

*Proceedings of Museums and the Web 2004 (MW2004), Arlington, Virginia, USA, March 31 – April 3, 2004 (forthcoming).*

## **MuseumFinland — Finnish Museums on the Semantic Web User's Perspective**

**Eero Hyvönen, Miikka Junnila, Suvi Kettula, Eetu Mäkelä, Samppa Saarela,  
Mirva Salminen, Ahti Syreeni, Arttu Valo, and Kim Viljanen**

**Helsinki Institute for Information Technology (HIIT)  
University of Helsinki  
P.O. Box 26, 00014 UNIV. OF HELSINKI, Finland**

**Email: [firstname.lastname@cs.helsinki.fi](mailto:firstname.lastname@cs.helsinki.fi)  
<http://www.cs.helsinki.fi/group/seco/>**

### **Abstract**

This paper presents a semantic portal, MuseumFinland, for publishing heterogeneous museum collections on the Semantic Web. The application is presented from the viewpoints of the end-user and the museums providing the contents. By semantic web techniques, it is possible to make collections semantically interoperable and provide the museum visitors with intelligent content-based search and browsing services to the global collection base. By using the MuseumFinland approach the museums with their semantically rich and interrelated collection content can create consolidated semantic collection portals together on the web.

### **Why Museums on the Semantic Web?**

A special characteristic of cultural collection contents is semantic richness. Collection items have a history and are related in many ways to our environment, to the society, and to other collection items. For example, a chair may be made of oak and leather, may be of a certain style, was designed by a famous designer, was manufactured by a certain company during a time period, was used in a certain castle together with other pieces of furniture, and so on. Other collection items, locations, time periods, designers, companies etc. can be related to the chair through their properties and implicitly constitute a complicated semantic network of associations. This semantic network is not limited to a single collections but spans over other related collections in other museums.

The Semantic Web (Berners-Lee et al., 2001; Fensel et al., 2002) is the next generation of the Web where the contents are meant not only to a human reader but for the machines to

interpret. The key idea is to represent the contents of the web by explicit metadata structures that conform to mutually agreed vocabularies, ontologies (Sowa, 2000). Semantic web technology (<http://www.w3.org/2001/SW/>) enables new possibilities when publishing museum collections on the web (Hyvönen et al., 2002):

- **Intelligent applications.** Firstly, intelligent applications based on the semantics of the collections can be created.
- **Collection interoperability in content.** Secondly, web languages, standards, and ontologies make it possible to make heterogeneous museum collections of different kind mutually interoperable. This enables, e.g., the creation of large inter-museum exhibitions.

To realize these ideas in practice, we have developed a semantic web portal called “MuseumFinland—Finnish Museums on the Semantic Web”. This system contains an inter-museum exhibition of cultural artifacts, such as textiles, pieces of furniture, tools etc. The contents for the pilot version come from the collections of the National Museum (<http://www.nba.fi>), Espoo City Museum (<http://www.espooli.fi/museo>), and Lahti City Museum (<http://www.lahti.fi/Kulttuuri/museot>). These museums are situated in different cities, use three different relational database schemas, data base systems, and collection management systems (called Musketti, Escoll, and Antikvaria, respectively).

In the following we first describe the knowledge-based services of the portal from the end-user’s viewpoint. The collections form for the end-user a seamless repository of web pages to search and browse with an ordinary web browser. After this it is shown how the heterogeneous distributed collections of the museums that participate in the system can be merged together in an interoperable way. For a participating museum, the portal provides a channel to publish content easily and independently from the museum’s collection database system. In conclusion, main results of the work are summarized, lessons learned discussed, and directions for further research outlined.

## A Semantic Search Engine

MuseumFinland provides the end-user with two major services.

- A **semantic view-based search engine** that is based on the underlying concepts and ontologies instead of simple keywords.
- A **semantic recommendation system** by which the user can find out explicit and implicit semantic associations within the global collection data, and use the associations for browsing the collections.

In this section the search engine is shortly discussed. Semantic recommendations are considered after this.

The metadata of collection objects in a museum database is described by using named properties, such as the 15 properties of the Dublin Core standard (<http://dublincore.org/>). The value of a property can be, for example, an integer representing the year of publication of a document or a free text description about the history of an artifact. When possible, it is beneficial to describe subject content by using *keywords* selected from controlled vocabularies or thesauri (Foskett, 1980). This keeps metadata descriptions coherent and in this way significantly eases information retrieval later.

