

# MuseoSuomi

## Suomen museot semanttisessa webissä

*Järjestelmä museovieraan ja museon  
näkökulmasta.*

Eero Hyvönen  
1.3.2004



# Tiivistelmä

*Artikkelissa esitellään MuseoSuomi – Suomen museot online semanttisessa webissä -järjestelmä webin käyttäjän näkökulmasta. MuseoSuomi on demonstraatio julkaisukanavasta, jonka kautta maamme eri museot voivat julkaista yhdessä kokoelmiaan webissä sisällöllisesti yhteismitallisella tavalla. Ideana on muodostaa kansalaisille webiin virtuaalisesti yksi yhtenäinen kansallinen museokokoelma, MuseoSuomi, joka sisältää eri museoista erityyppisiä aineistoja, kuten esinekokoelmia, taidetta ja maastossa olevia kulttuurikohteita. MuseoSuomi tarjoaa käyttäjälleen älykkään web-palvelun, jonka kautta voi 1) hakea tietoja samanaikaisesti kaikista järjestelmään kuuluvista kokoelmista ja 2) samoilla kokoelmissa sisältöjä yhdistäviä temaattisia suositteluja seuraten. Perinteisistä hakukoneista poiketen haku perustuu hakusanojen taustalla oleviin käsitteisiin ja taksonomioihin (ontologioihin), mikä helpottaa merkittävästi hakukyselyiden muodostamista ja johtaa tarkempiin hakutuloksiin. Myös suosittelulinkit perustuvat järjestelmän taustalla olevien kokoelmatietojen muodostamaan assosiativiseen semanttiseen verkostoon.*

*MuseoSuomen ensimmäinen pilottiversio otetaan koekäyttöön 8.3.2004. Siihen sisältyy n. 3000 kokoelmakohteen tiedot Kansallismuseon, Espoon kaupunginmuseon ja Lahden kaupunginmuseon kulttuurihistoriallisista esinekokoelmista. Taustalla olevia ontologisia yleis- ja yksilökäsitteitä on n. 10000.*

## **Tutkimusryhmä, -konsortio ja johtoryhmä**

MuseoSuomi-konsepti ja -järjestelmä on ideoitu ja kehitetty Helsingin yliopiston tietojenkäsittelytieteen laitoksen ja Tietotekniikan tutkimuskeskus HIIT:n Semantic Computing -tutkimusryhmässä. Sovelluksen lähtökohtana oli mm. Museoliiton 1990-luvulla visioima ajatus siitä, että Suomen eri museoiden kokoelmatietoja voitaisiin julkaista webin välityksellä kansallisen portaalin kautta samaan tapaan kuin esimerkiksi Australiassa<sup>1</sup> ja Kanadassa<sup>2</sup>. MuseoSuomi-hanke on osa Tekesin, Nokian, TietoEnatorin ja Espoon kaupunginmuseon, Museoviraston, Helsingin yliopistomuseon säätiön ja Antikvaria-ryhmän rahoittamaan Semantic Web -teknologian tutkimus- ja kehityshanketta 2002–2003. Projektin tutkimus- ja kehitysryhmään ovat kuuluneet prof. Eero Hyvösen johdolla (aakkosjärjestyksessä) Miikka Junnila, Suvi Kettula, Eetu Mäkelä, Vilho Raatikka, Samppa Saarela, Mirva Salminen, Ahti Syreeni, Arttu Valo ja Kim Viljanen. Projektin johtoryhmässä ovat toimineet Sten Björkman, Merja Heiskanen, Eero Hyvönen, Olli Immonen, Seppo Mustonen, Juha Mäkinen, Jaana Tegelberg, Kati Heinämies, Vesa Hongisto, Anneli Rantanen, Liisa Ropponen, Petteri Saarinen, Sirkka Valanto ja Kim Viljanen.

---

<sup>1</sup> Australian Museums Online, <http://www.amonline.net.au/>

<sup>2</sup> Artefacts Canada, <http://www.chin.gc.ca/>

## 1 WWW ja museot: uudet mahdollisuudet ja haasteet

---

### 1.1 Uusi julkaisukanava

Web tarjoaa uudenlaisen kanavan museokokoelmiin tutustumista ja tiedonhankintaa varten. Museossa käynnin fyysistä kokemusta näyttöpäätte ei korvaa, mutta mahdollistaa tätä täydentäviä tapoja kokoelmiin tutustumiselle:

- **Ajasta ja paikasta riippumaton tiedonvälitys.** Tietosisällöt voidaan aiempaa joustavammin saada tutkijoiden ja suuren yleisön käyttöön. Kokoelmiin tutustuminen ei edellytä museorakennukseen tulemistä tiettyyn aukioloaikaan.
- **Tietosisällön määrän lisääminen.** Laajat kokoelmat, joiden fyysinen näytteille asettaminen olisi mahdotonta, voidaan avata yleisölle WWW:n välityksellä.
- **Tietosisällön laadun parantaminen.** Tietotekniikka mahdollistaa tietojen haun, yhdistämisen ja esittämisen tavoilla, jotka eivät ole teknisesti tai taloudellisesti mahdollisia fyysisissä näyttelytiloissa.
- **Vuorovaikutteisuuden hyödyntäminen.** Tietojärjestelmien vuorovaikutteisuus (interactivity) ja sopeutuvuus (adaptivity, learning) mahdollistavat aiempaa joustavampia ja henkilökohtaisempia tapoja tutustua aineistoihin.
- **Audiovisualisuuden hyödyntäminen.** Tietotekniikka mahdollistaa sisältöjen esittelemisen audiovisuaalisilla keinoilla, kuten kuva-, musiikki- ja videotallenteilla.
- **Uudet kokoelmakohteet.** Tietotekniikka mahdollistaa uudentyyppisten aineistojen tallentamista museoihin osana kulttuuriperimäämme. Esimerkiksi Internet-sivustoja on määrätietoisesti ryhdytty tallentamaan historiallisina dokumentteina (Kansalliskirjaston toimesta).

Web-sovellusten kehittämisessä museoyleisölle on runsaasti käytännöllisiä teknisiä haasteita. Suurimmat haasteet liittyvät kuitenkin tietosisältöjen esittämiseen eikä tekniikkaan. Avainkysymys on, miten web-sivustoista saadaan kävijöiden kannalta aidosti mielenkiintoisia, hyödyllisiä ja helpokäyttöisiä. Tähän ei pelkkä perinteinen kokoelmatietojen hakumahdollisuus tietokannasta riitä. Siinä ei hyödynnetä kuin pieni osa niistä mahdollisuuksista, joita kokoelmien rikas sisältö voi tarjota.

