttinen yhdistäminen voidaan määritellä joustavasti kullekin aineistotyyppille erikseen semanttisten sääntöhahmojen avulla.

Kulttuurisammon prototyypissä sisältöjen konvertointi RDF-muotoon on tehty SeCo-tutkimusryhmän toimesta muistiorganisaatioiden toimittamista tietokannoista ja muista aineistosta. Ajatuksena on kuitenkin sisällöntuotannon muunnosten hajauttaminen eri organisaatioihin vastaavaan tapaan kuin esimerkiksi SMOL-järjestelmässä määrittelemällä julkisina rajapintoina metatietoformaattit aineistoille. SMOL:n tapauksessa käytössä on yksi XML-skeema. Kulttuurisammossa skeemoja on sisäisesti käytössä 16 eri aineistotyyppille. Näiden pohjalta ollaan laatimassa mallia järjestelmän ulkoisiksi metatietorajapinnoiksi. Rajapinnat määritellään RDF-muodossa.

Yksi ongelma aineistojen julkaisemisessa on niihin helposti syntyvät syntaktiset ja semanttiset virheet. Aineistojen semanttista validointia varten ollaan kehittämässä Vera-työkalua<sup>20</sup>. Se on jo osoittautunut varsin tarpeelliseksi Kulttuurisammon prototyypin aineistojen tarkistamisessa ja muokkaamisessa. Veran avulla sisällöntuottajat voivat tarkistaa, että kokoelmista tuotettu metatieto on rajapinnan metatietomäärittelyjen mukaista, esimerkiksi että maalauksen tekijä on henkilö eikä esimerkiksi jokin esine.

---

<sup>20</sup> <http://www.seco.tkk.fi/services/vera/>

## 5 Sisältöinfrastruktuuri: ontologiat ja ontologiapalvelut

Semanttisen webin tuleminen on käynnistänyt ja vilkastuttanut sanasto- ja ontologiatyötä monilla eri aloilla. Esimerkiksi terveysalalla on kehitetty jättimäinen, satojen tuhansien käsitteiden Snomed Clinical Terms<sup>21</sup> ja kulttuurialalla käytetään Getty-säätiön kulttuurikäsitteitä, taiteilijoita ja historiallisia paikkoja mallittavia tesauksia Art and Architecture Thesaurus AAT (n. 34 000 käsitettä ja 131 000 termiä), Universal List of Artist Names (ULAN) (n. 120 000 käsitettä ja 293 000 nimeä) ja Thesaurus of Geographical Names (TGN). (n. 912 000 paikkaa)<sup>22</sup>. Geonames-palvelu<sup>23</sup> sisältää miljoonia nykyisiä paikkoja. WordNet<sup>24</sup> (Fellbaum, 1998) on laajasti käytetty englannin kielen ontologinen rakenne, joka sisältää n. 155 000 englannin kielen sanaa koottuna 118 000 käsitteeksi (ns. synonyymijoukoksi, synset). Siitä on kehitetty versioita muillekin kielille (EuroWordNet<sup>25</sup>), muttei valitettavasti suomelle. Kansallisten semanttisen webin ontologioiden tulisi olla yhteentoimivia kansainvälisten käsitteistöjen kanssa. Toisaalta kansalliset käytännöt, sisällöt ja kielet joudutaan ottamaan huomioon.

Semanttisen webin sisällöllinen ontologiatyö on ollut paljolti alakohtaista ja perustunut olemassa olevien luokitusten käyttöön ja muokkaamiseen. Eri alojen välisen työn koordinaatiota ei ole juurikaan ollut. Eri sovelluksissa tarvittava ontologinen tieto on yleensä ainakin osittain tapauskohtaista, mistä johtuen aiemmin kehitettyjen ontologioiden uusiokäyttö toisissa sovelluksissa on haasteellisempaa kuin helposti kuvitellaan. Työn tuloksena semanttisen webin kehittäjien käytettävissä on kuitenkin jatkuvasti kasvava joukko ontologioita. Esimerkiksi ontologioiden hakuun kehitetty Swoogle hakukone<sup>26</sup> on indeksoinut yli 10 000 verkossa saatavilla olevaa ontologiaa. De facto -standardeja on muodostumassa suurten ontologiahankkeiden kuten WordNetin ja Getty säätiön sanastojen ympärille.

### 5.1 FinnONTO-ontologiatyön periaatteita

FinnONTO-hankkeessa ontologiatyö käynnistyi kansallisten käytäntöjen pohjalta ja perustuen olemassa oleviin asiasanastoihin. Asiasanastot (Aitchison et al., 2000) esitetään yleensä ISO 2788, BB 5723 (Iso-Britannia) ja ANSI/NISO Z39.19 (USA) standardien mukaisessa muodossa. Sanasto koostuu termeistä, joiden merkitys on kuvattu joukolla suhteita toisiin termeihin. Sanaston semanttiset suhteet jakaantuvat kolmeen pääryhmään: ekvivalenssisuhteisiin (KÄYTÄ, KORVAA), hierarkkisiin suhteisiin (LT laajempi termi, ST suppeampi termi) ja assosiativisiin suhteisiin (RT rinnakkaistermi).

FinnONTO-hankkeessa on pyritty hyödyntämään maksimaalisesti maassamme jo tehty sanastotyö. Näin tietokantoihin jo indeksoitu aineisto saadaan helpommin semanttisen käsittelyn piiriin ja sanastojen kehittäjien työ hyödynnettyä. Nykyisiä sanastoja ja luokituksia voidaan yrittää hyödyntää joko sellaisenaan semanttisessa webissä tekemällä

<sup>21</sup> <http://snomed.org/>

<sup>22</sup> [http://www.getty.edu/research/conducting\\_research/vocabularies/](http://www.getty.edu/research/conducting_research/vocabularies/)

<sup>23</sup> <http://www.geonames.org/>

<sup>24</sup> <http://wordnet.princeton.edu/>

<sup>25</sup> <http://www.ilc.uva.nl/EuroWordNet/>

<sup>26</sup> <http://swoogle.umbc.edu/>

mekaanisia muunnoksia esimerkiksi SKOS-formaattiin. FinnONTO:ssa monessa tapauksessa näin on meneteltykin. Yleisenä tavoitteena käytössä olevien asiasanatesaurusten muuntamisessa on kuitenkin täsmentää niiden semantiikkaa ja rakennetta siten, että niiden käyttö semanttisen webin sovelluksissa kuten päättelyssä ja visualisoinnissa tulee paremmin mahdolliseksi. Asiasanastojen kevyeen ontologisointiin kehitettiin yksinkertainen menetelmä (Hyvönen et al., 2008a), jonka avulla joukko sanastoja on muutettu semanttisen webin kannalta käyttökelpoisempaan muotoon. FinnONTO-hankkeen ehdotus asiasanastojen kehittäjille on astua askel eteenpäin ja kehittää asiasanastojen sijasta jatkossa yksinkertaisia ontologioita (light weight ontology) (Hyvönen, 2005).

## 5.2 Ontologiat

FinnONTO-hankkeessa kehitettyjä tietämyksen organisoinnin järjestelmiä (knowledge organization system, KOS<sup>27</sup>) voidaan jäsentää seuraaviin kategorioihin: ontologiat, luokitukset, tesauukset (hierarkkiset sanastot), asiasanastot ja sana/käsitelistat. Tarkka rajanveto on monasti epäselvää ja vaikeaa eikä tarkoituksenmukaistakaan. Ontologioita voidaan ryhmitellä edelleen luokkaontologioihin ja yksilöontologioihin. Luokkaontologiat sisältävät asiasanastojen tapaan yleiskäsitteitä luokkina (class), kuten ”kissat”, ”kansantuote” tai ”kylät”. Luokka edustaa siihen kuuluvia yksilöitä (individual, instance) ja määrittelee näille mahdollisesti yhteisiä ominaisuuksia. Esimerkiksi kaupungit-luokan yksilöitä voivat olla Helsinki tai New York, ja kaupungin sijainti voidaan ilmaista koordinaateilla ja sitä laajempi hallinnollinen alue valtio-yksilöön viittaavalla ominaisuudella.

Seuraavassa esitellään FinnONTO:n kansallisen semanttisen webin ytimeksi ehdottamaa KOKO-ontologiajärjestelmää, joka on käytössä Kulttuurisampo-portaalissa.

### 5.2.1 KOKO-luokkaontologiat

KOKO koostuu joukosta luokkaontologiata, jotka on muodostettu maassamme käytössä olevista asiasanastoista. Nämä on ensin muunnettu semanttisen webin OWL-muotoon ja siirretty Protege-ontologiaeditoriin<sup>28</sup>. Tämän jälkeen sanaston hierarkiat on täydennetty ja semanttiset suhteet (LT laajempi termi, ST suppeampi termi, RT rinnakkaistermi) täsmennetty ja korjattu käsityönä. Lopputulos on semantiikaltaan RDF(S)-suosituksen mukainen yksinkertainen luokkahierarkia, jonka luokista luotavien yksilöiden pitää olla kaikkien yläluokkiensa yksilöitä (transitive closure).

KOKO:n osaontologioista keskeisin on kansalliskirjaston Yleisen suomalaisen asiasanaston YSA pohjalta muodostettu Yleinen suomalainen ontologia YSO. YSA on nykyisin yli 25 000 termin sanasto, joka on laajassa käytössä maassamme eri aloilla sisällönkuvailussa. YSO muodostaa KOKO:n yläontologian (upper ontology), jota muut alakohtaiset ontologiat täsmentävät omilla alueillaan. Esimerkiksi maa- ja metsätalouden ontologia AFO luettelee YSO:n käsitteen ”madot” tarkempia alaluokkia, kuten ”sukkulamadot” ja ”ankeroiset”.

<sup>27</sup> [http://www.db.dk/bh/lifeboat\\_ko/concepts/knowledge\\_organization\\_systems.htm](http://www.db.dk/bh/lifeboat_ko/concepts/knowledge_organization_systems.htm)

<sup>28</sup> <http://protege.stanford.edu/>

Jokaiselle käsitteelle on annettu nimikkeen ohella yksilöivä pysyvä URI-tunniste HTTP-skeeman mukaan, mikä on vakiintunut semanttisen webin alalla käytännöksi. Nimet ovat muotoa:

[http://www.yso.fi/onto/ontologian\\_nimi/käsitteen\\_tunniste](http://www.yso.fi/onto/ontologian_nimi/käsitteen_tunniste)

Esimekiksi YSO-ontologian käsite ”nimet” on saanut tunniste:

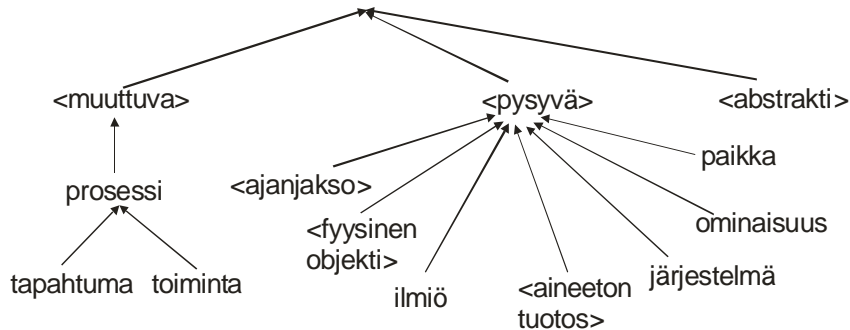
<http://www.yso.fi/onto/yso/p1638>

Käsitteiden tunnisteet ovat kirjain-numeroyhdistelmiä ilman mitään semantiikka, mikä edesauttaa niiden käyttöä kielirajojen yli ja pysyvyyttä. Sama paikallinen käsitteen tunniste kuten p1638 voi olla käytössä useissa eri ontologioissa.

Keskeisiä syitä HTTP-skeeman käyttöön URI-tunnisteissa ovat: 1) nimipalvelu (Domain Name System DNS) on jo maailmanlaajuisessa käytössä webissä ja mahdollistaa 1) domain-nimen perusteella tapahtuvan yksilöinnin sekä 2) URI:in liittyvän tietoverkon (RDF-määritelmä) hakemisen webistä, mikä on mm. yksi Tim Berners-Leen ja W3C:n kiteyttämistä Linked Data ajattelumallin periaatteista<sup>29</sup>. Muut eri yhteyksissä esitetyt nimeämisjärjestelmät ja URI-skeemat, kuten URN ja OID, ovat näissä suhteissa ongelmallisia.

### 5.2.2 KOKO-ontologian rakenne ja sisältö

YSO:n käsitteet on jaettu DOLCE-järjestelmän (Gangemi et al., 2002) perusteella kolmeen pääluokkaan (kuva 5.1): ”muuttuva” (perdurant), joka sisältää tapahtumiin ja toimintoihin liittyvät käsitteet, ”pysyvä” (endurant), joka sisältää mm. fyysisiin objekteihin, ajanjaksoihin ja paikkoihin liittyviä luokkia sekä ”abstrakti” (abstract), joka sisältää lähinnä suureisiin ja yksiköihin liittyviä käsitteitä. Prosessien erottaminen muusta käsitteistöstä on yksi keskeinen periaate tietämyksen esittämisessä tekoälyssä (Sowa, 2000) ja mahdollistaa mm. Kulttuurisammossa tapahtumakeskeisen sisältöjen mallintamisen ja skeemaintegraatiota (Ruotsalo, Hyvönen, 2007ab).

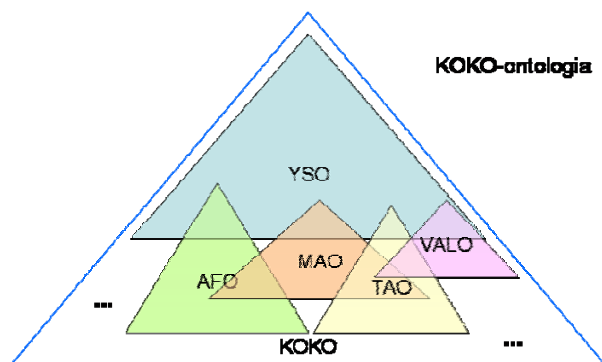


**Kuva 5.1.** YSO-ontologian ylimmät luokat. Ryhmittelevät käsitteet on merkitty kulmasuluilla.

<sup>29</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

Kuva 5.2 havainnollistaa YSO:n suhdetta sitä tarkentaviin alaontologioihin. Alaontologiat kehitetään lähtökohtaisesti YSO-ontologian laajennuksina peilaamalla niiden leikkaavat käsitteet YSO:n vastaaviin. Sääntönä on, että jokaisella alaontologian käsitteellä pitää olla vähintään yksi ekvivalentti tai yläkäsite YSO:ssa. Näin rakentuu yhtenäinen hierarkia ja semanttinen yhteys alaontologioiden ja YSON välille. Samalla syntyy silta YSO:n kautta myös eri alaontologioiden välille. Alaontologiat voivat leikata toisiaan myös suoraan, ei vain YSO:n kautta. Tällöin voidaan muodostaa myös suoria peilauksia eri alaontologioiden välille.