The search engine of MuseumFinland, called Ontogator, is based on the multi-facet search paradigm (Pollitt, 1998; Hearst et al., 2002). Here the keywords or concepts used for indexing are called *categories* and are organized systematically into a set of hierarchical, orthogonal taxonomies. For example, artifact types, such as furniture, cloths, weapons etc. can be into a taxonomy. The taxonomies are called subject *facets* or *views*. A search query in view-based search is formulated by selecting categories of interest from the different facets. For example, by selecting the category Clothing from an Artifact facet, category Cotton from a Material facet, and category 1800-1900 from a Time facet, the user can express the query for retrieving all trousers, skirts, and other clothing made of cotton in the 19<sup>th</sup> century. Intuitively, the query is a conjunctive constraint over the facets with disjunctive constraints over the sub-categories in each facet.

VIEW TYPE	VIEW NAME	ONTOLOGY
Object Views	Artifact	Artifacts
	Material	Materials
Creation Views	Creator	Actors
	Place of creation	Locations
	Time of creation	Times
Usage Views	User	Actors
	Place of usage	Locations
	Situation	Events
Collection View	Collection	Collections

**Table 1. Orthogonal View-Facets in MuseumFinland.**

MuseumFinland classifies the collection objects along 9 views organized in four groups (table 1). The Object Views describe the physical aspects of the collection item (artifact type and materials). The Creation Views tell who manufactured or created the object, as well as the location and time of the creation. The Usage Views indicate the user of the object, place of usage, and situations in which the object is used. Finally, the Collection View classifies the museums and collections participating in the portal.

ONTOLOGY	CONTENT	CLASSES	INDIVIDUALS
Artifacts	Classes for tangible collection objects	3227	0
Materials	Substances that the artifacts are made of	364	0
Actors	Persons, companies, organizations, and other agents	26	1715
Locations	Continents, countries, cities, villages, farms etc.	33	864
Times	Eras, centuries, etc. as time intervals	57	0
Events	Situations, events, and processes in the society	992	0
Collections	Museum collections included in the system	22	24

**Table 2. User interface for view-based multi-facet search in MuseumFinland.**

The views can be projected from a set of ontologies listed in the rightmost column of table 1. The contents of the ontologies and their sizes in the pilot version online at the moment are given in table 2. The *Artifacts* ontology is a taxonomy of the tangible collection objects such as pottery, cloths, weapons, etc. All exhibits in the system belong to some class in this ontology. The *Materials* ontology is a taxonomy of the artifact materials, such as steel, silk, tree, etc. The *Actors* ontology defines classes of agents, such as persons, companies etc., and individuals as instances of these classes. The *Events*

ontology include intangible happenings, situations, events, and processes that take place in the society, such as farming, feasts, sports, war, etc. *Locations* is an ontology representing areas and places on the Earth and in Finland in particular. The *Times* ontology is a taxonomy of various predefined historical periods, and the *Collections* ontology classifies the museums and collections in the portal. The Artifacts, Materials, and Events ontologies are subsets of a larger cultural ontology called MAO (6768 classes) that we created based on the Finnish cultural thesaurus MASA (Leskinen, 1997). MASA is widely used in Finnish museums for describing and classifying collection objects. It was a natural choice for the basis of a general cultural semantic web ontology and vocabulary. The ontology and vocabulary work underlying MuseumFinland is described in more detail in (Hyvönen et al., 2004).



Figure 1. The initial search interface of MuseumFinland with its nine facets.

Figure 1 shows the initial search interface of MuseumFinland. The nine facet hierarchies of table 1 are shown (in Finnish), such as Artifact (“Esinetyyppi”) and Material (“Materiaali”). For each facet hierarchy, the next level of sub-categories is shown as links. A query is formulated by selecting a sub-category by clicking on its name. When the user selects a category  $c$  in a facet  $f$ , the system constrains the search by leaving in the result set only such objects that are annotated in facet  $f$  with some sub-category of  $c$ . For example, figure 2 depicts the situation after selecting the sub-category Tools (“työvälineet”) from the Artifact facet (“Esinetyyppi”). The result set is shown on the right grouped by the sub-categories of Tools, such as Textile making tools (“tekstilityövälineet”) and Tools of folk medicine (“kansanlääkinnän työvälineet”). Hits in different the categories are separated by horizontal bars and can be scrolled independently in each category. In this case, all categories do not fit in the screenshot.