### 1.2 Museokokemus webissä

MuseoSuomen kehitystyön lähtökohtana on, 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. MuseoSuomen visiona on, että esimerkiksi Akseli Gallen-Kallelan taulua ”Kullervon sotaan lähtö” (kuva 1.1)



**Kuva 1.1.** Akseli Gallen-Kallelan teos Kullervon sotaan lähtö.

esittelevältä web-sivulta voitaisiin ”sotaa” teemana käyttäen luoda linkkejä eri kokoelmien ja (muisti) organisaatioiden aineistoihin, kuten

- eri taidemuseoiden muihin sota-aiheisiin tauluihin,
- kulttuurihistoriallisten museoiden kokoelmissa oleviin aseisiin ja varusteisiin, jotka ovat vastaavaa tyyppiä kuin Kullervolla taulussa,
- kirjastojen arkistoissa oleviin tutkimuksiin Kalevalasta ja Kullervo-hahmosta
- luonnonhistoriallisessa museossa oleviin aineistoihin susista ja hevosista tai
- webissä oleviin ulkoisiin, sotaa tai Kullervoä käsitteleviin sivustoihin.

Temaattisia virtuaalinäyttelyitä on tehty käsityönä sekä CD-levitykseen että webiin. Tämä onnistuu nykyisin pienille aineistoille kohtuullisilla kustannuksilla. MuseoSuomessa on kuitenkin kysymys kokoelmatietokantoihin perustuvista virtuaalinäyttelyistä, jossa voi olla kymmeniä tuhansia erityyppisiä kohteita eri museoiden kokoelmatietokannoista. Lisäksi kokoelmat kasvavat vuosi vuodelta. Käsintehdävä kiinteä teemoitus tai linkitys ei aineistojen monimuotoisuuden, laajuuden 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.

### 1.3 Webin käyttö tiedonhaussa

Webillä on kaksi pääasiallista käyttötapaa tiedonhaussa: 1) haku (search) Googlen ja Yahoon kaltaisilla hakukoneilla ja 2) samoilu (browse) sivuja yhdistäviä linkkejä seuraten. MuseoSuomen käyttöön liittyy vastaavasti kaksi keskeistä ongelma-aluetta:

Miten tiedon haku suurista kokoelmista tehdään riittävän helpoksi, kun käyttäjällä ei välttämättä ole hyvää käsitystä kokoelmien sisällöstä tai edes täsmällistä tiedonhaun tavoitetta? Jo järkevien hakusanojen keksiminen Googlen kaltaiselle hakukoneelle voi olla vaikeaa. Hakukoneen avulla olisi voitava löytää kokoelmista paitsi yksittäisiä esineitä, myös hahmottaa näistä muodostuvia kokonaisuuksia.

Miten kokoelmien selailu voidaan toteuttaa vastaavaan helppoon ja tuttuun tapaan kuin webissä? Webin sivuja yhdistävät linkit on (yleensä) luotu ja nimetty ihmisen toimesta ja saatu näin lukijoille mielenkiintoisiksi. MuseoSuomen kaltaisen järjestelmän haasteena on käyttäjälle mielenkiintoisen linkityksen automaattinen tuottaminen tekoälytekniikoilla.

## 2 MuseoSuomen tavoitteet

---

MuseoSuomi-konseptin keskeiset ideat järjestelmän loppukäyttäjän kannalta voidaan kiteyttää seuraavaan neljään päätavoitteeseen.

### 2.1.1 Suomen museoiden yhteiskokoelma webiin

Maamme museokokoelmat ovat nykyisin hajallaan satojen eri museoiden tietokannoissa. Käytössä on kymmeniä erilaisia museotietojärjestelmiä ja tiedonesitystapoja<sup>3</sup>. Useimpien museoiden kokoelmiin ei nykyisin ole mahdollista tutustua webin välityksellä.

Kokoelmatietojen hajautuneisuus, monimuotoisuus ja keskinäinen yhteensopimattomuus muodostavat kansallisella tasolla vakavan esteen tiedon saannille ja kokonaiskuvan luomiselle maamme kokoelmista. Loppukäyttäjälle mahdollisuus kokoelmiin tutustumiseen yhtenä loogisena kokonaisuutena olisi ensiarvoisen tärkeää. MuseoSuomi-hankkeen tavoitteena on muodostaa kansalaisten käytettäväksi webiin virtuaalisesti yksi yhtenäinen kansallinen museo-kokoelma. Projektin puitteissa on suunniteltu ja toteutettu tällaisen järjestelmän ensimmäinen prototyyppi.

MuseoSuomen ensimmäisessä vaiheessa yhdistetään heterogeenisissä tietojärjestelmissä olevia samantyyppisiä aineistoja. Pilottikohteena ovat kulttuurihistorialliset esinekokoelmat Kansallismuseosta, Espoon kaupunginmuseosta ja Lahden kaupunginmuseosta. Tämän jälkeen tavoitteena on ottaa mukaan erityyppisten museoiden ja sisällöntuottajien aineistoja, kuten

- taidemuseoiden taidekokoelmia,
- valokuvakokoelmia,
- tietoja ympäristön kulttuurikohteista, kuten muinaismuistoista, linnoista yms. sekä
- kirjastojen ja arkistolaitoksen dokumentteja.

Kunnianhimoisena tavoitteena on kaataa keinotekoiset raja-aidat sekä eri paikoissa ja erilaisissa tietojärjestelmissä sijaitsevien kokoelmien että sisällöltään erityyppisten kokoelmien välillä.

### 2.1.2 Sisältöperustainen semanttinen tiedonhaku

MuseoSuomessa tiedon haku ei perustu vain hakusanoihin vaan ennen kaikkea niihin liittyviin käsitteisiin. Esimerkiksi yhdellä hakusanalla ”päähine” voidaan löytää erityyppiset päähineet, kuten ”silinterit”, ”ylioppilas-

---

<sup>3</sup> Vesa Hongisto: Museoiden kokoelmien digitointiasteen kartoitus. Museovirasto, Helsinki, 2002.

lakit” tai ”harakat” ilman että käyttäjän tarvitsee pystyä luettelemaan näitä sanoja erikseen. Samoin ilmaisulla ”Etelä-Suomen lääni” voidaan viitata automaattisesti kaikkiin kyseisen alueen kuntiin kuten luonnollisessakin kielessä. Tällainen sanojen taustalla oleviin käsitteisiin perustuva tiedonhaku on lisäksi mahdollista tehdä samanaikaisesti kohteiden eri ominaisuuksien suhteen. Esimerkiksi valitsemalla kiinnostuksen kohteeksi ”päähineet”, kohteen käyttöpaikaksi ”Etelä-Suomen lääni” ja käyttäjäksi ”nainen” löytyvät kaikki ko. läänin eri kunnissa ja kylissä naisten käyttämät erityyppiset päähineet.