Käytännössä ontologioiden peilaus on FinnONTO:ssa tehty ensin alustavasti tietokoneella termien nimien perusteella. Tämän jälkeen rakenne on tarkistettu, korjattu ja muodostettu käsityönä Protege-editorin avulla. Kokemustemme mukaan alaontologian muodostaminen peilaamalla se YSO:oon helpottaa ontologiatyötä merkittävästi, kun YSO:n yhteydessä jo tehtyä hierarkisointityötä voidaan yleensä hyödyntää sellaisenaan.



**Kuva 5.2.** Yhteisöllinen kokonaisuontologia KOKO koostuu yläontologiasta YSO ja sitä tarkentavista alaontologioista, joilla on yhteisiä käsitteitä YSO:n ja toisten alaontologioiden kanssa. Ontologioiden keskinäiset leikkaukset kuvassa ovat vain viitteellisiä ja tarkoitettu yleiskuvan havainnollistukseksi.

Eri alaontologioiden kehitysversiot YSO:n (ja mahdollisesti muiden ontologioiden) suhteen peilattuna löytyvät ONKI-palvelusta omasta osiostaan. Tätä kirjoittaessa KOKO-järjestelmään on integroitu ensimmäiset versiot taulukossa 5.1 esitellyistä ontologioista, yhteensä yli 37 000 luokkakäsitteen ontologiat. Taulukossa 5.2. on lueteltu ontologioiden peilauksen tuloksena löydettyjen ja muodostettujen ekvivalenssisuhteiden määriä. Lisäksi ontologioiden käsitteiden välille on lisätty yläluokkasuhteita (subclassOf).

Nähtäväksi jää millaiseksi rajanveto ja työnjako eri ontologioiden välille lopulta muodostuu, kannattaako esimerkiksi pienempiä ontologioita missä määrin yhdistää. FinnONTO:n yleisenä ratkaisumallina kehitystyölle on käynnistää Web 2.0 -henkinen yhteisöllinen kehitysprosessi eri ontologioiden kehittäjäryhmien välillä, jota tuetaan ONKI-palveluiden kautta.

Alakohtaiset yhdistelmäontologiat on tarkoitettu vain ontologioiden kehittäjille. Sisällön kuvaajille ja halupalveluiden käyttäjille KOKO-ontologia näyttäytyy yhtenäisenä

ontologiana kuvan 5.3 havainnollistamalla tavalla. Annotoinneissa suositellaan käytettäväksi KOKO-ontologiassa olevia URI-tunnisteita ja rakenteita. Indeksointi hyvinkin tarkkoilla käsitteillä on mahdollista ilman että yhteys yleisempiin yläkäsitteisiin katkeaisi. Tämä edesauttaa eri tarkkuustasoilla kuvailtujen aineistojen yhteentoimivuutta. Kokonaisuudesta voidaan myös tarpeen mukaan poistaa tietyssä sovelluksessa tarpeettomia alaontologioita.

Nimi	Ontologian ala	Pohjana oleva asiasanasto	Koko
YSO	Yleinen suomalainen ontologia Yleiskäsitteet, monialainen	Yleinen suomalainen asiasanasto YSA. Kansalliskirjasto. <a href="http://vesa.lib.helsinki.fi/">http://vesa.lib.helsinki.fi/</a>	20 600
MAO	Museoalan ontologia	Museoalan asiasanasto MASA. Museovirasto (Leskinen, 1997).	6800
AFO	Maa- ja metsätalouden ontologia. Maa ja metsätalous	Agriforest tesaurus. Viikin tiedekirjasto.	5500
TAO	Taideteollisuusalan ontologia	Käsi- ja taideteollisuuden asiasanasto (Krogerus et al., 1997)	2600
VALO	Valokuvausalan ontologia	Valokuvakirjallisuuden asiasanasto (Ulla Karttunen). Valokuvaustekniikan asiasanasto. Suomen valokuvataiteen museo.	1900

**Taulukko 5.1.** KOKO-ontologian osaontologioita. Uusia alaontologioita liitetään järjestelmään muuntamalla asiasanastoja ontologioiksi.

Leikkaavat ontologiat	Yhteisten ekvivalenttien käsitteiden määrä
YSO + TAO	1990
YSO + MAO	2330
YSO + VALO	950
MAO + TAO	1190

**Taulukko 5.2.** KOKO ontologian osaontologiat jakavat runsaasti yhteisiä ekvivalenteja käsitteitä YSO:n kanssa ja keskenään.



**Kuva 5.3.** KOKO-ontologia näkyy indeksoijalle ja tiedon hakijalle yhtenäisenä kokonaisuutena (vrt. kuvan 2.2 kehittäjänäkymä).



### 5.2.3 KOKO-yksilöontologiat

FinnONTO:ssa ensimmäisenä kehitettyjä yksilöontologioita ovat paikkaontologiat ja toimijaontologiat (henkilöt ja organisaatiot), joita tarvitaan kulttuurialan portaaleissa. Paikkojen osalta keskeinen ontologia on Suomalainen paikkaontologia SUO (Henriksson et al., 2008). Se koostuu n. 700 paikkaluokasta, joista on sitten muodostettu suuri joukko yksilöitä Maanmittauslaitoksen paikannimirekisterin (PNR) pohjalta (n. 800 000 kotimaista paikkaa) ja kansainvälisistä paikka-aineistoista (GEONet Names Server GNS, miljoonia paikkoja). Suomen historiallisista kunnista ja niiden välisistä muutoksista on laadittu erillinen Suomen historiallinen paikkaontologia SAPO (Kauppinen, Hyvönen, 2007; Kauppinen et al., 2008), joka on yhdistetty SUO:hon.

Henkilöiden ja organisaatioiden osalta FinnONTOssa käytetään Getty-säätiön Union List of Artis Names (ULAN) käsitelistää, jota on tarkoitettu rikastaa kotimaisilla aineistoilla, kuten Suomalaisen kansallisbiografian tiedoilla, kotimaisilla auktoriteettitietokannoilla ja Kulttuurisampo-portaalin kokoelmien metatiedoissa esiintyvillä toimijoilla.

### 5.2.4 Muita ontologioita

KOKO-ontologioiden ohella hankkeessa on kehitetty ja tuotettu muitakin ontologioita, joita on sitten peilattu YSO:on (ja sitä kautta muihin alaontologioihin) eri tavoilla. Kulttuurisektorin sovellusten kannalta mainittakoon seuraavat:

**Iconclass**<sup>30</sup> on Hollannissa kehitetty laaja käsitteistö taiteen ikonograafiseen kuvailuun, joka on osittain suomennettu Valtion taidemuseon toimesta. Systeemi linkitettiin tätä kautta YSO-ontologiaan (Haaramo, 2006).

**Helsingin kaupunginkirjaston (HKK) luokitusjärjestelmä HKLJ**<sup>31</sup> on HKK:ssa käytetty laaja järjestelmä, jonka luokat on kuvailtu joukolla YSA:n termejä. Systeemistä on muodostettu oma, YSO:on linkittyvä ontologinen rakenne. HKLJ:ssä käytetyt YSA-termit on käännetty HKK:n toimesta englanniksi ja saatu tätä kautta liitettyä osaksi YSO:a – alkuperäinen Kansalliskirjaston YSA-sanasto on vain kaksikielinen. Kokonaisuutta on käytetty Opas-järjestelmässä (Vehviläinen 2006, Vehviläinen et al., 2008), josta on tuotteistettu sovellus Kirjasto.fi-portaalin Kysy kirjastonhoitajalta -palveluun.<sup>32</sup>

### 5.2.5 Asiasanastoja SKOS-muodossa

Lisäksi FinnONTO:ssa on muutettu asiasanastoja SKOS-muotoon ilman erillistä ontologisoitua, ja julkaistu niitä ONKI palveluna taulukon 5.3 mukaisesti.

<sup>30</sup> <http://www.iconclass.nl/>

<sup>31</sup> <http://hklj.kirjastot.fi/>

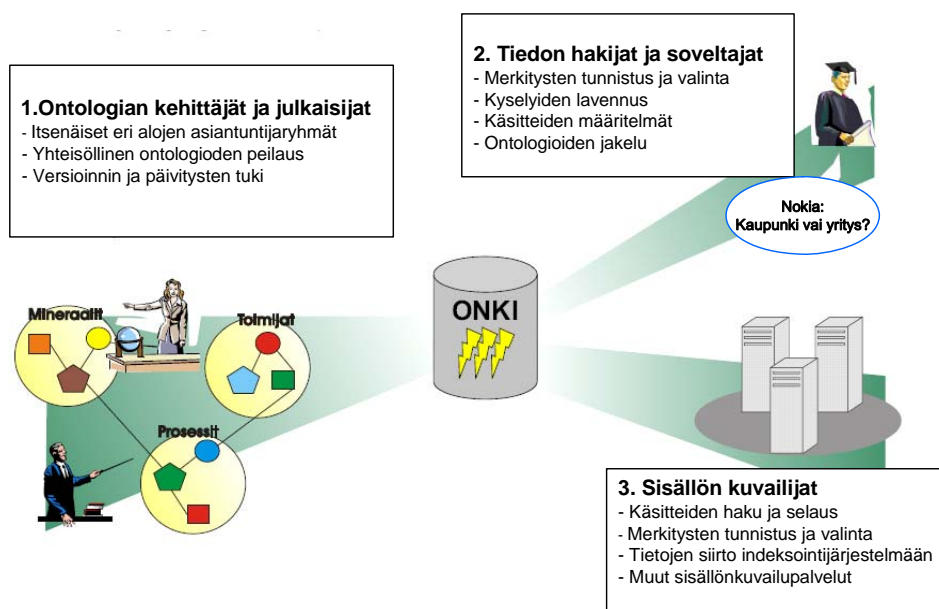
<sup>32</sup> <http://www.kirjastot.fi/fi-FI/tietopalvelu/>

Nimi	Sanasto ja alka	Lähde	Koko
YSA	Yleinen suomalainen asiasanasto	Kansalliskirjasto, <a href="http://vesa.lib.helsinki.fi/">http://vesa.lib.helsinki.fi/</a>	25600
NVAS	Valtioneuvoston asiasanasto	Valtioneuvoston kanslia	6300
Kaunokki	Kaunokirjallisuuden asiasanasto	<a href="http://www.kirjastot.fi/">http://www.kirjastot.fi/</a>	4400
MESA	Merenkulkualan asiasanasto	Merenkulkulaitos	1500
MUSA-CILLA	Musiikin asiasanasto	Kansalliskirjasto, <a href="http://vesa.lib.helsinki.fi/musa">http://vesa.lib.helsinki.fi/musa</a>	900

**Taulukko 5.3.** ONKI-palveluna julkaistuja SKOS-muotoisia asiasanastoja.

### 5.3 Kansallinen ontologiakirjastopalvelu ONKI

FinnONTO-vision mukaan pelkkä ontologioiden olemassaolo ei riitä kansalliseksi sisältöinfrastruktuuriksi, vaan lisäksi tarvitaan pysyviä palveluita, jotka mahdollistavat ontologioiden hyödyntämisen kustannustehokkaasti asiakasjärjestelmissä ja tukevat järjestelmän ylläpitoa ja kehittämistä.



**Kuva 5.4.** Kansallisen ontologiakirjastopalvelu ONKIn palvelut eri käyttäjäryhmille.

Ontologiat kehittyvät, lisääntyvät, yhdistyvät ja elävät ajan kuluessa, ja prosessit muuttuvat niitä kehittävien tahojen muuttuessa. FinnONTO-projektin ehdotus on käynnistää koordinoitu kansallinen ontologiakehitystyö nykyisten sanastojen kehittämishankkeiden luomalle perustalle. Aluksi ontologiatyötä voidaan tehdä FinnONTO-hankkeen toimesta ja rahoituksella pilotti-hankkeina. Tavoitteena on siirtää työn koordinointi hankkeen päättyessä tutkimushankkeelta siihen sitoutuneille

organisaatiolle, näillä näkymin vuoden 2009 tai 2010 lopussa. Uudet velvoitteet luonnollisesti edellyttäisivät työn jatkuvaa rahoitusta. Sanastotyö on maassamme perinteisesti ollut niukasti resursoitua, mutta tilanne parantunee semanttisen webin myötä selkeän hyödyntämisspolun syntyessä sanasto- ja sisältötyöstä käytännön verkkosovelluksiin.

Keskeinen uusi elementti kansallisesti koordinoitussa ontologiaprosessissa ovat uudet tekniset mahdollisuudet, joilla prosessia voidaan tukea ja saada sen tulokset kustannustehokkaasti hyödynnettäviksi eri sovelluksissa. FinnONTO-hankkeen tekninen ratkaisumalli ja alusta tähän on Kansallinen ontologiapalvelu ONKI. ONKI tarjoaa palveluita kolmelle käyttäjäryhmälle kuvan 5.4. mukaisesti:

1. Ontologian kehittäjille ja julkaisijoille on tarjolla keskitetty palvelu

<http://www.yso.fi/>

jonka kautta standardimuotoiset ontologiat, luokituksen ja sanastot voidaan julkaista välittömästi ja käytännössä ”ilmaiseksi” ilman lisätyötä: tyypillisesti tarvitaan vain ONKI SKOS -palvelimen pieni konfigurointi ja käynnistäminen. Tämän lisäksi palvelu tulee tarjoamaan yhteisöllistä Web 2.0 –henkistä tukea oman ontologian peilaamiseksi YSO:on tai sen alaontologioihin kansallista KOKO-ontologiaa varten sekä tuen ontologioiden versiointia ja päivityksiä varten.

**2. Tiedon hakijoille ja soveltajille** tarjotaan palveluja mm. hakusanojen merkitysten erotteluun. Ajatuksena on, että jos esimerkiksi Google-tyyppiseen hakukenttään kirjoittaa sanan ”johtaminen”, niin kenttään kytketty ONKI-palvelu voi tarjota semanttisen täydennyksen (Hyvönen, Mäkelä, 2006) avulla vaihtoehtoja sanan eri merkityksille johtaminen (sähkötekniikka), johtaminen (musiikin), johtaminen (matematiikassa), johtaminen (liiketaloudessa), jolloin eri valinnat voivat johtaa aivan erilaisiin tulosjoukkoihin. Toinen peruspalvelutyyppi, johon ontologioita voidaan käyttää, ovat erilaiset kyselyiden ontologiset laajentamiset (term expansion). Jos esimerkiksi hakusovelluksessa kysytään EU-maiden asioista, voidaan kysely laventaa paikkaontologian avulla TAI-kyselyksi Suomesta, Sloveniasta, Ranskasta jne. ilman että kysyjän tarvitsee tietää tai luetella kaikki EU-maat nimeltä. Tällaisesta palvelusta on tekeillä ensimmäinen demonstraatio Metsämuseo Luston Kantapuu.fi palveluun opetusministeriön rahoituksella.