The screenshot shows the MuseoSuomi search interface. At the top, there is a search bar with the text "Hae" and a "Uusi haku" button. Below the search bar, there are several facets for filtering results. The facets include:

- Esintyyppi:** [kalkki > työvälineet \(koko luokittelu\)](#) (245), [tehtäviätyövälineet](#) (125), [muut työvälineet](#) (29), [puutyövälineet](#) (2), [maalausvälineet](#) (2), [työkäsit](#) (143).
- Materiaali:** [kansainvälinen työvälineet](#) (1), [maalustyövälineet](#) (180), [pilkkon- ja leikontarvikkeet](#) (23), [kirjoitusvälineet](#) (10), [metallityövälineet](#) (29).
- Valmistaja:** [henkilöt](#) (20), [tuotemerkit](#) (3), [yritykset](#) (40).
- Valmistuspaikka:** [Etelä-Amerikka](#) (1), [Eurooppa](#) (26).
- Valmistusaika:** [aikakaudet](#) (142), [vuosiedet](#) (142).
- Käyttäjä:** [henkilöt](#) (58), [kaupungit](#) (1), [laitokset](#) (2), [yritykset](#) (8).
- Käyttöpaikka:** [Aasia](#) (1), [Eurooppa](#) (579).
- Käyttötilanne:** [kansalais-, harrastus- ja vapaa-aikatoiminta](#) (5), [juhlat ja seremoniat](#) (1), [matalous ja karjanhoito](#) (9), [valmistustekniikat](#) (9), [kohteille tehtävät toimenpiteet](#) (15), [elinkeinot](#) (21).
- Kokoelma:** [Espann kaupunginmuseon kokoelmat](#) (71), [Kansallismuseon kokoelmat](#) (643), [Lahden kaupunginmuseon kokoelmat](#) (47).

The main results area shows a list of tools with their IDs and names. The first two results are:

- [työväline \(NBA:11246:1847\)](#)
- [työväline \(NBA:11246:2074\)](#)

Below these, there are more results, including:

- [kirkonkellotyövälineet, kohteet 1-4/245 \(ryhmitteelle kohteet\)](#)
- [keuhkulehto, keuhkupum, kuuszel, huosah \(NBA:SU5069:26\)](#)
- [ruukinläpa \(ECM:100:1\)](#)
- [ruukinläpa \(NBA:1159\)](#)
- [flagabre, ruukinläpa \(NBA:4331:14\)](#)

At the bottom, there is a link for [Kansainvälinen työvälineet, kohteet 1-1/1 \(ryhmitteelle kohteet\)](#) and an image of a tool.

Figure 2. The search interface of MuseumFinland after selecting link Tools (“työvälineet”) in figure 1.

The facets are shown on the left. When answering the query, the result set for each direct sub-category in the facets seen on the screen is recomputed, and a number ( $n$ ) is shown to the user after the category name. It tells that if the sub-category is selected next, then there will be  $n$  hits in the result set. For example, in figure 2, the number 643 in the Collection facet on the bottom (“Kokoelma”) tells that there are 643 tools in the collections of the National Museum (“Kansallismuseon kokoelmat”).

A selection leading to an empty result set ( $n=0$ ) is removed from its facet (or alternatively disabled and shown in gray color, depending on the user’s preference). In this way, the user can be hindered from making a selection leading to an empty result set, and is guided toward selections that are likely to constrain the search appropriately. The query can be relaxed by making a new selection on a higher level of the facets or by dismissing the facet totally from the query.

In above, the category selection was made among the direct sub-categories listed in the facets. An alternative way is to click on the link Whole facet (“koko luokittelu”) on a facet. The system then shows all possible selections in the facet with hit counts. For example, in figure 3 the user selected in the situation of figure 2 the link Whole facet of the facet Time of creation (“Valmistusaika”). The system shows how the tools in the current result set are classified according to the Times facet. The facet is represented by a link hierarchy from which the search can be either constrained further or relaxed by

clicking on a category link. For example, by selecting the category 1840-1849, the tools manufactured during that decade are found.