### 2.1.3 Semanttinen suosittelu ja teemoitus linkeillä

MuseoSuomessa tiedon haku ei kohdistu vain kokoelmaesineisiin vaan tavoitteena on samalla paljastaa käyttäjälle, miten eri kokoelma-kohteet liittyvät toisiinsa. Yhteydet näkyvät käyttäjälle nimettyinä linkeinä, joita painamalla voi siirtyä kohteesta toiseen kuten webissä on tapana. Näin voidaan luoda semanttisesti linkitettyjä temaattisia kokonaisuuksia. Esinettä kuvaavaan web-sivuun voidaan esimerkiksi automaattisesti tuottaa linkit toisiin esineisiin, joita sama henkilö on käyttänyt tai lisätä linkit taideteoksiin, joita hän on luonut. Taulu voidaan linkittää toisiin teoksiin, joissa esiintyy samoja aiheita, jotka on luonut joku toinen saman koulukunnan edustaja, tai dokumentteihin, joiden aiheena on taulun maalarin tuotanto. MuseoSuomen tekniikalla tällaiset assosiaatiot voidaan määritellä joustavasti logiikan säännöillä. Näitä on helppoa muuttaa tai täydentää, jolloin koko näyttelyn linkitys muuttuu vastaavasti ilman että yksittäisten kokoelma-kohteiden tietoihin tarvitsee lainkaan kajota.

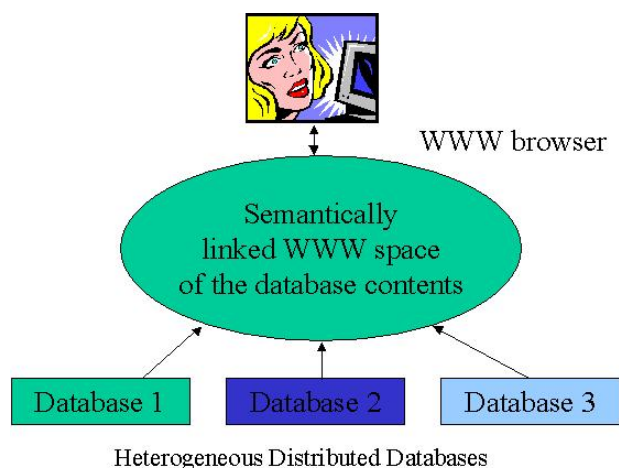
### 2.1.4 Kansallisen julkaisukanavan luominen

MuseoSuomen tavoitteena on viimekädessä kansallisen museoiden julkaisukanavan kehittäminen webiin. Sen kautta maamme eri museot voisivat julkaista yhdessä kokoelmiaan sisällöllisesti yhteismitallisella tavalla.

Webin menestyksen yksi syy on sisällön julkaisemisen helppous. Jokainen voi julkaista HTML- ja muita dokumentteja webissä helposti ja halvalla tallentamalla ne web-palvelimelle julkiseen hakemistoon. Julkaisuprosessin yksinkertaisuus on tavoitteena myös MuseoSuomen julkaisukanavassa. Järjestelmään liittyviltä museoita ei edellytetä teknisesti vaativaa kokoelmatietokantojen liittämistä webiin ja tällaisten palvelinjärjestelmien ylläpitoa. Riittää että museo tuottaa kokoelmistaan tietyn standardiformaatin ja sanastojen mukaista dataa. Toki tähänkin vaatimukseen liittyvät omat haasteensa: luetteloinnin on oltava systemaattista. Kun muunnos tietokannasta standardiformaattiin on ohjelmoitu ja sisällöntuotantoprosessi saadaan käyntiin, ei museossa jatkossa tarvita syvällisempää tietoteknistä osaamista. MuseoSuomi-hankkeessa on koemielessä ohjelmoitu muuntimet Musketti-, Antikvaria- ja Escoll-museotietojärjestelmille.

## 3 Kokoelmien looginen yhdistäminen

MuseoSuomi yhdistää eri museoiden kokoelmatiedot yhdeksi loogiseksi kokonaisuudeksi kuvan 3.1 mukaan. Käyttäjä näkee MuseoSuomen yhteisenä semanttisesti linkitettyinä web-sivustona, johon voidaan kohdistaa hakuja ja jota voidaan selailta linkejä seuraten.



**Kuva 3.1.** MuseoSuomen idea loppukäyttäjän näkökulmasta.

Ensimmäinen edellytys kokoelmatietojen mielekkäiden linkitysten ja teemaattisten kokonaisuuksien automaattiselle tuottamisella on kokoelmatietojen sisällöllinen yhteismitallinen esittäminen ja tietojärjestelmien yhteentoimivuus (interoperability). MuseoSuomi-hankkeen keskeisenä tutkimuskohteena onkin ollut kehittää menetelmiä, ohjelmistoja ja ontologioita, joiden avulla eri tietokantajärjestelmissä ja erilaisilla tiedonesitystavoilla tallennetut kokoelmatiedot voitaisiin yhdistää.

Tiedot on kyettävä yhdistämään sekä *rakenteen* että *sisällön* suhteen. Seuraavassa tarkastellaan lyhyesti kokoelmatietojen rakenteellista ja sisällöllistä yhdistämistä.

### 3.1 Rakenteellinen yhdistäminen (syntaksi)

Eri museoiden tietokannat ovat *hajautettuja*, eli sijaitsevat fyysisesti eri paikoissa ja *heterogeenisiä*, eli järjestelmät

- käyttävät erilaisia eri valmistajien tietokantajärjestelmiä ja
- niiden looginen rakenne (erilaiset taulukot, kentät, nimet jne.) voi olla erilainen.

Kokoelmatietokantojen yhdistäminen on siksi teknisesti varsin hankalaa.

Rakennetasolla kokoelmatietojen yhdistäminen tarkoittaa sitä, että tunnistetaan yhteiset keskeiset kokoelmien tietokentät, esimerkiksi ”esineyyppi” tai ”valmistaja”, ja sovitaan yhteisestä kenttien nimistä ja esitystavasta. Tarvitaan siis yhteisesti sovittu julkaistavien kokoelmatietojen esityskieli.