ONKI tarjoaa soveltajille valmiita, mahdollisimman korkean jalostusasteen toiminnallisia palveluita käyttöliittymätasolla Web 2.0 leijukkeina (widget) (AJAX-teknologia) ja perinteisinä Web Service –palveluina. Asiakaspäässä tarvittavan koodin luomista varten on ONKI-palvelussa tarjolla helppokäyttöisiä interaktiivisia verkkopalveluja. Mikäli asiakas tarvitsee koko ontologian omaa sovellustaan varten, on sen ajantasaisen version lähdekoodi ladattavissa ONKI:sta vastaavaan tapaan kuin perinteisissä koodistopalvelimissa. ONKI-vision mukaan suurin etu kuitenkin saavutetaan valmiiden palveluiden avulla. ONKI-palvelun kautta on esimerkiksi mahdollista liittää kymmeniä

ontologioita tai sanastoja omaan sovellukseen valmiina toiminnallisuutena leijuke-valitsemella, joka syntyy yhdellä rivillä Javascript-koodia.

**3. Sisällön kuvailijoille** ONKI tarjoaa palvelut indeksointikäsitteiden vaivattomaan hakuun semanttisen täydentäjän (semantic autocompletion) (Hyvönen, Mäkelä, 2006) avulla, joka pyrkii arvaamaan jokaisen kirjoitetun merkin jälkeen, mitä käsitettä ollaan hakemassa. Monipuolisen selaimen avulla voidaan tarkastella käsitteen semanttista ympäristöä parhaan käsitteen merkityksen selvittämiseksi ja siihen liittyvien muiden käsitteiden tarkastelemiseksi. Semanttista selailua varten on tarjolla IRMA-väline (Sinkkilä, 2008; Sinkkilä et al., 2008), joka yhdistelee toisiinsa liittyviä käsitteitä ontologian eri puolilta - myös useamman eri ontologian välillä, Tavoitteena on helpottaa indeksoijan työtä parhaiden käsitevalintojen tekemissä.

ONKI-palvelimista on kehitetty kolme eri versiota erilaisia aineistoja varten:

- **ONKI SKOS** (Tuominen, 2008; Viljanen et al., 2008b) on tarkoitettu SKOS-muotoisten sanastojen ja yksinkertaisten OWL-ontologioiden julkaisemiseen.
- **ONKI Geo** (Lindroos, 2008) on paikkatiedon ontologiapalvelin.
- **ONKI People** on kehitetty henkilö- ja organisaatio -ontologioita varten (Kurki, 2008).

Palvelimet tarjoavat:

- **ihmiskäyttäjälle** haku- ja selainliittymän (browser) ja valitsinkomponentin
- **sovelluksille** AJAX ja Web Service –rajapinnat

Seuraavassa esitellään näitä palveluita tarkemmin käytännön esimerkkien avulla.

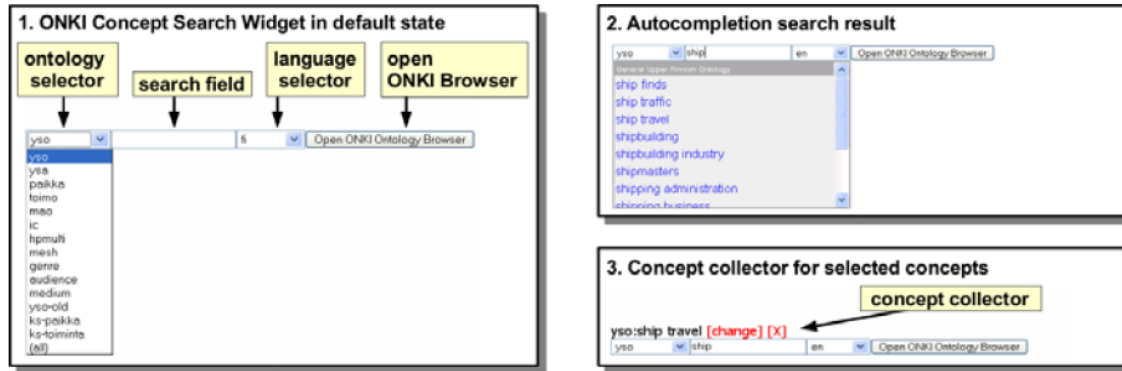
Ensin näytetään, miten ONKI-valitsin (selector widget) voidaan ottaa käyttöön kuvitteellisessa asiakassovelluksessa (Viljanen et al., 2008a). Tavoitteena on kehittää verkkosovelluksissa yleisesti käytetty HTML-lomake metatietojen syöttämistä varten valokuvista. Oletetaan esimerkiksi, että halutaan annotoida Helsingin sanomissa julkaistu valokuva nuorten autoilijoiden kolarista Pyhäjärvellä 5.9.2008. Valokuvien metadataskeema olkoon Dublin Core -standardin mukainen ja käytössä on kenttä dc:subject, joissa kerrotaan kuvan aihe (”nuoret” merkityksessä nuoret aikuiset; ”kolari” merkityksessä liikenneonnettomuus eikä Kolarin kunta), ja kenttä dc:coverage, jossa kerrotaan, missä kuva on otettu (yksi tietty Pyhäjärvi Suomen lukuisista Pyhäjärvistä paikkana eikä henkilönä).

ONKI voidaan kytkeä luettelointijärjestelmään luomalla metadatakentille - tässä kuvan aiheelle ja paikalle - syötekentät ja kytkemällä ne sopiviin ONKI-palvelimiin. Aiheen (asiasanojen) osalta syötekenttä voidaan kytkeä useampaakin ontologiaan ja paikkatiedon osalta paikkaontologiaan SUO. Tämä onnistuu lisäämällä ensin HTML-lomakkeen HEAD osioon viittaus ONKI-palvelussa olevaan Javascript kirjastoon vakiorivillä:

```
<script type="text/javascript" src="http://www.yso.fi/onki.js"></script>
```

Tämän jälkeen jokaiselle syötekentällä (input field) annetaan kytkentä valittuun ontologiaan. Esimerkiksi lomakkeen asiasanakentälle dc:subject voidaan muodostaa input-kenttä ja liittää se ONKI-ontologioihin seuraavalla rivillä:

```
<input id="dc:subject" onkeyup="onki['yso'].search()"/>
```



**Kuva 5.5.** ONKI-valitsin (widget). (Viljanen et al., 2008a)

Tuloksena syntyy lomakkeelle kuvan 5.5 mukainen syötekenttä (vasemmalla), joka jakautuu neljään osaan vasemmalta oikealle: 1) Vasemmalla olevasta ontologian valitsimesta voidaan valita käytettävä ontologia alavetovalikosta; oletusarvona on input-kentän parametrina annettu YSO. Haku voidaan kohdistaa myös samanaikaisesti kaikkiin eri ontologioihin. 2) Hakukenttään voidaan hakea käsitteitä valitu(i)sta ontologi(ois)ta semanttisella täydennyksellä (kuvassa oikealla ylhäällä). Kun valinta on tehty, käsitteeseen liittyvä tunniste ja nimike siirtyvät lomakkeelle hidden-kentän arvoksi, ja nimike näkyy käyttäjälle syötekentän yläpuolella käsitekeräimessä (collector) (kuvassa oikealla alla). Valittu käsite voidaan poistaa tai sitä muuttaa keräimen linkkejä painamalla. 3) kielivalitsemista voidaan valita käytettävä kieli – ONKI:n useimmat ontologiat ovat monikielisiä (jopa 12 eri kieltä). 4) Open ONKI Browser –napista voi käynnistää ONKI-selaimen, jolloin uuteen ikkunaan avautuu täysi selainnäkyä ontologian käsitteiden hakua ja selailua varten ”Nouda käsite” –painikkeella varustettua. Painiketta painamalla valittu käsite siirtyy selainikkunasta tietoineen lomakkeelle samalla tavalla kuin semanttista täydentäjää käyttämällä. Kaikki tämä keskitetyn palvelimen ajantasainen toiminnallisuus saadaan sovellukseen vain yhdellä rivillä koodia HTML -sivulla. Kuvassa 5.6. on esitetty tilanne, jossa käyttäjä on löytänyt ontologiasta liikenneonnettomuudet-käsitteen, nyt englannin kielellä. Nouda käsite –painike näkyy ikkunan oikeassa ylänurkassa.

Input-kentästä voidaan avata myös ONKI Geo -palvelimen karttapalvelu (Lindroos, 2008). Kirjoittamalla ”Pyhäjärvi” käyttäjä voi valita Suomen 49 Pyhäjärvestä oikean tulkinnan tai vaihtoehtoisesti kertoa onnettomuuden tarkan paikan kartalla, jolloin lomakkeelle saadaan koordinaattitiedot. Kuvassa 5.7 näkyy tilanne, jossa käyttäjä on Pyhäjärvi-nimen ohella rajannut hakua piirtämällä Pohjois-Suomen yllä rajatun alueen.

ONKI Geo -palvelu on ONKI valitsimen tapaan itsekin ns. mash-up-sovellus, joka hyödyntää Google Maps -palvelua karttanäyttöjen osalta.



**Kuva 5.6.** ONKI selain. YSO-ontologiasta on selaimella löydetty käsite liikenneonnettomuudet, kielenä englanti. Nouda käsite –nappia painamalla (oikealla ylhäällä) käsitteen nimi ja URI siirtyvät verkosta asiakassovellukseen..

Vastaavalla tavalla voidaan luoda input-kenttä paikkatiedon lisäämiseksi kuvaan. Tällöin

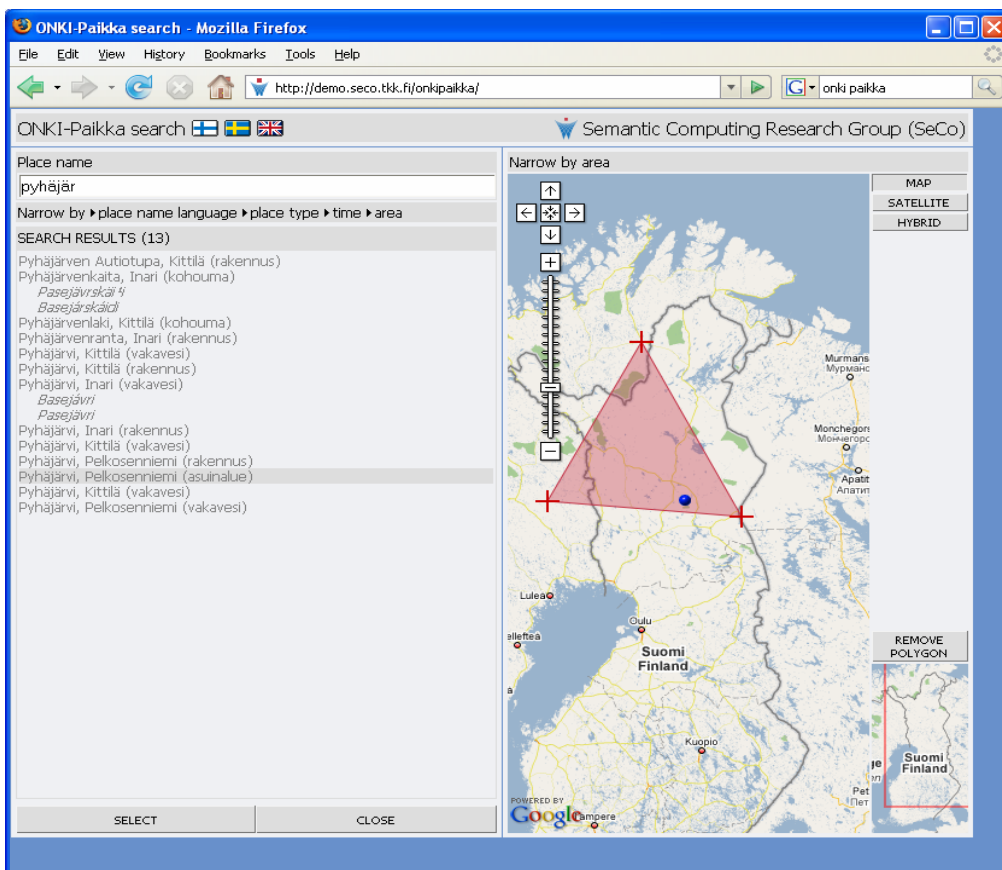
Esimerkki osoittaa, että ONKI:n avulla voidaan helpottaa merkittävästi vaikkapa luettelointisovelluksen ohjelmointia, jos halutaan saada käyttöön ajantasaiset sanasto- ja paikkatietopalvelut. Sovelluksen oma asia on päättää, miten ONKI-palvelimelta saadut tunnisteet, nimikkeet ja koordinaatit talletetaan esimerkiksi tietokantaan ja hyödynnetään. Yksilöivien käsitetunnisteiden kautta järjestelmällä kuvailut valokuvat voidaan automaattisesti kytkeä oikealla tavalla ilman sekaannuksia kansallisen semanttisen webin muihin sisältöihin.

ONKI-valitsinta voidaan sovellustilanteen tarpeiden mukaan konfiguroida monipuolisesti, esimerkiksi poistaa ontologian valintamahdollisuus, jos halutaan käyttää vain tiettyä ontologiaa, tai asettaa oletusarvo käytettävälle kielelle. Koodirivin automaattista muodostamista varten ONKI-palvelussa on erityinen interaktiivinen koodigeneraattori, jolla erilaisia valitsimia on helppo luoda ja kokeilla.

ONKI palvelut voidaan ottaa käyttöön myös perinteisen Web Service –rajapinnan kautta. Tällöin asiakasovellus kutsuu ONKI-palvelimen WSDL-rajapintaa HTTP- ja SOAP-protokollia käyttäen. SOAP-kutsujen ja WSDL-rajapinnan määritysten generointiin on tarjolla oma interaktiivinen palvelu.

Kulttuurisammossa ONKI-palveluita on hyödynnetty SAHA-editoriin yhdistettynä eräiden paikkatietoaineistojen annotoinnissa. Jatkossa erimuotoisten aineistojen

metatietorajapinnat on tarkoitus avata ulkoisille sisällöntuottajille Web 2.0 –hengessä vastaavaan tapaan. FinnONTO-ontologioita ja palveluita on alettu hyödyntää paitsi FinnONTO:n omissa projekteissa TerveSuomi ja Kulttuurisampo, myös muiden tahojen verkkosovelluksissa, kuten Kirjasto.fi-portaalin Kysy kirjastohoitajalta –palvelussa<sup>33</sup> ja yhteisöllisessä Tilkut-portaalissa<sup>34</sup>.



**Kuva 5.7.** ONKI Geo -palvelu yhdistää SUO paikkaontologian ja Google Maps –palvelun. Kuvassa käyttäjä etsii Suomen Pyhäjärvi-nimisiä paikkoja ja on rajoittanut hakuja Pohjois-Suomessa olevaan alueeseen. Tumma piste ilmaisee valitun paikan kartalla.