The screenshot shows the MuseoSuomi website interface. At the top, there are logos for the Helsinki Institute for Information Technology and the University of Helsinki. The main header reads "MuseoSuomi - Suomen museot semanttisessa webissä". Below the header, there is a navigation bar with links: "Uusi haku | Takaisin hakusivulle | näy kaikki kategoriat | ohjeet | tietoa ohjelmasta". The main content area is titled "Hakuehdot" and shows a search filter: "Kategoria: Esinetyyppi > työvälineet (ryhmittele kohteet) (poista)". Below this, the category "Kategoria: Valmistusaika" is expanded, showing a hierarchical list of time periods with item counts in parentheses:

- aikakaudet (142)
  - esihistoriallinen aika (2)
    - meroväikasaika (1)
    - rautakausi (2)
      - kansainvaellusaika (1)
      - viikinkiaika (1)
    - varhaismetallikausi (2)
  - historiallinen aika (142)
    - keskiaika (1)
    - uusi aika (142)
      - sodat (23)
        - I maailmansota (12)
        - II maailmansota (13)
- vuosikaudet (142)
  - 1600-luku (4)
  - 1700-luku (21)
    - 1700-1749 (7)
    - 1750-1799 (2)
  - 1800-luku (49)
    - 1800-1809 (8)
    - 1810-1819 (7)
    - 1820-1829 (11)
    - 1830-1839 (8)
    - 1840-1849 (11)
    - 1850-1859 (11)
    - 1860-1869 (15)

**Figure 3. The Time facet hierarchy classifying the result set of tools in figure 2.**

In this way, the user

- can easily formulate the query using the right categories exposed to her as links, and
- can get easily overviews of the database contents along different classifications in different situations.

User studies (Lee et al., 2003; English et al., 2003) have recently been carried out to show that if the user does not know precisely what objects she is looking for, then the multi-facet search method with its browsing the shelves sensation is clearly preferred over keyword search or using only a single facet. The latter approach is commonly used for finding resources on the web, e.g., in Yahoo and in the Open Directory Project (<http://dmoz.org>).

However, if the user is capable of expressing her information need straightforward in terms of keywords, then a Google-like keyword search interface is usually faster and preferred. To support word-based search, too, an additional search engine was implemented in MuseumFinland. This engine is used for two purposes at the same time:

- For searching categories to be used in multi-facet search
- For searching collection objects with matching metadata values in the conventional way

The problem of finding relevant categories in the facets is a search problem of its own when dealing with thousands of categories. The user may then type in keywords in the search text field labeled "Haku" in the left upper corner of figure 1. In response, all

categories whose names match with the input keywords (substring match) are selected and shown to the user as links in the form:

*FacetName* > *category*

For example, a search with the string “ase” gives the link set shown in figure 4.

**Haku**


**Hakuehdot**

**Hakusana: ase (ryhmittele)**

**Kohteet ryhmiteltyinä hakusa (näytä ilman ryhmittelyä) (k)**  
Esinetyyppi > [aseet](#), kohteet 1-4

**Hakusana: ase (poista)**  
Esinetyyppi > [aseet](#) (23),  
Esinetyyppi > [aseet ja ampumatarvikkeet](#) (39),  
Materiaali > [asettaatti](#) (1)

**Esinetyyppi (koko luokittelu) (ryhmittele kohteet)**  
[aseet ja ampumatarvikkeet](#) (39), [astiat ja taloustarvikkeet](#) (16),  
[asumisen esineet](#) (15), [henkilökohtaiset esineet](#) (15),  
[kulkuvälineet ja kulkuvälineet noineen](#) (12)



**Figure 4. Using the word search for finding categories.**

Here the query matched the categories “guns” (“aseet” in Finnish) and “guns and shooting equipment” (“aseet ja ampumatarvikkeet”) in the facet Artifact (“Esinetyyppi”), and the category “acetate” (“asettaatti”) in the Material facet (“Materiaali”). By selecting a link, the multi-facet search can be executed. For example, by clicking on “aseet” in figure 4, the 23 guns in the RDF repository are retrieved.

When executing word-based search, the search engine also performs a combined keyword and multi-facet search in the following way. The *union* of the found categories (cf. figure 4) is used as a query in the multi-facet search resulting in a result set R1. In addition, a conventional keyword match search is performed for (some) property values of the objects, resulting in another result set R2. The search result shown to the user is  $R1 \cup R2$ . In our example, guns and shooting equipment together with objects made of acetate are retrieved.

By default, the categories shown in the search are grouped by the last selection, but the system also supports grouping based on arbitrary categories and keywords. For example, in figure 5, the results of the search of figure 2 are shown grouped by museum collection. This provides the user a quick and intuitive view on what kind of tools there are in each collection of the participating museums.