XML (eXtensible Markup Language) on tietotekniikassa voimakkaasti yleistynyt muodollinen tapa sopia tällaisista kielistä. XML on metakieli, kielten määrittelemisen kieli, jonka avulla eri yhteisöt kuten museoala voivat helposti ja täsmällisesti määritellä omia sovelluskohtaisia tiedonesityskieliä. WWW-tekniikat perustuvat nykyisin tyypillisesti erilaisiin XML-standardeihin. XML on koko webin kielellinen perusta ja sitä käytetään yhä enemmän myös museoiden tietojärjestelmissä.

XML:n käytöllä avulla saavutetaan mm. seuraavia etuja.

- **Avoimet standardit.** XML-kielet ovat talletusmuodoltaan yksinkertaista tekstiä. Työkaluja (myös ilmaisia public domain -ohjelmistoja) XML-aineistojen käsittelyyn on runsaasti saatavilla, eikä formaattia ole sidottu yksittäisten valmistajien tuotteisiin. Tällaisten ns. *avointen standardien* käyttö vähentää museoiden riippuvuutta yksittäisistä järjestelmätoimittajista ja tuotteista, lisää valintamahdollisuuksia ja laskee kilpailun lisääntyessä hintoja.
- **Kokoelmatietojen siirto.** Kokoelmatietoja tullaan säilyttämään ja jalostamaan pitkiä aikoja. Tämä on ristiriidassa tietotekniikan kiihkeään kehityksen kanssa, jossa yritykset ja järjestelmät syntyvät ja kuolevat nopealla tempolla. Museot joutuvat siksi yhä uudelleen muuntamaan ja siirtämään tietokantojaan uusiin järjestelmiin. Avointen standardien kautta tämä on yleensä sekä joustavaa että taloudellista.
- **Eri museoiden kokoelmien yhdistäminen.** Avoimet standardit tarjoavat niin ikään keinoja, joilla eri kokoelmien tietojen yhdistäminen helpottuu. Tilanteesta riippuen yhdistäminen voi merkitä museoiden kokoelmien kokoamista yhdeksi tietokannaksi tai erillisten kokoelmien *yhteiskäyttöä* verkon yli yhdessä sovittujen esitysmuotojen ja tiedonsiirtoikäntöjen avulla.
- **Julkaiseminen eri muodoissa.** XML:n käytön tärkeä etu on, että sen avulla esitetyistä kokoelmatiedoista voidaan helposti (ns. XSLT-muunnoksilla) tuottaa erilaisia ilmiasuja, esimerkiksi WWW-sivuja tai painettuja julkaisuja.

XML-kielten avulla museoiden on mahdollista sopia täsmällisesti yhteistä järjestelmäriippumattomista standardeista kokoelmatietojen tallentamisessa. Kun yhteisestä kielestä on sovittu, helpottuu kokoelmien siirto, yhdistäminen ja WWW-julkaiseminen merkittävästi.

MuseoSuomi-järjestelmässä museotietojen yhdistäminen perustuu rakennetasolla yhteisesti sovittavaan XML-kieleen, jolla ilmaistaan webissä julkaistavat kokoelmatiedot.

### 3.2 Sisällöllinen yhdistäminen (semantiikka)

XML on moderni ratkaisu tietojen yhdistämiseen rakenteen eli syntaksin tasolla. Syvällisempi tietojen sisällöllinen yhdistäminen ja yhteiskäyttö edellyttävät, että kokoelmatiedoissa käytetyt *sanastot* saadaan merkitykseltään eli semantiikaltaan yhteismitalliseksi. Esimerkiksi samanlainen päähine voi olla toisessa museossa olla luetteloituna ”liinalakki”, toisessa taas ”tanu”. Hakusanalla ”tanu” jäisivät ”liinalakkeina” luetteloidut esineet löytämättä.

Sanaston lisäksi on sovittava tavoista, joilla tietojen välisiä semanttisia suhteita merkitään. Sitä ennen tietysti on vielä sovittava, mitkä semanttiset



suhteet ylipäänsä ovat hyödyllisiä ja kiinnostavia aineiston tulevan käytön kannalta webissä. Yksi keskeinen mm. asiasanoituksessa käytetty semanttinen suhde on yläkäsite-alakäsite-suhde (hyponymia), joka ilmaisee vaikkapa sen tosiasian, että ”tuolit” ovat ”huonekaluja”. Jos tämä semanttinen suhde tiedetään, voidaan käsitteellä ”huonekalu” löytää kokoelmista tuolien ohella myös ”pöydät”, ”kaapit” ym. huonekalutyypit. Meronymia puolestaan on osan ja kokonaisuuden välinen semanttinen suhde. Esimerkiksi ”pöytälevy” on ”pöydän” osa ja ”Ahvenanmaa” kuuluu ”Suomeen”. Kokoelmatiedoista löytyy paljon semanttisia suhteita, kuten että ”valmistaja”-ominaisuus liittyy ”esineen” johonkin ”toimijaan”, ”valmistuspaikka”-ominaisuus ”paikkaan” jne.

Kokoelmatietojen sisällöllinen yhdistäminen edellyttää, että käytettävissä on kokoelmatiedoissa esiintyvien esinetyyppien ja muiden käsitteiden sekä näiden välisten suhteiden täsmällinen määrittely eli *ontologia*. Ontologioiden käyttö luetteloinnin ja tiedonhaun perustana on tulevaisuuden trendi museoalalla<sup>4</sup>.

Ontologia	Sisältö	Koko
<b>Museoalan ontologia MAO</b>	Museoalan käsitteistö. Kehitty MASA-tesauruksesta.	n. 6900 yleiskäsitettä taksonomisesti luokiteltuna.
<b>Esineet</b>	Konkreettisten kokoelmaobjektien taksonomia.	MAO:n osa, 3227 luokkaa.
<b>Materiaalit ja aineet</b>	Taksonomia materiaaleista, joista konkreettiset objektit on valmistettu.	MAO:n osa, 364 luokkaa.
<b>Tapahtumat</b>	Yhteiskunnan tapahtumien ja prosessien taksonomia.	MAO:n osa, 992 luokkaa.
<b>Toimijat</b>	Yksilöiden, yritysten ym. toimijoiden yksilöiden ja käsitteiden ontologia.	26 käsitettä ja 1715 yksilöä.
<b>Paikat</b>	Paikkakäsitteiden (kylät, kaupungit jne.) ja näiden yksilöiden ontologia.	33 käsitettä ja 864 yksilöä (hyponymia ja meronymia).
<b>Ajat</b>	Aikakausien ontologia aikaintervalleina.	57 käsitettä.
<b>Kokoelmat</b>	Museoiden ja kokoelmien ontologia.	22 käsitettä ja 24 yksilöä.