<sup>33</sup> <http://www.kirjastot.fi/fi-FI/tietopalvelu/>

<sup>34</sup> <http://owela.vtt.fi/tilkut/>

## 6 Kulttuurisampo portaali ihmisille ja koneille

---

### 6.1 Asiakaskäyttöliittymän perustoiminnot: haku ja selailu

Kulttuurisammon sisältöihin voi tutustua sekä hakemalla (search) kohteita että selailemalla (browse) siitä eri tavoin. Tyypillisessä käyttötapauksessa tehdään ensin hakua kohdejoukon rajaamiseksi, jonka jälkeen kohteita sekä niihin liittyviin asioihin aletaan tutustumaan selailun avulla.

#### Haku: pikahaku, nakymähaku ja muuta tavat

Portaalin vasemmassa ylänurkassa on aina käytettävissä yksinkertainen ”pikahaku” kenttä, jota käytetään perinteisten hakukoneiden tapaan kirjoittamalla siihen hakusanoja. Kulttuurisammossa haku tehdään kuitenkin ”semanttisesti” ontologioita hyödyntäen, joka näkyy käytännössä mm. seuraavilla tavoilla:

- *Semanttinen kyselyn täydennys.* Jokaisen kirjoitetun merkin jälkeen semanttinen täydentäjä yrittää ”arvata” ne mahdolliset sanamerkit, joihin käyttäjä pyrki ja näyttää koko ajan sen mukaista hakutulosta ruudulla.
- *Tulosten semanttinen ryhmittely.* Hakutulosten hahmottamisen avuksi ne voidaan ryhmitellä eri aineistotyyppisiin (maalaukset, biografiat, esineet jne.) sekä sen mukaan, millaisen semanttisen suhteen kautta objekti on löytynyt (esimerkiksi Akseli Gallen-Kallela taideteoksen kuvaamana kohteena, tekijänä, elämäkertansa henkilönä jne.
- *Tulosten semanttinen visualisointi.* Tulosjoukkoa voidaan vaivattomasti visualisoida kartoille, aikajalalle ja ontologisille luokituksille.
- *Monikielisyys.* Ontologiat ovat lähtökohtaisesti kieliriippumattomia, mikä mahdollistaa kieltenväliset kyselyt. Näin esimerkiksi suomenkielellä kuvatut ”tuolit” voidaan löytää ruotsin kielisellä hakusanalla ”stol” tai englanniksi ”chair”.

Kehittyneempiä semanttisia hakuja voidaan tehdä erillisessä Hae ja jäsenä –osiossa sekä eräissä muissa temaattisissa osioissa, joita selostetaan alla tarkemmin.

#### Selailu: semanttiset suosittelut

Portaalin yhtenä peruseriaatteena on: aina kun käyttäjä on tehnyt valinnan tutustua johonkin sisältökohteeseen, hänelle pyritään muodostamaan kohdetta laajempi näkemys aiheesta liittämällä kohteen sivuun suosittelusääntöjen avulla muita kohteita ja tietoa. Suosittelut sekä niihin liittyvät perustelut esitetään kohdesivujen alareunassa. Idea on sama kuin MuseoSuomessa, mutta Kulttuurisammon aineistot ovat huomattavasti rikkaampia ja monimuotoisempia ja tässä suhteessa vielä otollisempia semanttiseen suositteluun kuin MuseoSuomen esineet.

Suosituksia muodostuvat dynaamisesti ja täysin automaattisesti Kulttuurisammon kulloisenkin tietämiskannan perusteella ja perustuvat W3C:n standardoimalla SPARQL-kyselykielellä muodostettuihin suosittelusääntöihin. Tämä merkitsee kahta tärkeää seikkaa:



*Linkitys päivittyy automaattisesti.* Heti kun Kulttuurisampoon tulee uusi sisältökohde, linkittyy se automaattisesti kaikkiin muihin mahdollisiin sisältöihin sääntöjen määrittelemällä tavalla, ja muut sisällöt linkittyvät vastaavasti uuteen tiedon muruseen.

*Linkityslogiikkaa voidaan muuttaa helposti.* Sääntöjen muuttaminen ja lisääminen on hyvin helppoa verrattuna aiempaan MuseoSuomen logiikkaperustaiseen säännöstöön (Hyvönen et al., 2005; Viljanen 2006; Viljanen et al., 2006), jonka käyttäminen edellytti enemmän ohjelmointitaitoja. Tosin yksinkertaisuudesta on maksettava hintansa: SPARQL asettaa monia rajoituksia mutkikkaiden sääntöjen muodostamiselle. Koska aineisto on esitetty semanttisesti joustavalla tavalla RDF-muodossa, on mutkikkaiden sääntöjen luominen mahdollista ilman, että itse aineistoa tarvitsee muuttaa. Toki metadatan ja ontologioiden laatu asettaa reunaehdot sille, millainen suosittelu on semanttisesti mahdollista.

The screenshot shows the Kulttuurisampo website interface in a Mozilla Firefox browser. The page title is "Kulttuurisampo - Semanttisia näkökulmia suomalaiseen kulttuuriin". The browser address bar shows "http://demo.seco.tkk.fi/kulsa/". The website header includes navigation links: "Kieli: suomi", "Ohje", "Pääsivu", "Tietoa portaalista", and "Lähetä palautetta". Below the header is a search bar and a list of categories: "Pikahaku", "Jäsennetynä", "Karttahuja ja selailu", "Yhteyshaku", "Hae ja jäsennä", "Kokoelmat", "Suomen historia", "Taidot", "Elämäkerrat", "Kalevala", and "Karjala".

The main content area features a grid of thematic navigation tiles:

- Karttahuja ja selailu:** "Selaille nykyistä ja historiallisia paikkoja, alueita ja karttoja sekä niihin liittyvää esineistöä."
- Yhteyshaku:** "Kuinka Akeli Gallen-Kallela liittyy Ranskan keisariin Napoleonin ensimmäiseen?"
- Hae ja jäsennä:** "Aseta aineisto laajempaan kehukseen tarkastelemalla sitä eri näkökulmista."
- Kokoelmat:** "Mitä kokoelmia Kulttuurisampo on niellyt?"
- Suomen historia 1850-1919:** "Selaille historiallisia tapahtumia."
- Taidot:** "Perehdy semanttisesti kuvattuihin taitoihin."
- Elämäkerrat:** "Selaille semanttisesti rikastettuja elämäkertoja."
- Semanttinen Kalevala:** "Selaille Kalevalaa semanttisessa webissä."
- Tietoa Karjalasta:** "Wikipedian Karjala-aineistoa semanttisesti rikastettuna."

Below the grid, there is a section titled "Viimeksi katsottuja kohteita:" with several clickable items:

- museoesine : [Irtotasku, nainen](#)
- museoesine : [luomapuu](#)
- kaupunginosa : [Bemböle](#)
- museoesine : [peli](#)
- dokumenttielokuva : [Suutari Wirlander valmistaa nahkasaappaat](#)

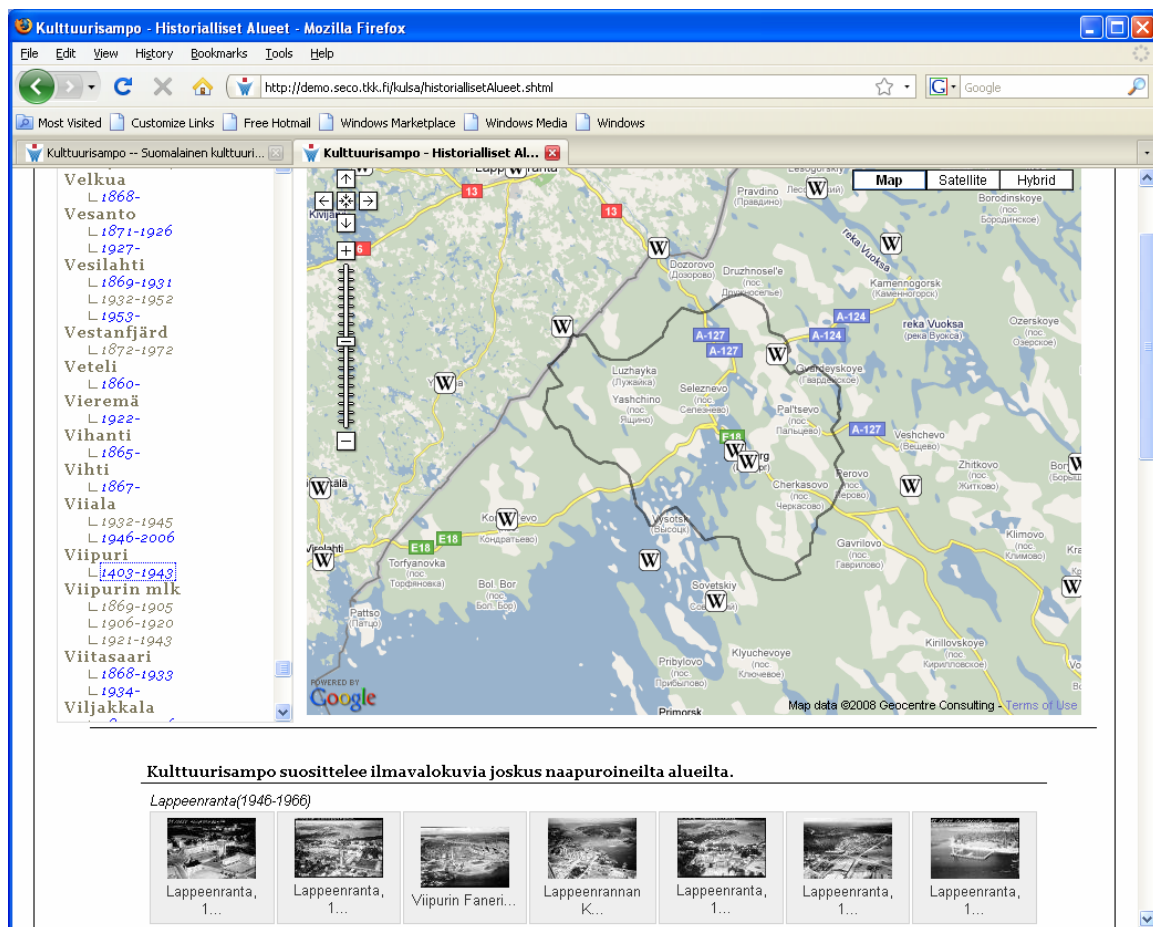
Two image thumbnails are visible below the links: one showing a wooden sculpture and another showing a book or document.

**Kuva 6.1.** Kulttuurisammon aloitussivu ja yhdeksän temaattista näkökulmaa.

## 6.2 Temaattiset näkökulmat

Kulttuurisammon asiakasliittymän uutena ideana on tarjota käyttäjälleen temaattisia näkökulmia suomalaiseen kulttuuriin. Näkökulmat ovat eräänlaisia silmälaseja, joiden kautta aineistot näyttäytyvät käyttäjälle eri valossa. Järjestelmän ensimmäisessä versiossa on toteutettu

yhdeksän näkökulmaa, jotka ovat esillä portaalin etusivulla kuvan 6.1 mukaisesti ja aina myöhemmin valittavissa yläpalkista. Seuraavassa esitellään lyhyesti eri näkökulmien ydinideat:



**Kuva 6.2** Historialliset alueet välilehdellä voi hakea Suomen historiallisia kuntia rajoineen, nähdä ei paikkoihin liittyviä Wikipedian artikkeleita ja tutustua eri paikoissa kansalaisten ottamiin valokuviiin. Kuvassa näkyy Viipuri 1403-1943 ja osa siihen liittyvistä semanttisista suositteeluista.

### Karttahuaku ja selailu

Tässä näkymässä sisältöihin voi tutustua Google Maps -karttojen avulla eri tavoin:

- **Hae esineitä kartalla** välilehdellä Kulttuurisammon kokoelmakohteet piirtyvät kartalle niihin liittyvien koordinaatti- tai muiden paikkatietojen perusteella. Tämä näkymän avulla voi esimerkiksi helposti löytää kulttuurikohteita omalta kotiseudullaan, esineitä joita on valmistettu tai käytetty siellä, tai paikallisia kansanrunoja ja –sävelmiä. Hakua voi rajata 11 eri semanttisen suhteen perusteella, jolla aineisto liittyy paikkoihin. Suhteita ovat esimerkiksi (kulttuurikohteen) ”kohteen sijainpaikka”, (valokuvan, maalauksen tms.) ”kuvaspaikka” tai (esineen tms.) valmistuspaikka.