### Kohteet ryhmiteltyinä kategorian *Kokoelma* mukaisesti

(näytii ilman ryhmittelyä) (kaikkien 762 kohteen tiedot)

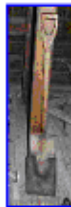
[Espoon kaupunginmuseon kokoelmat](#), kohteet 1-4/71 (ryhmittele kohteet)



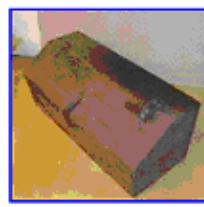
rukinlapa (ECM:100:1)



kuokka:kuokka (ECM:1274:1)



rautaterälapio:rautaterälapio  
(ECM:1296:1)



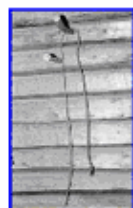
työkalulaatikko:työkalulaatikko  
(ECM:1488:1)

(edellinen) / (seuraava)

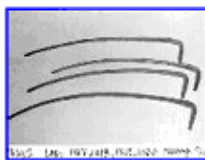
[Kansallismuseon kokoelmat](#), kohteet 1-4/643 (ryhmittele kohteet)



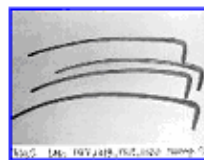
kuokka (NBA:10655:2190)



nahkaremmi  
(NBA:11246:1314)



vikatteen terä  
(NBA:11246:1317)



vikatteen terä  
(NBA:11246:1318)

(edellinen) / (seuraava)

[Lahden kaupunginmuseon kokoelmat](#), kohteet 1-4/47 (ryhmittele kohteet)



Figure 5. Search results from figure 2 after selecting the link “group by” (“ryhmittele kohteet”) from the Museum Collection facet (“Kokoelma”).

## A Semantic Recommendation System

At any point during multi-facet search the user can select any hit found by clicking on its image. The corresponding collection object is then shown as a web page, such as the one in figure 5. It depicts a special part, distaff (“rukinlapa” in Finnish), used in a spinning wheel. The page contains the following information and links:

1. The image(s) of the object is (are) depicted on the left.
2. The metadata of the object shown in the middle on top.
3. All facet categories of the object are listed in the middle bottom as hierarchical link paths. A new search can be started by selecting any link from there.
4. A set of semantic links on the right provided by a semantic recommendation system.





# MuseoSuomi

- Suomen museot semanttisessa webissä -



UNIVERSITY OF HELSINKI

---

Uusi haku | Takaisin hakusivulle | ohjeet | tietoa ohjelmasta

---

**ruukinlapa**



**Nimike:** ruukinlapa  
**Valmistuspaikka:** Suomi  
**Valmistusaika:** 1793-1793  
**Käyttöpaikka:** Suomi, Bemböle, Espoo, Suomi, Vanhakartano, Espoo, Suomi  
**Asiasana:** KEHRUU, KORISTEVEISTO, PUUMERKKI, VUOSILUKU  
**Museokokoelma:** Museokokoelma  
**Vastuumuseo:** Espoon kaupunginmuseo  
**Asiasanasato:** Espoon kaupunginmuseon sanasto  
**Esiineen numero:** ECM.100.1  
**ID:** 1001

**Esinetyyppi:**

- [työvälineet](#) (767) > [tekstiiliäestövälineet](#) (245)
- > [kehrurin ja langanvalmistuksen työvälineet](#) (116) > [kehrurivälineet](#) (66) > [kuontalopittimet](#) (45)
- > [ruukinlavat](#) (36)

**Valmistuspaikka:**

- [Eurooppa](#) (2656) > [Suomi](#) (2370)

**Valmistusaika:**

- [1800-luvut](#) (3050) > [historiallinen aika](#) (3047) > [uusi aika](#) (3047)
- > [vuosikadat](#) (3045) > [1700-luku](#) (36)

**Käyttöpaikka:**

- [Eurooppa](#) (3959) > [Suomi](#) (3932)
- > [Eurooppa](#) (3959) > [Suomi](#) (3932) > [Etelä-Suomen lääni](#) (2133) > [Uusima-Nyland](#) (648)
- > [Espoo](#) (407)
- > [Eurooppa](#) (3959) > [Suomi](#) (3932) > [Etelä-Suomen lääni](#) (2133) > [Uusima-Nyland](#) (648)
- > [Espoo](#) (407) > [Bemböle](#) (14)