**Taulukko 3.1.** MuseoSuomen ontologiat (helmikuussa 2004).

Taulukko 3.1 esittää yhteenvedon MuseoSuomen tämänhetkisistä ontologioista. Keskeisin ontologia on Museoalan ontologiaksi MAO ristitty taksonomia. Se on kehitetty Museoviraston Museoalan asiasanasto MASA:n pohjalta ryhmittelemällä tämän käsitteet systemaattisesti taksonomisiksi hierarkioiksi ja täydentämällä sanastoa uusilla, kokoelmissa esiintyvillä ja

<sup>4</sup> Esimerkiksi ICOM:n piirissä kehitetään yleistä Conceptual Reference Model (CRM) –ontologiaa.

muilla käsitteillä (n. 900 uutta käsitettä). MuseoSuomen ontologiat ”Esi-  
neet”, ”Materiaalit ja aineet” ja ”Tapahtumat” perustuvat suoraan MAO:n  
vastaaviin alihierarkioihin. Toimijat-ontologian avulla voidaan tehdä yksi-  
käsitteiseksi esimerkiksi samannimiset eri henkilöt (esimerkiksi eri ”Matti  
Virtaset”). Paikat-ontologialla taas voidaan erottaa toisistaan vaikkapa  
maamme eri ”Pyhäjärvet”.

Ontologioiden esittämiseksi web-sovelluksissa käytetään XML:ää korke-  
ammantasoista RDF-standardia<sup>5</sup>. RDF on yksinkertainen tapa esittää meta-  
tietoa nimettyinä ominaisuuksia. Suositukseen liittyvän RDF Scheman  
avulla voidaan sopia RDF kuvauksissa käytettävästä sanastosta ja ominai-  
suuksista. RDF on webin teknologista kehitystä koordinoivan W3C-  
konsortion suositus<sup>6</sup>.

Käytännössä tietojen semanttinen yhdistäminen MuseoSuomessa tarkoittaa  
sitä, että rakennetasolla tunnistettujen yhteisten tietokenttien arvoina olevat  
kielelliset ilmaukset voidaan yhdistää liittämällä ne ontologian käsitteisiin.  
Esimerkiksi esinetyyppi-kentän merkkijonoarvot ”liinalakki” ja ”tanu”  
voidaan tunnistaa samaksi käsitteeksi, jos nämä termit esiintyvät ontologi-  
assa saman käsitteen synonyymisinä suomenkielisinä niminä. Näin eri mu-  
seoiden tai eri luetteloiden käyttämä erilainen terminologia voidaan saada  
yhteismitalliseksi ilman, että paikallisesti käytettyä sanastoa tarvitsee vält-  
tämättä muuttaa. Riittää, että ontologian käsitteisiin viittaamalla on määri-  
tely, mitä luetteloinnissa käytetyillä termeillä tarkoitetaan.

### 3.3 Sisällön rikastaminen

Luetteloinnin yhteydessä esineistä talletetaan metatietoa, kuten esineen  
tyyppi (nimi), valmistusmateriaali, käyttäjät jne. Asiasanojen ja luokitusten  
avulla esineeseen liitetään kontekstuaalista metatietoa. Esimerkiksi asiasa-  
na ”joulu” kertoo kohteen liittymisestä tavalla tai toisella jouluun – asiasa-  
na ei kerro tarkemmin assosiaation luonteesta. Kun yksittäisen kokoelma-  
objektin tietokentän arvo, kuten esineen tyyppi ”keinutuoli”, yhdistetään  
ontologian vastaavaan käsitteeseen, yhdistyy kyseinen objekti samalla au-  
tomaattisesti ontologian virittämään verkostoon, johon on aiemmin liitetty  
muitakin kokoelmatietoja. Näin syntyy suuri semanttinen verkko, *tietä-  
myskanta* (knowledge base), jossa kokoelmatiedot liittyvät toisiinsa onto-  
logiassa määriteltyjen käsitteiden ja suhteiden avulla. Syntyneiden ontolo-  
gisten yhteyksien kautta kokoelmatietojen sisältö rikastuu automaattisesti  
– kokonaisuus on enemmän kuin osiensa summa. MuseoSuomen älykäs si-  
sältöperustainen tiedonhaku perustuu tähän semanttisesti rikastuneeseen  
sisältöön.

Kielitieteessä semanttisen tason yläpuolella on vielä pragmaattinen taso,  
jolla kuvataan kielen merkitysten käyttöä ja tarkoituksia. Myös museoko-  
koelmilla on pragmaattinen ulottuvuutensa, joka kertoo esimerkiksi sen,  
miten ja millaisissa yhteyksissä kokoelmaesineitä on yhteiskunnassa käy-  
tetty. MuseoSuomen tietämyskanta tarjoaa mahdollisuuden kokoelmatieto-

<sup>5</sup> <http://www.w3.org/RDF/>

<sup>6</sup> RDF(S):n pohjalta ollaan standardoimassa vielä korkeammantasoista ontologiakieltä OWL, joka  
perustuu formaaliin terminologialogiikkaan. MuseoSuomi-järjestelmässä on rajoitettu RDF(S)-  
tason peruselementteihin metatietojen esittämisessä.

jen semanttiseen rikastamiseen paitsi ontologisten yhteyksien myös loogisten sääntöjen avulla. Näiden avulla voidaan päätellä kokoelmakohteille uusia metatietoja, kuten oletusarvoista pragmaattista kontekstietoa yleisellä tasolla. Voidaan esimerkiksi määritellä yleinen sääntö:

”ylioppilaslakia” käytetään ”ylioppilasjuhlat” –tyyppisissä ”tapahtumissa”

Tämän jälkeen järjestelmä ymmärtää, että juhla-tapahtumissa käytettäviä kohteita haettaessa pitää palauttaa hakutuloksena myös kaikki ylioppilaslakit, koska ”ylioppilasjuhlat” ovat eräs juhla-käsitteen alityyppi ja ylioppilaslakit ovat ”ylioppilasjuhliin” liittyviä esineitä. Päätelmän tekemiseksi ei luettelointitiedoissa tarvitse yksittäisen lakin käyttötarkoitusta erikseen olla kerrottu. Sääntöjen avulla voidaan siis helpottaa metatiedon tuottamistyötä. Jos kokoelma-aineistosta halutaan myöhemmin luoda uudenlaisia yhteyksiä eri sovelluksiin tai näyttelyihin, voidaan näitä tuottaa joustavasti samasta aineistosta vain sääntöjä muuttamalla.