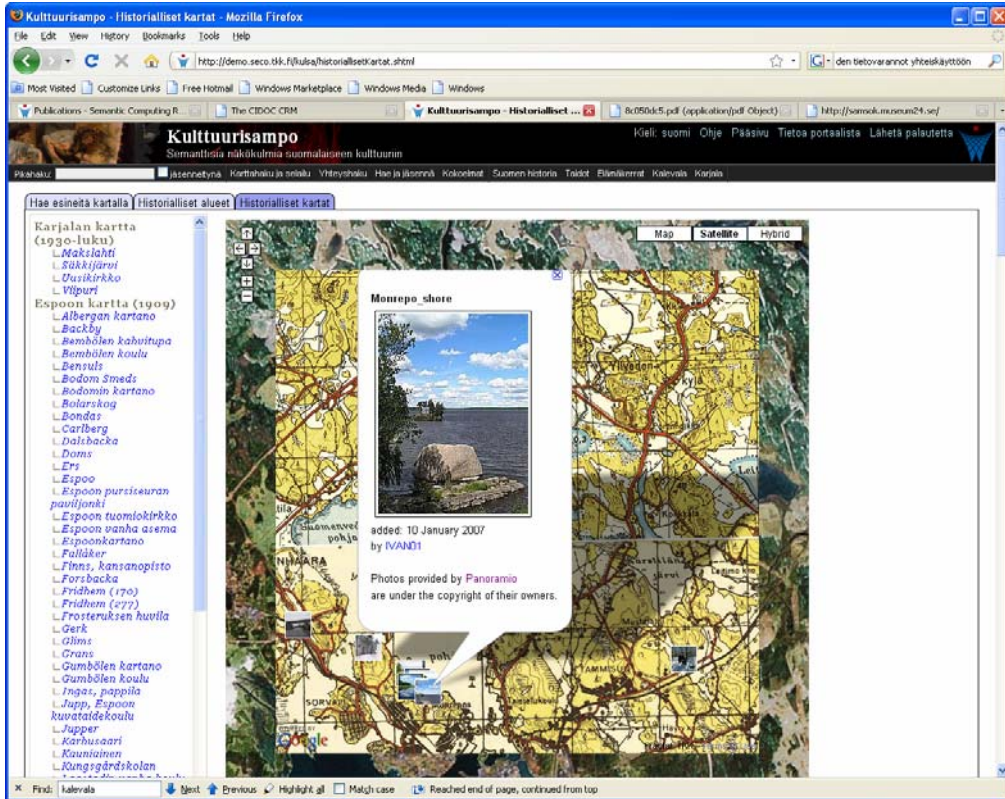
- **Historialliset alueet** välilehdellä voi hakea Suomen ajallisessa paikkaontologiassa SAPO mallinnettuja historiallisia ja nykyisiä alueita (tällä hetkellä Suomen kunnat 1860-luvulta alkaen). Demonstraatioon on Geologian tutkimuskeskuksen toimesta digitoitiin kuntien rajat polygoneina kahdelta aikajaksolta. Kuvan 7.1 käyttöliittymäkuvassa historialliset kunnat näkyvät vasemmalla olevassa valikkolistassa, jossa sinisinä linkkeinä näkyville alueille olemassa oleva polygoni. Kuvassa käyttäjä on valinnut historiallisen Viipurin ”Viipuri 1403-1943”, jonka rajat on visualisoitu kartalle. Samalla Kulttuurisampo tarjoaa käyttäjälle suositellulinkkejä alueeseen ja sen historiaan liittyen kartan alla, tässä kuvakaappauksessa näkyy Veljekset Karhumäki Oy:n vanhoja ilmavalokuvia läheisiltä alueilta. Tähän karttanäkymään on lisäksi projisoitu Wikipedia-aineistossa olevia artikkeleita, joissa kerrotaan kartalla näkyvistä paikoista, esimerkiksi Viipurin linnasta ja lähialueen paikoista kuten Ihantalasta ja Antreasta.
- **Historialliset kartat** välilehdellä voi tutustua historiallisiin alueisiin suhteuttamalla vanhoja karttoja nykyisiin Googlen karttoihin ja ilmakuviin. Järjestelmään on esimerkkinä digitoitu Maanmittauslaitoksen Karjalan vanhat kartat ennen toista maailmansotaa sekä Geologian tutkimuskeskuksen vanha geologinen kartta Espoon alueelta. Asiakas voi hakea paikkoja vanhojen paikannimien perusteella, joita nykyisissä kartoissa ei enää ole ja nähdä muutoksia, joita karttoja kuvaavilla alueilla on tapahtunut, esimerkiksi uuden kylän tai tien. Kuvassa 7.3 tarkastelun kohteena on historiallinen Viipuri. Keskellä oleva neliö on eräänlainen ”historiallinen ikkuna”, joka näyttää samanaikaisesti päällekkäin sekä nykyistä satelliittikuvaa että sotia edeltänyttä vanhaa karttaa Viipurin alueelta. Neliössä näkyvää tarunhoitoista nimeä ”Monrepos” ei nykyisistä kartoista enää välttämättä löydy. Lisäksi näkymään on liitetty kansainvälinen Panoramio-valokuvapalvelu. Karttapalvelun ideana on auttaa ihmistä hahmottamaan luovutetun Karjalan aluetta ja nimistöä suhteessa nykyiseen Venäjälle kuuluvaan alueeseen. Vasemmalla olevasta linkkistasta voidaan valita erilaisia kartalla näkyviä paikkoja katsottavaksi. Näytön alaosa löytyvät jälleen semanttiset suositellulinkit muihin aineistoihin.
- **Lähellä olevat kohteet** välilehdellä voi karttaa hiirellä osoittamalla hakea valittua paikkaa lähellä olevia kohteita.

## Yhteyshaku

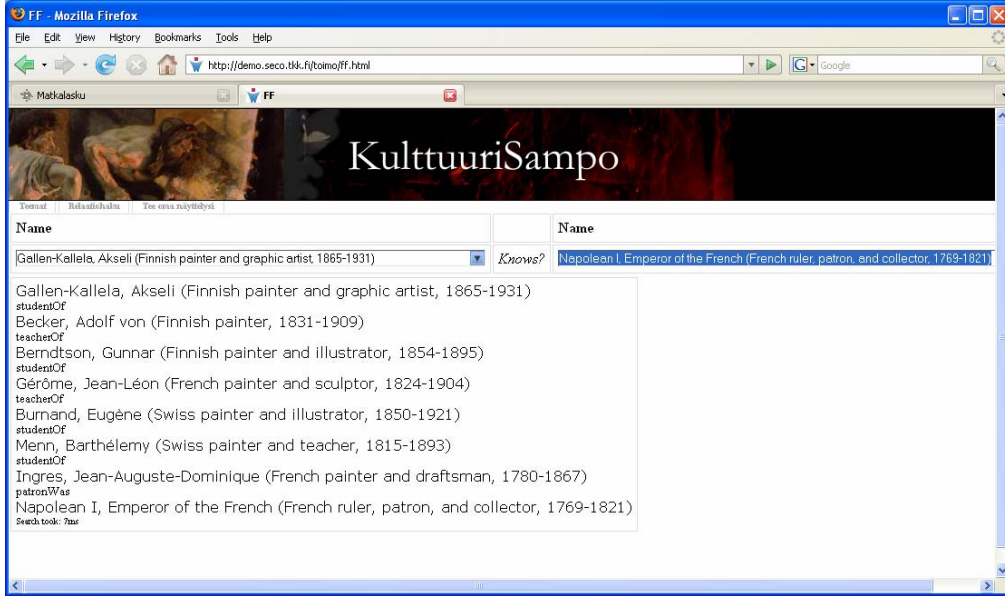
Kulttuurisammon yhteyshaun ideana on etsiä sisältökohteiden sijasta näiden välisiä kiinnostavia yhteyksiä. Järjestelmän aineistona on J. Paul Getty -säätiön n. 120 000 taiteilijaa ja kulttuurialan toimijaa sisältävä ULAN-tietokanta<sup>35</sup>. Tietokanta sisältää runsaasti tietoa mm. taiteilijoiden välisistä sosiaalisista suhteista, esimerkiksi kuka oli kenenkin lapsi tai oppilas. Aineisto on muutettu semanttisen webin muotoon. Yhteyshaussa käyttäjä kirjoittaa semanttisen täydentäjän avulla kaksi nimeä niille varattuihin kenttiin, jonka jälkeen järjestelmä kertoo miten valitut henkilöt liittyvät toisiinsa (Hyvönen, Kurki, 2007). Kuvassa käyttäjä on valinnut Akseli Gallen-Kallelan ja Ranskan keisari Napoleonin. Relatatiohaun tuloksena on seitsemän

<sup>35</sup> [http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)

ihmissuhteen polku, joka yhdistää Gallen-Kallelan eri aikaan eläneeseen Napoleoniin. Sen mukaan Gallen-Kallela oli Adolf von Beckerin oppilas, joka opetti Gunnar Berndtsonia, joka oli Jean-Leon Geromen oppilas, joka opetti Eugen Burnardia, joka oli Barthelemy Mennin oppilas, joka oli Jean-Auguste-Dominique Ingresin oppilas, jonka suojelija keisari Napoleon oli.



**Kuva 6.34** Käyttöesimerkki luovutetun Karjalan vanhojen karttojen (keskellä) ja Google Maps -palvelun yhdistämisestä kansainväliseen valokuvapalveluun. Vanha suomalainen kartta näkyy läpiluuhtavana nykyisen Googlen kartan (tässä ilmakuvan) päällä. Käyttäjä on löytänyt alueelta historiallisen Monrepos'n puiston Viipurista ja katsoo venäläisen valokuvaajan Panoramio-palvelun kautta julkaisemaa uutta valokuvaa puiston rannalta.



**Kuva 6.4.** KulttuuriSammon yhteyshaku on löytänyt seitsemän askeleen yhteyden Akseli Gallen-Kallelan ja Ranskan keisari Napoleonin välillä (Hyvönen, Kurki, 2007).

Kulttuurisammon yhteyshakua on tarkoitus jatkossa monipuolistaa hakemaan yhteyksiä vaikkapa paikkojen ja henkilöiden välillä. Systemiltä voisi silloin esimerkiksi kysyä miten Akseli Gallen-Kallela liittyy Pariisiin, jolloin vastauksena oli mm. tieto siitä että hän maalasi Ateneumin kokoelmiin kuuluvan Aino-triptyykin Pariisissa vuonna 1903.

Yhteyshaku on esimerkki Kulttuurisammon yleisestä filosofiasta, jossa hakumenetelmiä pyritään kehittämään kohti erilaisia ongelmanratkaisun menetelmiä. Asiakkaallehan haku ei ole primaarinen päämäärä vaan ainoastaan sekundaarinen keino yrittää tyydyttää tietotarpeita ja ratkaista taustalla olevia ongelmia.

## Hae ja jäsennä

Tämä näkökulma tarjoaa monipuolisen ns. semanttisen näkymähakupalvelun (faceted search, view-based search), jossa kohteita voidaan hakea ontologisesta luokituksista tehtyjen valintojen avulla hieman vastaavaa tapaan kuin MuseoSuomessa. Uusina ominaisuuksina ovat käytössä semanttinen tekstin täydennys ja aiempaa monipuolisemmat mahdollisuudet rajata hakuja eri roolien avulla, jotka Kulttuurisammon metatietoskeemoissa yhdistävät tietoja toisiinsa. Esimerkiksi Akseli Gallen-Kallela voi liittyä haettavaan kohteeseen maalauksen tekijänä, muotokuvassa olevan henkilönä, biografian kohteena olevan henkilönä, merkkijonona tekstissä jne. Uutta on myös mahdollisuus visualisoida hakutulos paitsi ontologisille näkymille, myös Googlen kartoille tai MIT:n Simile-projektissa kehitetylle aikajanelle<sup>36</sup>.

Käyttöliittymään on kehitteillä malli kaksikulotteisesta ”näyttelygeneraattorista”, jonka avulla voi esimerkiksi visualisoida miten Suomalaisissa kokoelmissa olevat Japanissa valmistetut esineet jakautuvat eri esinetyyppeihin eri aikakausina (Mäkelä et al., 2007).

<sup>36</sup> <http://code.google.com/p/simile-widgets/>

## Kokoelmat

Kokoelmat näkökulma tarjoilee Kulttuurisammon sisällöt organisaatiolähtöisesti hankkeessa mukana olevien museoiden, kirjastojen, arkistojen ym. sisällöntuottajien mukaan ryhmiteltynä. Jokaiselle organisaatiolle on automaattisesti tuotettu oma Kulttuurisampo-kotisivu, jota kautta sen omiin kokoelmiin ja normaalille kotisivulle pääsee vaivattomasti käsiksi.

Kulttuurisampo on ollaan jatkossa kehittämässä Web 2.0 hengessä yhteisölliseen suuntaan, jolloin sisällöntuottajiksi voisi tulla mukaan vaikkapa kotiseutuyhdistyksiä, erilaisia järjestöjä tai yksityisiä keräilijöitä. Ajatuksena on, että selkeän yksilöllisen näkyvyyden tarjoaminen valtakunnallisessa semanttisessa verkossa oman ”kotisivun” kautta kannustaa osaltaan kulttuurialan toimijoita sisältötalkoisiin.

## Suomen historia

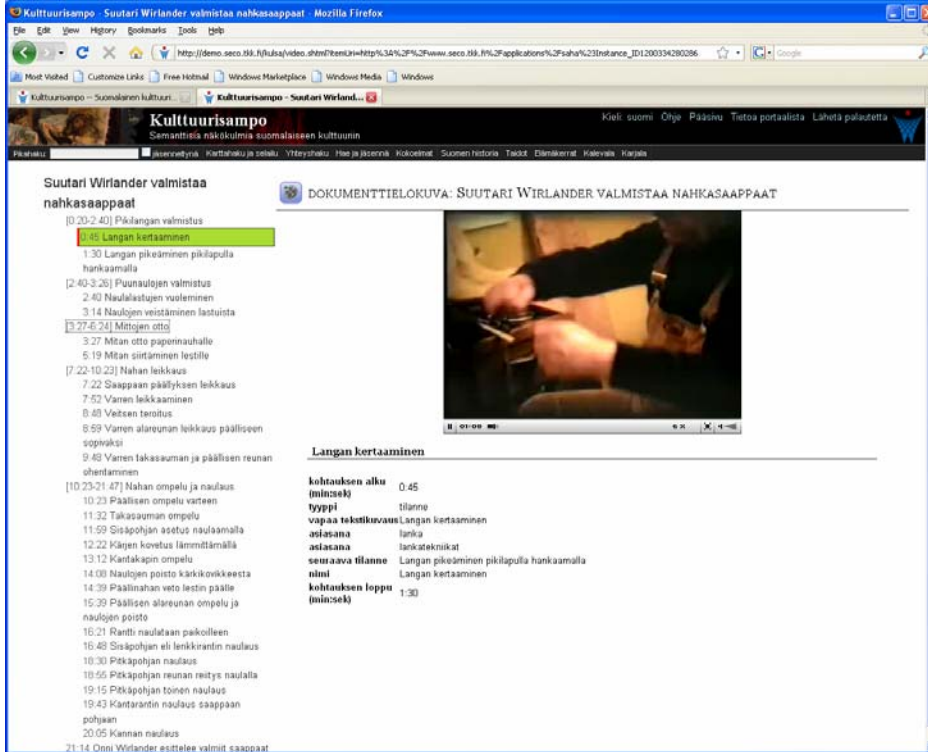
Tämän teeman keskiössä on aikajana ja sille projisoitu Suomen historiaontologian HISTO tapahtumat. Aineisto perustuu Suomen historiantutkijoiden yhteisön Agricola - Suomen historiaverkko -portaalissaan toteuttamaan aikajanaan<sup>37</sup>. Kulttuurisammossa tapahtumiin liittyy semanttinen sisällönkuvailu, jota kautta Kulttuurisammon historiallisiin tapahtumiin liittyviä sisältöjä voidaan näyttää historiallisesta näkökulmasta tarkasteltuna.

## Taidot

Taidot-näkökulman museologisenä ideana on ajatus siitä, että museoissa pitäisi ryhtyä tallentamaan systemaattisesti luetteloituja tietoja immateriaalisista ilmiöistä kuten käsityötaidoista (Kettula, 2006). Kulttuurisammossa ”taidot” voivat olla yleisiä kuvauksia kulttuurisista prosesseista tai tarinoista (Junnila et al., 2007). Prototyypissä on tästä kaksi esimerkkiä: Maatalousmuseo Sarassa toimitettu prosessikuvaus vuoden kierrosta maatilalla, jonka tavoitteena on kertoa maatalouskulttuurin ilmiöistä kaupunkilaistuvalla väestölle, ja Taideteollisessa korkeakoulussa laadittu kuvaus keramiikan valmistuksesta, jota voidaan käyttää vaikkapa opetuksen apuna. Semanttisen prosessikuvauksen perusteella Kulttuurisampo osaa automaattisesti luoda prosessin visualisoivan sivun ja yhdistää siihen aineistoja suosittelusäännöllä, esimerkiksi posliinisia esineitä tai työkaluja museoista keramiikan valmistuksen eri vaiheisiin.