### 3.4 Temaattinen yhdistäminen

Pragmatiikan tasolla tarjoutuu virtuaalinäyttelyn tekijälle tulevaisuudessa mielenkiintoisia uusia mahdollisuuksia tietojen temaattisen yhdistämisen kannalta. Tarkastellaan esimerkiksi ”verkolla kalastamista”. Kyseessä on tapahtuma, johon voi liittyä monia esineitä, ihmisiä, paikkoja, aika ym. tekijöitä eri rooleissa. Esimerkiksi ”kalaverkko” liittyy apuvälineen ominaisuudessa ”kalastamiseen”. Tiettyä kalaverkkoa voi olla käytetty tiettyyn aikaan tiettyssä paikassa kalastamiseen joidenkin henkilöiden toimesta, ehkä tiettyä museossa olevaa venettä apuna käyttäen. Pragmaattinen taso, tässä kalastaminen tapahtumana, tarjoaa näin mahdollisuuden monien tietokannassa olevien eri esineiden ja toimijoiden temaattiseen yhdistämiseen. Museossa kävijän kannalta pragmaattinen taso on erityisen kiinnostava, koska vasta käyttöön liittyvä kontekstietä tekee monet esineet mielenkiintoisiksi ja saa ne ”elämään”.

## 4 Semanttinen tiedonhaku MuseoSuomessa

---

MuseoSuomen käyttäjälle on tarjolla kolme tapaa hakea tietoa: *näkymähaku* (view-based search), *käsitehaku* ja *semanttinen samoilu*.

### 4.1 Näkymähaku

Näkymällä (view) tarkoitetaan kokoelma-aineiston hierarkkista luokittelua jonkin näkökulman, kuten esinetyypin, materiaalin, valmistuspaikan tms. suhteen. Kuvassa 4.1 on esitetty esimerkkinä osa MuseoSuomen esinetyyppi-näkymästä. Luokittelun kategoriat ovat hakukoneen käynnistäviä linkkejä ja kunkin (ali)kategorian perässä oleva luku kertoo niihin kuuluvien esineiden lukumäärän kokoelmissa. Esimerkiksi linkkiä ”taltat (8)” painamalla hakutulokseen tulee kahdeksan mahdollisesti eri kokoelmissa olevaa talttaa. Näkymähierarkian ideana on antaa käyttäjälle yleiskuva museoaineiston jakautumisesta eri luokkiin yhden näkymän suhteen ja mahdollisuus suorittaa haku minkä tahansa kategorian suhteen.



**Kuva 4.1.** Esineityyppien hierarkkinen näkymä. Näkyvässä on osa ”työvälineet” luokittelua linkkeinä.

Näkymän sisältö	Näkymä	Taustaontologia
Esineen tiedot	Esineityyppi	Esineet (MAO)
	Materiaali	Materiaalit (MAO)
Valmistustiedot	Valmistaja	Toimijat
	Valmistuspaikka	Paikat
	Valmistusaika	Ajat
Käyttötiedot	Käyttäjä	Toimijat
	Käyttöpaikka	Paikat
	Käyttötilanne	Tapahtumat (MAO)
Kokoelmatiedot	Kokoelma	Kokoelmat

**Taulukko 4.1.** MuseoSuomen näkymähierarkiat.

MuseoSuomessa on käytössä 9 eri näkymää jaoteltuna esinetietoihin, valmistustietoihin, käyttötietoihin ja kokoelmatietoihin taulukon 4.1 mukai-

sesti. Näkymät luodaan automaattisesti niiden taustalla olevien ontologioiden (taulukko 3.1) ja kokoelma-aineistojen perusteella.

MuseoSuomen näkymähaussa avautuu käyttäjälle kuvan 4.2 kaltainen sivu, jonka avulla voidaan tehdä näkymähakuja.



**Kuva 4.2.** MuseoSuomen moninäkymähaku

Vasemmassa sarakeessa ovat näkymät avattuna siten, että näkymän kategorian alikategoriat on lueteltu linkkeinä kunkin näkymän kohdalla. Ruutuun mahtuvat tässä näkymät Esinetyyppi, Materiaali, Valmistaja, Valmistuspaikka, Valmistusaika, Käyttäjä ja osa Käyttöpaikkaa. Tässä esinetyypiksi on valittu Astiat ja taloustarvikkeet, jonka alikategoriat Taloustarvikkeet ja Astiat näkyvät seuraavan tason valintamahdollisuuksina. Oikealla olevassa alueessa näytetään haun tuloksena löytyneet kokoelma-kohteet alityypeittäin ryhmiteltynä (tässä Taloustarvikkeet ja Astiat).

Tässä hakuikkunassa hakua voidaan rajoittaa valitsemalla mistä tahansa näkymästä haluttu alikategoria vastaavaa linkkiä painamalla. Järjestelmä hakee silloin kaikki samanaikaisesti eri näkymien valintoja vastaavat esineet ja tulostaa tuloksen. Päinvastoin kuin kuvassa 4.1 hakua ei suoriteta yhden vaan samanaikaisesti monen näkymävalinnan suhteen. Esimerkiksi valitsemalla valmistuspaikaksi ”Pohjois-Amerikka” saadaan tulokseksi kaikki Pohjois-Amerikassa valmistetut astiat ja taloustarvikkeet. Aiemmin tehtyjä kategoriavalintoja voidaan höllentää näkymän otsakkeessa näkyviä yläkategorioita vastaavia linkkejä painamalla. Linkki ”uusi haku” poistaa kerralla kaikki valinnat.

Jokaisen valinnan jälkeen järjestelmä laskee odotettavissa olevien hakutulosten määrän jokaisen mahdollisen jatkovalinnan suhteen ja näyttää määrät suluissa vastaavan linkin perässä. Jos kategorian valinta johtaa tyhjiin joukkoon, estetään valinnan tekeminen kokonaan. Näin voidaan estää se, että käyttäjä joutuisi sellaisiin turhauttaviin tilanteisiin, joissa haku ei tuota lainkaan osumia. Tilanteissa, joissa haku tuottaa hyvin suuren määrän osumia, näytetään tuloksista vain osa ja jatkolinkki, jolla lisää osumia saadaan tarvittaessa näkyville.

Näkymähaku antaa käyttäjälle sanaston, jonka avulla hän osaa järkevästi ja oikein esittää hakukyselyjä järjestelmän tuntemien käsitteiden avulla. Tämä on hyödyllistä silloin, kun käyttäjä ei osaa muotoilla täsmällisesti tiedontarvettaan. Tarvittaessa koko näkymähierarkia nykyvalinnasta alaspäin saadaan näkyviin valintahierarkiana. Tämä tapahtuu painamalla näkymän linkkiä ”koko luokittelu”. Tällöin avautuu kuvan 4.1 kaltainen valintapuu, josta näkyy kaikki valintamahdollisuudet linkkeinä sekä valintoja vastaavien osumien määrät etukäteen laskettuna. Näkymäpuun avulla voi hahmottaa kerralla aineiston jakautumisen näkymän kategorioihin. Esimerkissämme valmistusajan ”koko luokittelu” linkkiä painamalla saataisiin näkyviin Pohjois-Amerikassa valmistettujen astioiden ja taloustarvikkeiden jakauma eri vuosille.