Toinen kulttuurisammon taitotyyppi on yksittäisen kulttuurisen prosessin dokumentaatio videona. Ainoana esimerkkinä tästä on Espoon kaupunginmuseon tuottama video ja sen semanttinen prosessikuvaus, jossa suutari Wirlander valmistaa nahkasaappaat kuvassa 7.5. Videoita voidaan katsella portaalissa prosessikuvauksesta automaattisesti tuotetun sisällysluettelon avulla. Kun YouTubeen kaltaisissa palveluissa videot ovat kokonaisuuksia, voi haku Kulttuurisammossa kohdistua myös videon eri osiin, jotka voidaan käynnistää erikseen katseltaviksi. Kulttuurisammon tuottamat semanttisen suosittelut muihin aineistoihin voivat vaihtua kohdan mukaan samalla kun videota katsellaan.

<sup>37</sup> <http://agricola.utu.fi/hist/kronologia/index.php>



**Kuva 6.5** Semanttista videonkatselua Kulttuurisammossa. Suutari Wirlander valmistaa nahkasaappaat Espoossa.

## Elämäkerrat

Yksi perinteinen näkökulma kulttuuriin muodostuu suurmiesten ja muiden henkilöiden elämän kautta. Kulttuurisammossa tämä näkökulma on toteutettu Suomalaisen kirjallisuuden seuran Kansallisbiografian verkossa olevien elämäkertojen kautta. Ideana on, että kun asiakas lukee elämäkertaa, voidaan siinä esiintyviin käsitteisiin ja hahmoihin tuottaa automaattisesti semanttisesti perusteltuja suosittelulinkkejä muihin Kulttuurisammon aineistoihin. Esimerkiksi tekstissä esiintyviä paikkoja voidaan näyttää kartalla, liittää niihin ilmavalokuvia, henkilön luoma taideteos saattaa löytyä kokoelmista, henkilön sosiaalinen verkosto voidaan visualisoida jne.

## Semanttinen Kalevala

Kalevala on suomalaisen kulttuurin keskiössä semanttisen linkittymisen kannalta ja valittiin alusta asti yhdeksi Kulttuurisammon teemaksi. Lisäksi koko järjestelmä nimettiin Kalevalan salaperäisen sammon mukaan, joka usein mielletään kaikkea arvokasta tuottavaksi mystiseksi koneeksi. Tämä on Kulttuurisammonkin tavoite.

Jo Kulttuurisammon ensimmäisessä versiossa v. 2005 kuvailtiin Kalevalasta pari runoa hyvinkin yksityiskohtaisella tapahtumamallilla. Semanttisen Kalevalan ideana (Hyvönen et al., 2007c) on jakaa narratiivinen teksti sellaisiksi tapahtumajaksoiksi, jotka ovat mielekkäitä itsenäisiä kokonaisuuksia. Jaksot vastaavat eräissä Kalevalan painoksissa olevia runojen

sisällysluettelon kohtia ja muodostavat tapahtumien mukaisia laajempia jaksoja. Jaksot ovat Kulttuurisammon hakukohteita ja niitä on annotoitu semanttisesti. Ajatuksena on, että kun asiakas lukee Kalevalaa, voidaan kuhunkin jaksoon tuottaa suosittelulinkkejä. Jaksoon voidaan esimerkiksi linkittää Suomen Kansan Vanhoja Runoja, joita Lönnrot käytti, tai näyttää kulttuurihistoriallisissa kokoelmissa olevia vanhoja suksia, jos jaksossa kerrotaan hirven hiihtämisestä. Lisäksi jaksoon voidaan liittää muutakin metatietoa, esimerkiksi kielellisesti vaikean kohdan tulkintaa nykykielellä.

## Tietoa Karjalasta

Kulttuurisammon yhtenä sisällöllisenä fokusalueena on ollut kerätä aineistoja liittyen Karjalaan, Kalevalaan ja 1800-luvun maataloudesta eläneeseen Suomeen. Aineistojen valinnalla on pyritty edistämään niiden keskinäistä linkittymistä. Luovutetun Karjalan esineistö kulkeutui ja hajasijoitettiin toisessa maailmansodassa evakuoinnin yhteydessä useaan maamme eri museoon, joita on nyt mukana sisällöntuottajina KulttuuriSampo-hankkeessa opetusministeriön tuella Antikvaria-ryhmän kautta. Web tarjoaa julkaisukanavan, jonka avulla hajasijoitetut Karjalan kulttuurisisällöt voidaan koota verkon avulla virtuaalisesti yhteen. Näin Karjala saadaan takaisin semanttisessa webissä. Vastaavanlainen kulttuurisisältöjen hajautuminen eri museoihin ja maihin on hyvin yleistä ja ongelmallista mm. Keski-Euroopassa, puhumattakaan entisistä siirtomaista, joiden kulttuurihistoriallisesti arvokkain esineistö löytyy usein entisten emämaiden suurista museoista.

Kulttuurisammon Karjala-aineisto sisältää museoiden kokoelmia Karjalasta (mm. Viipurista), vanhoja maanmittauslaitoksen Karjalan karttoja, Wikipediassa olevia Karjala-aiheisia tekstisivuja ja kartoille projisoituja paikkasivuja. Tietoa Karjalasta teemassa aineistoa tarkastellaan Karjala-aiheisten Wikipedia-sivujen kautta, joita linkitetään Kulttuurisammon muihin sisältöihin.

Ehkäpä joskus tulevaisuudessa Kulttuurisampoon voidaan semanttisen webin kautta yhdistää myös nykyisissä venäläisissä museoissa olevia Karjala-aineistoja.

### 6.3 Portaalin käyttö ulkopuolelta palveluina: leijukkeet

Kulttuurisampo ei ole tarkoitettu vain ihmiskäyttäjille vaan portaalin palveluita voivat käyttää myös toiset verkossa olevat sovellukset. Kulttuurisampo ei ole vain yhteinen julkaisualusta, johon sisällöntuottajat luovuttavat sisältöjään. Ideana on samalla myös rikastaa omia sisältöjä muiden toimijoiden aineistoilla ilmaiseksi ja hyödyntää syntynyttä uutta kokonaisuutta myös Kulttuurisammon ulkopuolella, esimerkiksi museon omilla verkkosivuilla. Kulttuurisampoon on siksi kehitteillä sisällöntuotantokanavan ohella myös ”paluukanava” portaalista sisällöntuottajille, jota kautta semanttisesti rikastuneet aineistot voidaan käyttöön sisällöntuottajien ja muiden organisaatioiden verkkosovelluksissa vaivattomasti ns. semanttisina leijukkeina (widget).

FinnONTO-hankkeessa kehitetyn leijuke-konseptin idea on hieman vastaava kuin Googlen AsSense-mainospalvelussa, jossa mille tahansa verkkosivulle voidaan tuoda mainoslinkkejä sivun sisältöön perustuen lisäämällä haluttuun sivun kohtaan HTML-koodin sekaan yksinkertainen Javascript-koodi. Kulttuurisammon leijukerajapinnan avulla voidaan vastaavalla



tavalla tuottaa mielivaltaiselle verkkosivuille linkkisuosituksia tai hakutuloksia Kulttuurisampo-portaalista. Yhtenä käyttöesimerkkinä on kaupungin matkailusivu, jolle voidaan automaattisesti tuottaa linkkejä kaupungin alueella oleviin kulttuurikohteisiin. Vastaavaan tapaan museon tai lehden tiettyä taiteilijaa käsittelevälle sivulle voidaan tuoda linkkejä henkilön luomiin taideteoksiin eri museoissa, yhdistää sivu taiteilijan elämäkerrallisiin tietoihin jne.

Leijukkeen tarjoaman palvelun käyttö asiakassovelluksessa on erittäin yksinkertaista eikä edellytä varsinaista sovellusohjelmointi – vain rivi Javascript-koodia tarvitaan Kulttuurisammon valmiin toiminnallisuuden yhdistämiseksi omaan mash-up sovellukseen. Tätä ideaa havainnollistava demonstraatio leijukkeen käytöstä, jossa Suomi.fi-portaalin sivulle on tuotettu aineistoja TerveSuomi-portaalista, on kokeiltavissa verkkossa<sup>38</sup>.

---

<sup>38</sup> <http://www.seco.tkk.fi/applications/tervesuomi/>

## 7 Yhteenveto: miten tästä eteenpäin?

---

Muistiorganisaatioiden yhteisportaalien kehittämisessä on runsaasti teknisiä ja käytännöllisiä haasteita. Loppukäyttäjän kannalta avainkysymyksiä ovat 1) riittävän monipuolisen ja hyödyllisen sisällön saaminen portaaliin ja se, 2) miten asiakasliittymästä saadaan aidosti mielenkiintoinen, hyödyllinen ja helppokäyttöinen. Kulttuurisampo hankkeessa sisältökysymystä on ratkaistu

- ottamalla aineistotyypeiksi lähtökohtaisesti kaikki mahdollinen kulttuurisisältö esineistä kulttuuripaikkoihin ja aineettomiin prosesseihin sekä tarinoihin,
- kokoamalla hankkeen sisällöntuottajiksi mukaan poikkeuksellisen laaja, erilaisia näkökulmia edustava konsortio muistiorganisaatioita,
- laajentamalla sisällöntuotantoyhteisö klassisten muistiorganisaatioiden ulkopuolelle muihin kulttuurisisältöjä tuottaviin organisaatioihin (esim. Maanmittauslaitos, Geologian tutkimuskeskus, Yleisradio),
- hyödyntämällä sisältönä nopeasti kasvavia verkkoyhteisöjen aineistoja ja –palveluita (esim. Agricola – Suomen historiaverkko, Wikipedia, Geonames)
- kehittämällä sisällöntuotantoprosessia varten ontologiaperustainen infrastruktuuri ja palveluita (Kansallinen ontologiapalvelu ONKI) sekä erilaisia työkalu (esim. muunnosohjelmia olemassa oleviin aineistovarantoihin ja annotoinnin apuvälineitä).

Älykkään asiakasliittymän ongelmaa taas on lähestytty

- tutkimalla ja kehittämällä uusia ”älykkäitä” haku- ja selailuparadigmoja (esim. semanttinen täydennys, näkymäperustaisen haun erilaisia muotoja, suosittelujärjestelmät),
- kehittämällä ja hyödyntämällä uusia visualisointimenetelmiä (esim. dynaamiset kartat, aikajanat, sosiaaliset verkostot, puunäkymät ja jäsentelyt)
- käyttämällä selainkäyttöliittymässä uusinta Web 2.0 AJAX-tekniologiaa.

### 7.1 Sisällöntuotannon mahdollisuudet ja haasteet

Kokemustemme perusteella ehkä suurin käytännön haaste Kulttuurisammon kaltaisessa hankkeessa liittyy tietosisältöjen tuotannon systematisointiin ja harmonisointiin. Kysymys on yksinkertaisesti siitä, miten museoissa, arkistoissa, kirjastoissa ja muissa sisältöä tuottavissa organisaatioissa voitaisiin tuottaa tehokkaasti metatietoa yhteismitallisessa, semanttisesti tulkittavissa olevassa muodossa niin, ettei sisältöjä tarvitsisi myöhemmin korjailta käsin. Nykyinen asiansanoitukseen ja vapaisiin tekstikenttiin perustuva kuvailu ei ole riittävän täsmällistä tietokoneen käytettäväksi, vaan tarvitsee tulkikseen ihmistä. Uusissa Web 2.0 -sovelluksissa käyttöön otetulla vapaalla ”tägäyksellä” on sama ongelma: eri henkilöt käyttävät samoja termejä eri merkityksissä eikä ”tägien” välisiä suhteita huomioida. Esimerkiksi ”lasilla” voidaan viitata joko ikkunaan, materiaaliin, linssiin, juomalasiin tms. Seurauksena on

semanttinen sekaannus, joka suuremmilla ja alojen välisillä (cross-domain) aineistoilla muodostuu yhä häiritsevämmäksi estäen täsmällisen tietämyksen hallinnan.

Ongelmaan on kaksi periaatteellista ratkaisutapaa.

- **Portaalivetoinen malli.** Sisällöntuotannossa jatketaan nykyiseen malliin ja yritetään täsmentää ja korjata aineistoja yhteisportaalin puolella myöhemmin. Toki kaikki kannattavat ainakin juhlapuheissa luettelointikäytäntöjen kehittämistä ja harmonisointia. Tämä malli on sikäli houkutteleva resurssipulan vaivaamille muistiorganisaatioille, että mitään ei tarvitse tehdä itse: aineistojen ontologinen tarkennus ja siivous voidaan jättää portaalin ylläpitäjien murheeksi. Etuna on, että kustannukset syntyvät vasta aineistojen julkaisuvaiheessa, ja toivon mukaan toinen taho, ts. portaaliorganisaatio, vielä hoitaa ne.
- **Yhteisöllinen malli.** Toinen strategia on, että sisällöntuotannon prosesseja sisältöä tuottavissa organisaatioissa aletaan systemaattisesti harmonisoida ja kehittää semanttiseen suuntaan niin, että aineisto saadaan julkaisukuntoon jo luettelointivaiheessa. Tällöin aineistot voidaan parhaimmillaan julkaista yhteisportaalissa nappia painamalla ilman lisätyötä ja -kustannuksia portaalin ylläpitäjille. Esimerkkinä tästä lähestymistavasta on kansallinen TerveSuomi-portaali<sup>39</sup>.

Useimpien portaalihankkeiden lähtökohtana on strategia (1). Tämä strategia on myös ainoa mahdollisuus silloin, kuin palvelun tarjoaja ei voi vaikuttaa sisällöntuotantoon, esimerkiksi vaikka Googlen kaltaiset yleiset hakukoneet. Muistiorganisaatioiden yhteisportaali on Kulttuurisammon vision mukaan kuitenkin yhteisöllinen Web 2.0 -järjestelmä, jonka jäseninä on ensivaiheessa ollut vakiintuneita kansallisia organisaatioita. Olennaista tällaisen järjestelmän toimivuudelle on yhteinen tahtotila semanttisen yhteispalvelun kehittämiseksi, jonka ylläpito ei ole ainoastaan portaalipalvelun kehittäjien intressissä, kuten Googlen tapauksessa. Lähtökohtana Kulttuurisammossa on ollut strategia (2), mutta aineistotyypistä ja tilanteesta riippuen sovelletaan joustavasti myös strategiaa (1).

Yhteisöllisen mallin perustana oleva tahtotila voi syntyä vain silloin, kun järjestelmään sisältyy riittävän motivoiva sisäinen ”palkitsemisjärjestelmä”. Kulttuurisammon palkitsemisjärjestelmä antaa osallistujille seuraavia etuja:

**Sisältöjen parempi löydettävyys.** Kokoelmakohteille saadaan parempi *näkyvyys ja löydettävyys* kansallisen portaalin kautta. Tämä etu ei liity vain portaalin sisällä tapahtuviin hakuihin. Kulttuurisammon kohteet indeksoituvat yleisiin hakukoneisiin ja tulevat korkean linkittymisasteensa takia nousemaan esimerkiksi Googlen page ranking -algoritmissa korkeammalle kuin museoiden omilla sivuilla. Esimerkiksi MuseoSuomessa n. 2/3 asiakkaista tulee portaaliin perinteisten hakukoneiden kautta.