## 4.2 Käsitehaku

Näkymähauulla voi muodostaa yleiskuvan tarjolla olevasta aineistosta sekä muotoilla helposti tehokkaita kyselyitä. Näkymähaku ei kuitenkaan korvaa perinteistä sanahakua vaan on sen täydennys. Sanahaku on tarpeellinen erityisesti tilanteissa, joissa käyttäjä pystyy välittömästi ilmaisemaan tarpeensa sanoina. Jos esimerkiksi etsitään Asko Oy:n valmistamia kohteita, on helpompi kirjoittaa hakukenttään suoraan sanat Asko Oy kuin etsiä Asko Oy-nimistä toimijaa toimijat-ontologian satojen yritysten joukosta.

MuseoSuomeen on siksi toteutettu myös käyttäjän kannalta perinteisen sanahaun kaltaisesti toimiva hakukone. Kone hakee perinteiseen tapaan kaikki kokoelmakohteet, joiden kuvailussa annettu hakusana, esimerkiksi ”Helsinki” esiintyy. MuseoSuomessa tämä hakutapa hyödyntää kuitenkin myös näkymien kategorioita vastaavien ontologisten käsitteiden verkostoa. Esimerkiksi koska hakusana ”Helsinki” sopii paikkanäkymän ”Helsinki” kategorian nimeen, kohdistuu haku myös kohteisiin, joiden käyttö tai valmistuspaikka on jokin Helsingin osa, kuten ”Suomenlinna”.

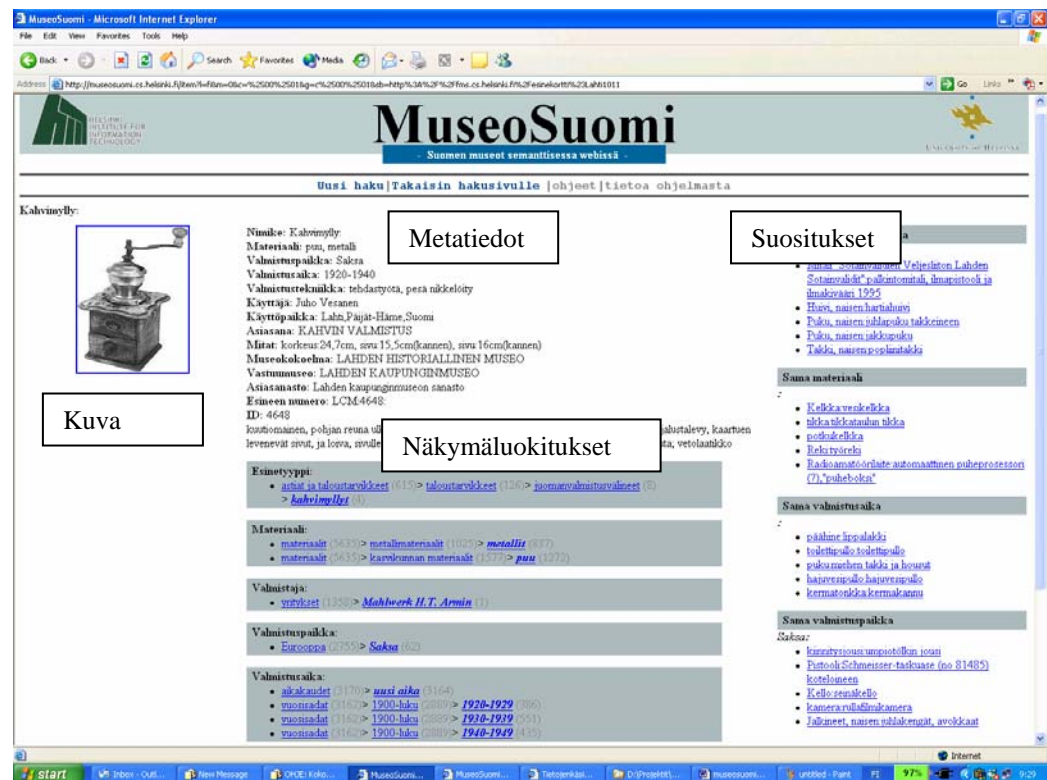
MuseoSuomen käsitehakukone näyttää aina käyttäjälle kaikki eri näkymien kategoriat (polkuineen), joiden nimessä (tai kategorialle erikseen määritellyssä synonyymilistassa) kyseinen sana esiintyy. Valitsemaan jokin näistä, siirtyy ko. valinta normaaliin tapaan vastaavan näkymän uudeksi valintarajoitteeksi.

## 4.3 Semanttinen samoilu

Näkymä- ja käsitehaun avulla käyttäjä voi löytää itseään kiinnostavan, mielekkäästi rajatun kohdejoukon. Tämän jälkeen hän voi hiirellä valita hakutuloksesta jonkin esineen tarkempaa tutustumista varten, jolloin MuseoSuomi näyttää kohteen kuvan ja metatiedot omalla *kohdesivullaan* (ku-

va 4.3.). MuseoSuomen yhtenä innovaationa on tässä yhteydessä tarjota käyttäjälle ”semanttinen selailumahdollisuus” assosiatiivisten suosittelulinkkien avulla. MuseoSuomi voi esimerkiksi suositella linkkejä saman henkilön käyttämiin muihin esineisiin, samaa materiaalia oleviin esineisiin jne. Suositukset perustuvat tietämyskannan semanttisiin yhteyksiin. Näihin perustuen MuseoSuomen kehittäjä voi määritellä esineitä eri tavoin yhdistäviä loogisia sääntöjä, jotka sitten näkyvät loppukäyttäjälle nimettyinä linkeinä.

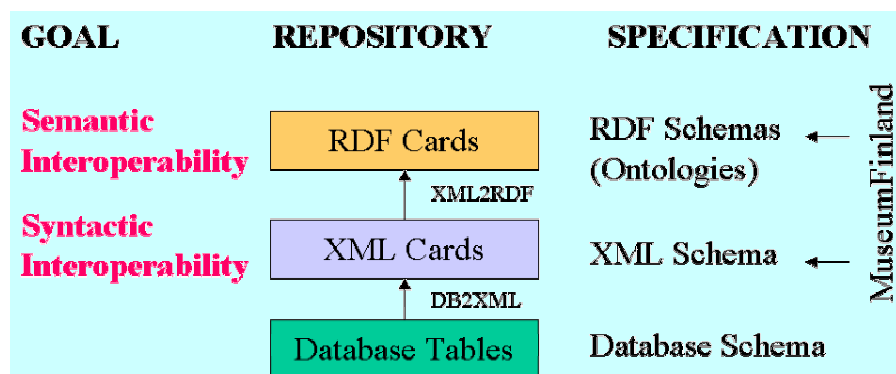
Kohteen (meta)tietojen ja linkkisuositusten lisäksi kohdesivun alareunassa näytetään kohteeseen liittyvät ontologiset luokittelut linkeinä, joita painamalla käynnistyy uusi kategoriaa vastaava näkymähaku.



**Kuva 4.3.** Jokaiselle museokohteelle on oma web-sivunsa, jossa näkyy valitun kohteen kuva, metatiedot (keskellä), semanttiset suositukset toisiin kohteisiin (oikealla) ja ontologiset näkymäluokitukset (alaosassa).

## 5 Kansallisen julkaisukanavan luominen

Kuva 5.1 havainnollistaa sitä prosessia, jolla yksittäisen museon kokoelmatiedot voidaan siirtää MuseoSuomeen. Ensin tuotetaan julkaistavasta kokoelmatiedosta (DB2XML-muunnos) yhteisesti sovitun XML-kielen mukaisia esinekortteja (kuvasssa XML Cards). Tietoteknisen toteutuksen kannalta kunkin museon omalle vastuulla jää muuntimen toteuttaminen omasta kokoelmatietokannasta XML-muotoon. Projektin puitteissa on esimerkkinä toteutettu muuntimet Museoviraston Musketille, Antivariajärjestelmälle ja Espoon kaupunginmuseon Escoll-järjestelmälle. Osia näistä muuntimista voidaan hyödyntää järjestelmään mahdollisesti jatkossa liittyvien muiden museoiden muuntimissa.



**Kuva 5.2.** Kokoelmatietojen siirtäminen MuseoSuomi-järjestelmään.

Seuraavassa vaiheessa muunnetaan XML-kortit MuseoSuomen käyttämään RDF-muotoon (RDF Cards). Tätä muunnosta varten on tarjolla kaksi työkalua:

1. *Terminaattori*-ohjelmiston avulla voidaan määritellä museon kokoelmissa esiintyvät uudet termit, joita MuseoSuomi-järjestelmä ei vielä tunne. Näin syntyy termit ontologisiin käsitteisiin yhdistäviä ns. *termikortteja*.
2. *Annomobiili*-ohjelmiston ja termikorttien avulla XML-esinekortit voidaan muuttaa RDF-muotoon MuseoSuomea varten.

Näiden välineiden käyttäminen ei edellytä ohjelmointitaitoa, mutta jonkin verran kouluttautumista työkalujen käyttöön.

Valmis RDF-muotoinen metadata voidaan sitten toimittaa esimerkiksi CD:llä MuseoSuomeen. XML/RDF rajapinnan avulla museo voi helposti ja täydellisesti kontrolloida niitä tietoja, jotka se haluaa ja voi julkaista. MuseoSuomeen liittyminen ei aseta vaatimusta museon oman museojärjestelmän käyttämisestä webin kautta, mikä aiheuttaisi museoille kustannusten ohella tietosuoja, palomuuuri yms. ongelmia.

## 6 MuseoSuomen jatkokehitys

Hankkeessa kehitettyjä ja kehitettäviä ratkaisuja on tarkoitus hyödyntää Museoviraston syksyllä 2003 käynnistämässä Suomen museot online – hankkeessa.

MuseoSuomeen tullaan lisäämään uudentyyppisiä ontologioita ja aineistoja (mm. taide ja valokuvat) osana Tekesin kansallista ”Suomalaiset semanttisen webin ontologiat” –tutkimushanketta 2003–2005:

<http://www.cs.helsinki.fi/group/seco/ontologies/>

Hankkeessa on mukana 16 julkisen hallinnon organisaatiota ja yritystä.



## 7 Lisätietoja

---

MuseoSuomi-projektin kotisivu:

<http://www.cs.helsinki.fi/group/seco/museosuomi/>

MuseoSuomi-järjestelmää on esitelty tarkemmin alla olevassa englanninkielisessä julkaisussa:

E. Hyvönen, M. Junnila, S. Kettula, E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, K. Viljanen: MuseumFinland - Finnish Museums on the Semantic Web. User's Perspective. Proceedings of Museums and the Web 2004 (MW2004), Arlington, Virginia, USA, March 31 – April 3, 2004 (forthcoming).

<http://www.cs.helsinki.fi/u/eahyvone/publications/MuseumFinland.pdf>

MuseoSuomi-konseptin ensimmäistä sovellusta Helsingin yliopistomuseon valokuva-arkistoon on kuvattu alla olevassa artikkelissa:

E. Hyvönen, S. Saarela, K. Viljanen: Application of Ontology Techniques to View-Based Semantic Search and Browsing. Proceedings of the First European Semantic Web Symposium, May 10-12, 2004, Springer-Verlag (forthcoming).

<http://www.cs.helsinki.fi/u/eahyvone/publications/promoottori.pdf>

MuseoSuomen sisältötyöstä on ilmestynyt myös artikkeli:

E. Hyvönen, M. Salminen, S. Kettula, M. Junnila: A Content Creation Process for the Semantic Web. Paper, submitted, 2004.

<http://www.cs.helsinki.fi/u/eahyvone/publications/contentCreation.pdf>

MuseoSuomi-aineistoja voidaan julkaista semanttisessa webissä myös käyttämällä projektissa kehitettyä SWeHG-työkalua:

E. Hyvönen, M. Holi, K. Viljanen, A. Valo: A Logic-Based RDF to HTML Generator for Spinning the Semantic Web. Paper, submitted, 2004.

[http://www.cs.helsinki.fi/u/eahyvone/publications/swehg\\_ecai2004.pdf](http://www.cs.helsinki.fi/u/eahyvone/publications/swehg_ecai2004.pdf)

## 8 Yhteystiedot

---

Semanttisen laskennan tutkimusryhmän kotisivu (SeCo):

<http://www.cs.helsinki.fi/group/seco/>

Yhteyshenkilö: Prof. Eero Hyvönen, Helsingin yliopisto ja HIIT, [eero.hyvonen@cs.helsinki.fi](mailto:eero.hyvonen@cs.helsinki.fi), <http://www.cs.helsinki.fi/eero.hyvonen/>