**Oman näkyvyyden paraneminen.** Organisaatiolle ja muille sisällöntuottajille saadaan *omaa yksilöllistä näkyvyyttä* omien kotisivujen kaltaisten ratkaisujen kautta.

<sup>39</sup> <http://www.tervesuomi.fi/>

**Yhteistyön tulosten kustannustehokas hyödyntäminen.** Jokainen voi itse hyödyntää yhteistyöllä semanttisesti rikastettua Kulttuurisammon tietämuskantaa leijukkeiden kaltaisina kustannustehokkaina palveluina. Tämä tehostaa sekä muistiorganisaatioiden omaa toimintaa, että lisää jatkossa aineistojen näkyvyyttä muiden kulttuurisisältöä tarvitsevien portaalien kautta (esim. turismiportaalit, mobiilit navigointipalvelut, verkkolehdet jne.).

**Valjastamattomien yhteisöllisten voimavarojen hyödyntäminen.** Järjestöjen ja kansalaisten sisällöntuotanto on jatkossa tarkoitus kytkeä yhä voimakkaammin mukaan portaalien Web 2.0 -hengessä. Kuten esimerkiksi Wikipedia on osoittanut, Web 2.0 –tyyppisessä yhteisöllisyydessä piilee valtavia uusia mahdollisuuksia sisältöjen tuotannolle ilman merkittäviä lisäkustannuksia. Kulttuurisammon ensimmäisessä prototyypissä tämä on mahdollista kirjoittamalla Wikipediaan artikkeleita paikoista ja liittämällä näihin koordinaattitietoja. Systemissä on myös jo rakennettuna valmius ottaa vastaan asiakkaiden kommentteja ja lisätietoja sisältökohteista.

Kulttuurisampo-hankkeen tavoitteena on, että näiden etujen siivittämänä maahamme syntyy kansallinen, yhteisöllinen kulttuurialan semanttinen portaalit, jossa kohtaavat sekä kulttuurisisältöjen tuottajat että kuluttajat molemmissa rooleissaan kulloistensa mukaan.

Käsitykseni mukaan vision toteutumisen edellytyksenä on, ettei vastuuta aineistojen semanttisen sisällönkuvaflusta säilytetä vain kansallisen portaalien ylläpitäjien kontrolle strategian (1) mukaisesti pidemmällä tähtäimellä. Kestävällä perustalla oleva kehitys edellyttää pääsääntöisesti strategian (2) käyttöönottoa, ts. että jatkossa sisältöjen semantiikan harmonisointikysymyksiin tartutaan määrätietoisesti yhteisen tavoitteen saavuttamiseksi. Muuten tuloksena syntyy joko ylläpitokustannuksiltaan liian kallis portaalit tai semantiikaltaan ja toiminnallisuudeltaan liian köyhä järjestelmä.

Kulttuurisammon kaltaisessa semanttisessa portaalissa tarvittava yhteentoimiva sisällönkuvaflut edellyttää käytännössä vain hieman aiempaa täsmällisempää sopimista metatietokäytänteistä kokoelmien luetteloinnissa (skeemaintegraatio) ja semanttisen webin ontologioihin liittyvien käsitteiden (tai muuten merkityksiä yksilöivien tunnisteiden) käyttämistä kuvaflussa vapaamuotoisten tekstien sijasta (arvointegraatio). Välttämättä tämä ei merkitse suuria muutoksia tai kustannuksia, ja FinnONTO-hankkeessa kehitetty kansallinen ontologinen sisältöinfrastruktuuri ONKI-palveluineen ja tekniset apuvälineet ovat jo nyt käytettävissä.

Vakiintuneiden ja kohtuullisesti toimimien käytäntöjen muuttamiseen liittyy kuitenkin aina sekä teknisiä että poliittisia kysymyksiä, muutosvastarintaa, insinööri- ja humanististen tieteiden välisiä jännitteitä, eikä niitä ole syytä aliarvioida. Joissain tapauksissa pienetkin muutokset voivat aiheuttaa yhteentoimivuuteen liittyviä ongelmia organisaatioiden sisällä esimerkiksi aiemmin kuvailtujen sisältöjen ja tietojärjestelmien kanssa. Työ kannattaa siksi aloittaa poimimalla matalalla olevat hedelmät ensin, ts. niiden toimijoiden osalta, joilla on merkittäviä kansallisia sisältöjä hallussaan, tekninen valmius muutoksiin ja rohkeutta ottaa vastaan uuden ja nopeasti kehittyvän web-teknologian mahdollisuudet ja haasteet.

## 7.2 Kansallinen digitaalinen kirjasto ja Kulttuurisampo

Uusi kansallinen panostus muistiorganisaatioiden kokoelmien digitaaliseen yhteisjulkaisuun on Opetusministeriön Kansallinen digitaalinen kirjasto –hanke (KDK), joka on tarkoitus suunnitella ja toteuttaa vuosina 2008-2010<sup>40</sup>. Sen tavoitteet ovat:

1. yhteisen hakupalvelun kehittäminen kirjastojen, arkistojen ja museoiden keskeisille sähköisille tietovarannoille
2. kirjastojen, arkistojen ja museoiden keskeisten, priorisoitujen aineistojen digitointi ja liittäminen osaksi yhteistä hakupalvelua
3. sähköisten kulttuuriperintöaineistojen pitkäaikaissäilytysratkaisujen tuottaminen
4. kulttuuri- ja kulttuuriperintöaineistojen digitointia ja sähköistä saatavuutta sekä sähköisten kulttuuriperintöaineistojen pitkäaikaissäilytystä koskevan osaamisen lisääminen

Kansallinen digitaalinen kirjasto on tarkoitus yhdistää ruotsalaisen SamSök-järjestelmän tapaan yhtenä sisällöntuottajana eurooppalaiseen Europeanaan. Europeanan ensimmäinen versio on valmistumassa marraskuussa 2008. Suomesta hankkeessa ovat mukana Kansallisarkisto ja Kansalliskirjasto.

Kansallinen digitaalinen kirjasto -hanketta tulee rahoittamaan Opetusministeriö. Kulttuurisammon kehitys puolestaan lähti liikkeelle teknologian tutkimuksen haasteista, joihin kulttuurisisältöjen semanttisen verkkojulkaisemisen haasteet tarjosivat erinomaisen tutkimus- ja kehitysalustan. Työn tärkein rahoituskanava löytyi siksi Tekesin, muistiorganisaatioiden ja yritysmaailman suunnalta.

Hankkeilla on voimakas synergia kansallisella tasolla, joka pitäisi jatkossa osata hyödyntää. Onhan Kulttuurisampo-hankkeessa jo valmistunut prototyyppi Kansallinen digitaalinen kirjasto –hankkeen tavoitteesta (1) toteuttaa muistiorganisaatioiden yhteinen hakupalvelu. Kulttuurisammon ratkaisumalli ja toiminnallisuus on laajempi kuin mitä KDK-hankkeen suunnitelmissa on alustavasti kaavailtu. Kulttuurisammossa on portaaliratkaisun ohella kehitetty ratkaisumalli sisällöntuotannolle kansallisen FinnONTO-sisältöinfrastruktuurin pohjalta. Nämä ratkaisut liittyvät osittain Kansallisen digitaalisen kirjaston tavoitteeseen (2). Kolmanteen tavoitteeseen, digitoinnin priorisointiin tai pitkäaikaissäilytyksen ratkaisuihin (3), Kulttuurisampo-hankkeessa ei ole otettu lainkaan kantaa. Tavoitteen (4) osalta oppia on saatu koko FinnONTO-hankkeen ajan ja sitä on dokumentoitu kymmeniksi opinnäytetöiksi ja kansainvälisiksi julkaisuksi, jotka kaikki löytyvät sähköisessä muodossa Semanttisen laskennan tutkimusryhmän julkaisujen kotisivulta <http://www.seco.tkk.fi/publications/>.

<sup>40</sup>[http://www.minedu.fi/OPM/Kulttuuri/kulttuuripolitiikka/linjaukset\\_ohjelmat\\_ja\\_hankkeet/digitaalinen\\_kirjasto/index.html](http://www.minedu.fi/OPM/Kulttuuri/kulttuuripolitiikka/linjaukset_ohjelmat_ja_hankkeet/digitaalinen_kirjasto/index.html)

### 7.3 Miten tästä eteenpäin?

Webin tulevaisuus on semanttinen. Käsitteiden mukaan suomalainen Kansallinen digitaalinen kirjasto kannattaa siksi toteuttaa W3C:n semanttisen webin standardien ja periaatteiden mukaisesti MuseoSuomen ja Kulttuurisammon viitoittamalla tiellä. Toki järjestelmä voidaan tehdä perinteisin menetelminkin, mutta silloin Kulttuurisammossa jo tehdyn työn hyödyntäminen ei ole täysimääräisesti mahdollista. Ei-semanttinen ratkaisumalli tulisi jo valmistuessaan olemaan vanhanaikainen. Ehdotukseni onkin, että Kansallista digitaalista kirjastoa alettaisiin kehittää arvioimalla tässä artikkelissa esiteltävä Kulttuurisampo-visio, ratkaisumalli, kokemukset ja prototyyppi ja edetä hankkeessa rohkeasti semanttisen webin keinoin.

Maassamme on vireillä hieman vastaavantyyppinen kansallinen hanke terveyteen liittyvien sisältöjen julkaisualustasta semanttisessa webissä: sosiaali- ja terveysministeriön rahoittama ja Kansanterveyslaitoksen koordinoima TerveSuomi.fi<sup>41</sup> (Hyvönen et al., 2007; Suominen et al., 2007; Suominen 2007). Sen prototyyppi<sup>42</sup> kehitettiin ja valmistui Kulttuurisammon tapaan Semanttisen laskennan tutkimusryhmän toimesta FinnONTO-hankkeessa ja järjestelmä otettiin koekäyttöön julkisessa verkossa 12.9.2008. Kansanterveyslaitos on nyt prototyypin perusteella toteuttamassa prototyypistä tuotantoversiota, sekä omin voimin että kahden yrityksen avulla. Prototyypin perusteella kehitävä ”tuotantoversio” julkistetaan helmikuussa 2009 Kansanterveyslaitoksen tarjoamana kansallisena palveluna.

Vastaava strategiaa kannattaa harkita myös kulttuurialalla. Se mahdollistaisi hankkeessa tarvittavan, nopeasti edistyvän web-teknologian kehittämisen yhteistyössä yliopistojemme kanssa, joiden intressissä ovat uudet teknologiset ratkaisut ja hyvien tietoaisteiden hyödyntäminen tutkimustyössä. Samalla kehitystyön intressipohjaa ja rahoista voidaan laajentaa yritysmaailmaan ja Tekesin kaltaisiin merkittäviin kansallisiin rahoittajiin vastaavalla tavalla kuin FinnONTO:n Kulttuurisampo-hankkeessa. Näin vapautuisi enemmän resursseja mm. sisällöntuotannon rahoittamiseen. Semanttisessa webissä aivan erityisesti pätee sanonta ”sisältö on kuningas”, content is king.

Kulttuurisampo on semanttisessa rikkaudessaan ja yhteisöllisyydessään ainutlaatuinen järjestelmä maailmassa. Kulttuurisammon takojien puolesta toivon, että FinnONTO-hankkeessa kehitetty kansallisen sisällöntuotannon, infrastruktuurin ja portaaliratkaisun Kulttuurisampo-malli ja sen prototyyppi johtavat sellaisen kansallisen tahtotilan syntymiseen, jonka tavoitteena on semanttiseen webiin ja Web 2.0 -konseptiin perustuva, yhteisöllinen, eri kulttuurisektorit ja asiakkaat yhdistävä kansallinen kulttuurisisältöjen julkaisujärjestelmä.

<sup>41</sup> <http://www.terveysuomi.fi/>

<sup>42</sup> <http://demo.seco.tkk.fi/terveysuomi/app>

## Kiitokset

---

Yli 30 tutkijaa on osallistunut artikkelissa esiteltyyn tutkimustyöhön sen eri vaiheissa. Kulttuurisampo-mallin kehitys- ja toteutustyö on tehty Semanttisen laskennan tutkimusryhmässä SeCo, joka on Teknillisen korkeakoulun mediatekniikan laitoksen ja Helsingin yliopiston tietojenkäsittelytieteen laitoksen yhteinen tutkimusryhmä. FinnONTO-hanketta ja Kulttuurisammon taontaa rahoittavat nykyisin 2008-2010 Tekes sekä 38 muun yrityksen ja julkisen organisaation muodostama konsortio, johon kuuluvat aakkosjärjestyksessä lueteltuina: AAC Global Oy, Antikvaria-ryhmä, CelAmanzi Oy, Connexor Oy, Coronaria Oy, eduskunta, Ego Beta Oy, Elisa Oyj, Espoon kaupunginmuseumo, Geologian tutkimuskeskus GTK, Helsingin kaupunginkirjasto, Helsingin yliopiston Viikin tiedekirjasto, Kansallisarkisto, Kansalliskirjasto, Kansanterveyslaitos, Koulutuskeskus Dipoli, Maa- ja metsätalousministeriö, Maanmittauslaitos, M-Cult ry, Merenkululaitos, Museovirasto, Mysema Oy, Osuuspankkikeskus OSK, PKT-Säätiö, Profium Oy, Rautaruukki Oyj, Sanastokeskus TSK, Sininen Meteoriiitti Oy, Stakes, Suomen Asiakastieto Oy, Suomalaisen kirjallisuuden seura SKS, Suomen valokuvataiteen museo, Taideteollisen korkeakoulun kirjasto, Terveiden edistämisen keskus (Tekry), TietoEnator Oyj, Valtion taidemuseo, Valtiovarainministeriö / ValTI ja Wärtsilä Oyj.

Lisäksi hanketta tukee merkittäväällä tavalla Suomen Kulttuurirahasto 2008-2009, jota kautta yhteistyöverkkoon on kytkeytynyt edellä mainittujen muistiorganisaatioiden lisäksi historianantutkijoiden yhteisö Agricola –Suomen historiaverkko, Kansalliskirjaston Mikkelin mikrokuvaus- ja konservointilaitos, Lusto Suomen metsämuseo ja Yleisradio. Työtä tehdään Teknillisellä korkeakoululla osin EU:n rahoittamaan SmartMuseum –projektiin<sup>43</sup> liittyen, jossa mukana oleviin muistiorganisaatioihin kuuluvat Firenzen Institute and Museum of History of Science (tuleva Galileo Galilei –museumo) ja Malta Heritage. Yhteistyötä on tehty tutkijavaihdoin Amsterdamin VU Universityn kanssa, joka toteuttaa suunnitelmien mukaan semanttisen version eurooppalaisesta digitaalisesta kirjastosta, Europeana-portaalista.

Kiitokset kaikille hedelmällisestä yhteistyöstä.

---

<sup>43</sup> <http://www.smartmuseum.eu/>

## Kirjallisuutta

---

J. Aitchison, A. Gilchrist, D. Bawden (2000): Thesaurus construction and use: a practical manual. Europa Publications, London.

O. Alm (2007): Tekstidokumenttien automaattinen ontologiaperustainen annotointi. Pro Gradu, Helsingin yliopisto. <http://www.seco.tkk.fi/publications/2007/alm-gradu-2007.pdf>

R. Baeza-Yates (1999): Modern information retrieval. Addison-Wesley, New-York.

Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, Matthew (eds.) (2006): Definition of the CIDOC Conceptual Reference Model. Version 4.2.1, October 2006. [http://cidoc.ics.forth.gr/docs/cidoc\\_crm\\_version\\_4.2.1.pdf](http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.1.pdf)

Jérôme Euzenat, Pavel Shvaiko (2007): Ontology matching. Springer-Verlag.

C. Fellbaum (1998): WordNet. The MIT Press, Massachusetts.

D. Fensel (2003): Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin. 2nd Edition, Springer-Verlag.

Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider (2002): Sweetening ontologies with DOLCE. In: Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Springer-Verlag.

Mikko Haaramo (2006): Iconclass-luokittelujärjestelmän ontologisointi ja soveltaminen. Diplomityö, Teknillinen korkeakoulu. <http://www.seco.tkk.fi/publications/>

Riikka Henriksson, Tomi Kauppinen, Eero Hyvönen (2008): Core Geographical Concepts: Case Finnish Geo-Ontology. Location and the Web (LocWeb) 2008 workshop, 17th International World Wide Web Conference WWW 2008, ACM International Conference Proceeding Series, Vol. 300.

Eero Hyvönen, Mirva Salminen, Suvi Kettula, Miikka Junnila (2004): A Content Creation Process for the Semantic Web. Proceedings of OntoLex 2004, Lisbon, Portugal. <http://www.seco.tkk.fi/publications/>

E. Hyvönen (2005): Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita. Signum, nro 5. <http://www.seco.tkk.fi/publications/2005/hyvonen-miksi-asiasanastot-eivat-riita-2005.pdf>

Eero Hyvönen (2008a): FinnONTO-malli kansallisen semanttisen webin sisältöinfrastruktuurin perustaksi - visio ja sen toteutus. Teknillinen korkeakoulu ja Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>

Eero Hyvönen (2008b): Semantic Portals for Cultural Heritage. In: (Staab, Studer, 2008), forth-coming.



Eero Hyvönen (2008c): Semantic Interoperability of Content and Services. A National Cross-domain Content Infrastructure for the Semantic Web. In: Phillip Sheu, Heather Yu, C. V. Ramamoorthy, Arvind K. Joshi, and Lotfi A. Zadeh (Eds.). *Semantic Computing*. IEEE Press / Wiley, New York, forth-coming.

E. Hyvönen, A. Valo, V. Komulainen, K. Seppälä, T. Kauppinen, T. Ruotsalo, M. Salminen, and A. Ylisalmi (2005a): Finnish National Ontologies for the Semantic Web - Towards a Content and Service Infrastructure. *Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005)*, Madrid, Spain, short papers.

E. Hyvönen, E. Mäkelä, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula (2005b): MuseumFinland - Finnish Museums on the Semantic Web. *Journal of Web Semantics*, Vol. 3, No. 2.

Eero Hyvönen, Eetu Mäkelä (2006): Semantic Autocompletion. *Proceedings of the first Asia Semantic Web Conference (ASWC 2006)*, Springer-Verlag.

Eero Hyvönen, Kim Viljanen and Osma Suominen (2007a): HealthFinland - Finnish Health Information on the Semantic Web. *Proceedings of the 6th International Semantic Web Conference (ISWC 2007)*, Springer-Verlag.

Eero Hyvönen, Tuukka Ruotsalo, Thomas Häggström, Mirva Salminen, Miikka Junnila, Mikko Virkkilä, Mikko Haaramo, Eetu Mäkelä, Tomi Kauppinen and Kim Viljanen (2007b): CultureSampo - Finnish Culture on the Semantic Web: The Vision and First Results. In: K. Robering (ed.): *Information Technology for the Virtual Museum*. LIT Verlag, Berlin.

Eero Hyvönen, Joeli Takala, Olli Alm, Tuukka Ruotsalo and Eetu Mäkelä (2007c): Semantic Kalevala - Accessing Cultural Contents Through Semantically Annotated Stories. *Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, Nov, 2007.

Eero Hyvönen, Katri Seppälä, Kim Viljanen ja Matias Frosterus (2007d): Yleinen suomalainen ontologia YSO - kohti suomalaista semanttista webiä. *Tietolinja*, no 1, 2007.

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

Thomas Häggström (2007): Toimintakeskeisen semanttisen moninäkömäähaun toteutus ja evaluointi kulttuurialan portaalisovelluksessa. *Diplomityö*, Teknillinen korkeakoulu.

Miikka Junnila (2006): Tietosisältöjen semanttinen yhdistäminen toimintakuvausten avulla. *Pro Gradu -työ*, Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>

Miikka Junnila, Eero Hyvönen, Mirva Salminen (2007): Describing and Linking Cultural Semantic Content by Using Situations and Actions. Teoksessa Klaus Robering (Ed.): Information Technology for the Virtual Museum. LIT Verlag, Berlin.

Ulla Karttunen (2008): Valokuvakirjallisuuden asiasanasto. Suomen valokuvataiteen museo. [http://www.fmp.fi/fmp\\_fi/muvieras/kirjasto/asiasan/index.htm](http://www.fmp.fi/fmp_fi/muvieras/kirjasto/asiasan/index.htm)

Tomi Kauppinen, Riikka Henriksson, Jari Väätäinen, Christine Deichstetter and Eero Hyvönen: Ontology-based Modeling and Visualization of Cultural Spatio-temporal Knowledge. Developments in Artificial Intelligence and the Semantic Web - Proceedings of the 12th Finnish AI Conference STeP 2006, October 26-27, 2006.

Tomi Kauppinen, Christine Deichstetter, Eero Hyvönen (2007a): Temp-O-Map: Ontology-based Search and Visualization of Spatio-Temporal Maps. Demo track at the European Semantic Web Conference ESWC 2007, Innsbruck, Austria, Springer, June 4-5, 2007.

Tomi Kauppinen, Eero Hyvönen (2007b): Modeling and Reasoning about Changes in Ontology Time Series. Ontologies. In Rajiv Kishore, Ram Ramesh, Raj Sharman (eds.): A Handbook of Principles, Concepts and Applications in Information Systems, Springer-Verlag.

Tomi Kauppinen, Jari Väätäinen, Eero Hyvönen (2008): Creating and Using Geospatial Ontology Time Series in a Semantic Cultural Heritage Portal. S. Bechhofer et al.(Eds.): Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, Springer-Verlag.

Suvi Kettula (2006): Käsiyöprosessit museossa semanttisen tiedonhaun lähteenä ja kohteen. Teoksessa L. Kaukinen, Collanus M. (toim.): Tekstejä ja kangastuksia. Puheenvuoroja käsityöstä ja sen tulevaisuudesta. Artefakta 17, Akatiimi, Tampere.

Suvi Kettula (2008a): Kansakunnan muisti tarvitsee termistöjä ontologioiden muodossa. Terminfo 2/2008.

Suvi Kettula (2008b): Valmisteilla oleva jatko-opintotyö MuseoSuomen termistöön ja ontologioihin liittyen. Helsingin yliopisto, kotitalous- ja käsityötieteiden laitos.

Ville Komulainen: Public Services for Ontology Library Systems. Pro Gradu -työ, Helsingin yliopisto, 2007. <http://www.seco.tkk.fi/publications/>

Tellervo Krogerus, Mirja Kälviäinen, Riitta-Liisa Leskinen, Anne Mikkanen (toim.) (1997) Käsi- ja taideteollisuuden asiasanasto. Taitemia 6, Kuopio. Verkkooversio: <http://www.designkuopio.fi/kirjasto/asiasanasto/>

Jussi Kurki (2008): Tekeillä oleva Pro Gradu -työ ONKI People -järjestelmään liittyen. Helsingin yliopisto.

Eetu Mäkelä (2006): View-Based Search Interfaces for the Semantic Web. Pro Gradu -työ, Helsingin yliopisto.

Eetu Mäkelä, Eero Hyvönen, Samppa Saarela (2006): Ontogator - A Semantic View-based Search Engine Service for Web Applications. Proceedings of the 5th International Semantic Web Conference (ISWC 2006). Springer-Verlag.

Eetu Mäkelä, Tuukka Ruotsalo, Eero Hyvönen (2007a): Domain-Centric View-Based Search. Poster proceedings of the 6th International Semantic Web Conference, Busan, Korea, November 11-15, 2007.

Eetu Mäkelä, Kim Viljanen, Olli Alm, Jouni Tuominen, Onni Valkeapää, Tomi Kauppinen, Jussi Kurki, Reetta Sinkkilä, Teppo Käsälä, Robin Lindroos, Osma Suominen, Tuukka Ruotsalo and Eero Hyvönen (2007b): Enabling the Semantic Web with Ready-to-Use Web Widgets. Proceedings of the First Industrial Results of Semantic Technologies Workshop, ISWC2007, November 11, 2007.

Eetu Mäkelä, Osma Suominen and Eero Hyvönen (2007c): Automatic Exhibition Generation Based on Semantic Cultural Content. Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea, November 12, 2007.

Eetu Mäkelä, Tuukka Ruotsalo and Eero Hyvönen (2007d): Domain-Centric View-Based Search. Poster proceedings of the 6th International Semantic Web Conference, Busan, Korea, November 11-15, 2007.

R. L. Leskinen (toim.) (1997): Museoalan asiasanasto. Museovirasto, Helsinki.

Robin Lindroos (2008): Paikkatiedon ontologiapalvelu. Diplomityö, Teknillinen korkeakoulu. <http://www.seco.tkk.fi/publications/>

Tuukka Ruotsalo and Eero Hyvönen (2007a): A Method for Determining Ontology-Based Semantic Relevance. Proceedings of the International Conference on Database and Expert Systems Applications DEXA 2007, Regensburg, Germany, Springer, September 3-7, 2007.

Tuukka Ruotsalo, Eero Hyvönen (2007b): An Event-based Approach for Semantic Metadata Interoperability. Proceedings of the 6th International Semantic Web Conference (ISWC 2007), Springer-Verlag.

Samppa Saarela (2004): Näkymäpohjainen RDF-haku. Pro Gradu -työ. Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>

Mirva Salminen (2006): Kuvien ja videoiden semanttinen sisällönkuvailu. Pro Gradu -työ, Helsingin yliopisto.

G. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, B. Omelayenko, J. van Ossenbruggen, R.

- Siebes, J. Taekema, J. Wielemaker, and B. Wielinga. MultimediaN E-Culture demonstrator. Proceedings of the Fifth International Semantic Web Conference (ISWC 2006), Springer-Verlag, 2006.
- Reetta Sinkkilä (2008): Käsitteen kontekstiperustainen valinta semanttisessa webissä. Pro Gradu -työ, Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>
- Reetta Sinkkilä, Eetu Mäkelä, Tomi Kauppinen and Eero Hyvönen (2008): Combining Context Navigation with Semantic Autocompletion to Solve Problems in Concept Selection. First International Workshop on Semantic Metadata Management and Applications, SeMMA 2008 (at (ESWC 2008). CEUR Workshop Proceedings, vol. 346, pp. 61-68, <http://CEUR-WS.org>.
- S. Staab, R. Studer (toim.) (2003, 2008, 2<sup>nd</sup> edition): Handbook on Ontologies. Springer-Verlag.
- J. Sowa (2000): Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, California.
- Osma Suominen, Kim Viljanen, Eero Hyvönen (2007): User-centric Faceted Search for Semantic Portals. Proceedings of the European Semantic Web Conference (ESWC 2007), Springer-Verlag.
- Osma Suominen (2008): Käyttäjäkeskeinen moninäkömähaku semanttisessa portaalissa. Pro Gradu -työ, Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>
- H. Suonuuti (2001): Guide to Terminology. NordTerm Publication 8, Tekniikan Sanastokeskus TSK, Helsinki.
- Taideteollinen korkeakoulu (1995): Taideteollisuuden asiasanasto. Taideteollisen korkeakoulun kirjasto, Helsinki.
- J. Tuominen. (2008): Helppokytkenäiset ontologiapalvelut semanttisessa webissä. Pro Gradu -työ, Helsingin yliopisto, ilmestyvä.
- Mikael Vakkari, Jyrki Simovaara ja Sirkka Valanto (2004); Kamut 2 – Muistiorganisaatioiden tietovarannot yhteiskäyttöön. Toteuttamismahdollisuudet ja toimenpidesuositus. KAMUT 2 - yhteishankkeen loppuraportti. 26.1.2004, Museovirasto. <http://www.nba.fi/tiedostot/8c050dc5.pdf>
- O. Valkeapää (2006): Verkkoresurssien ontologiaperustainen annotointi. Diplomityö, Teknillinen korkeakoulu, <http://www.seco.tkk.fi/publications/>.
- O. Valkeapää, O. Alm and E. Hyvönen (2007): A Framework for Ontology-based Adaptable Content Creation on the Semantic Web. Journal of Universal Computer Science, Vol. 13, No. 12.

Antti Vehviläinen (2006): Ontologiapohjainen kysymys-vastauspalvelu (Ontology-based question-answer service). Diplomityö, Teknillinen korkeakoulu. <http://www.seco.tkk.fi/publications/>

Antti Vehviläinen, Eero Hyvönen, Olli Alm (2008): A Semi-automatic Semantic Annotation and Authoring Tool for a Library Help Desk Service. *Emerging Technologies for Semantic Work Environments: Techniques, Methods, and Applications*, IGI Group, Hershey, USA.

Kim Viljanen (2006): Monilähteinen suositteleva semanttisessa webissä. Pro Gradu –työ, Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>

Kim Viljanen, Teppo Käsälä, Eero Hyvönen and Eetu Mäkelä (2006): ONTODELLA - A Projection and Linking Service for Semantic Web Applications. *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006)*, Krakow, Poland, pp. 370-376, IEEE Press.

Kim Viljanen, Jouni Tuominen, Teppo Käsälä, Eero Hyvönen (2008a): Distributed Semantic Content Creation and Publication for Cultural Heritage Legacy Systems. *Proceedings of the 2008 IEEE International Conference on Distributed Human-Machine Systems*, IEEE Press.

Kim Viljanen, Jouni Tuominen, Eero Hyvönen (2008b): Publishing and Using Ontologies as Mash-Up Services. *Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW2008), 5th European Semantic Web Conference 2008 (ESWC 2008)*, Tenerife, Spain.

Jari Väätäinen: Ajallisesti muuttuvan paikkatiedon hallinta. *Mediatekniikka, EVTEK, 2007*. <http://www.seco.tkk.fi/publications/>

Yiwen Wang, Natalia Stash, Lora Aroyo, Peter Gorgels, Lloyd Rutledge, Guus Schreiber (2008): Recommendations Based on Semantically-enriched Museum Collection. *Journal of Web Semantics* (in press).