**Käyttötilanne:**

- [valmistustekniikat](#) (297) > [häsiyty](#) (230) > [tekninen työ](#) (3) > [väjätö](#) (3) > [koristeväsite](#) (3)
- > [valmistustekniikat](#) (297) > [häsiyty](#) (230) > [tekstiilitö](#) (219) > [kuututyo](#) (63) > [kehruus](#) (52)

**Kokoelma:**

- [Espoon kaupunginmuseon kokoelmat](#) (1369)

**Sama käyttöpaikka**

**Bemböle:**

- [jämsävuolin](#)
- [opetusväline peli](#)
- [opetusväline peli](#)
- [opetusväline peli](#)
- [opetusväline peli](#)

**Espoo:**

- [korokirja korokirja kangasta](#)
- [leimuki lapsen lyhytelämäinen leimuki](#)
- [neulelakki naisen neulelakki](#)
- [hartiavaste naisen pitsinen hartivaste](#)
- [pöytälaiva yläosa, jalka naisen pöytälaiva yläosa](#)

**Suomi:**

- [ruokailina ruokailina, damasetti](#)
- [kattailina kattailina, etupietokirjonta](#)
- [pöytälaiva pikkuhina, kirjottu](#)
- [pöytälaiva ristipietokirjontainen pöytälaiva](#)
- [kattailina bestillina, kirjottu](#)

**Samanlaisia esineitä**

**kehruus:**

- [jakkara kehrurijakkara](#)
- [rullatuoli rullatuoli](#)
- [puola kehruripuu](#)
- [puola kehruripuu](#)
- [kukki kehruripuu](#)

**koristeväsite:**

- [pöytälaiva puinen kangaspuiden pöytälaiva](#)
- [tuoppi puutuoppi](#)

**puumerkit:**

- [punnari](#)

**Sama valmistusaika**

1793:

- [diagonaali ruukinlapa](#)

**Figure 6. Web page depicting a collection object, its metadata, facet categories, and semantic recommendation links to other collection object pages.**

Semantic recommendations reveal to the end-user a most interesting aspect of the collection items: the implicit semantic relations that relate collection data with their context and with each other. The recommendation links provide a *semantic browsing* facility to the end-user. For example, in figure 6 there are links to objects used at the same location (categorized according to the name of the common location), to objects related to similar events (e.g., objects used in spinning, and decorative objects, because distaffs are usually beautifully decorated), to objects manufactured at the same time, and so on. Since a decoratively carved distaff used to be a typical wedding gift in Finland, it is also possible to recommend links to other objects used as wedding gifts, such as wedding rings. In MuseumFinland, such associations can be exposed to the end-user as link groups whose titles and link names explain to the user the reason for the recommendation. The possibilities for creating such associations are intriguing. Of course, only links that can be inferred based on the metadata and ontologies available can be created.

Recommendations are defined in terms of flexible logical predicate rules using the methods described in (Hyvönen et al., 2003, 2004). The links can be *explicit* or *implicit*. Explicit links correspond to the RDF statements (triples) in the underlying knowledge base and are directly based on the collection domain ontologies (classes and their properties) and the actual collection data (instance data). For example, an instance of a painting may have the RDF property “creator” linking the art work to an individual artist. Implicit links can be defined in terms of explicit ones but are not present in the RDF graph. For example, if there are explicit links linking children with their mothers and fathers, then implicit links such as “grandfather” or “cousin” can be defined.

The semantic recommendation system of MuseumFinland is implemented as a logic server called “Ontodella”. This system is based on the HTTP server version of SWI-Prolog (<http://www.swi-prolog.org>) (Wielemaker et al., 2003). The MuseumFinland system itself is Cocoon-based server (<http://cocoon.apache.org>) that queries with the Ontogator search engine server and Ontodella server with XML/RDF messages. It is possible to do this over HTTP.

There is also a prototype implementation of MuseumFinland that can be used with WAP 2.0 compatible mobile telephones. The current prototype recreates all functionality of the web interface in a layout more suitable to the limited screen space of mobile devices, as seen in figure 7. When the user makes a selection for the multi-facet search, impossible category choices leading to empty results can be pruned out. This is a very useful feature for devices that have a small screen to display choices. Future work will add further mobile-specific features, most notably support for mapping the geographical location of the phone to the Location facet. This provides a quick access to objects that were created, were used, or reside near the current location of the mobile user.

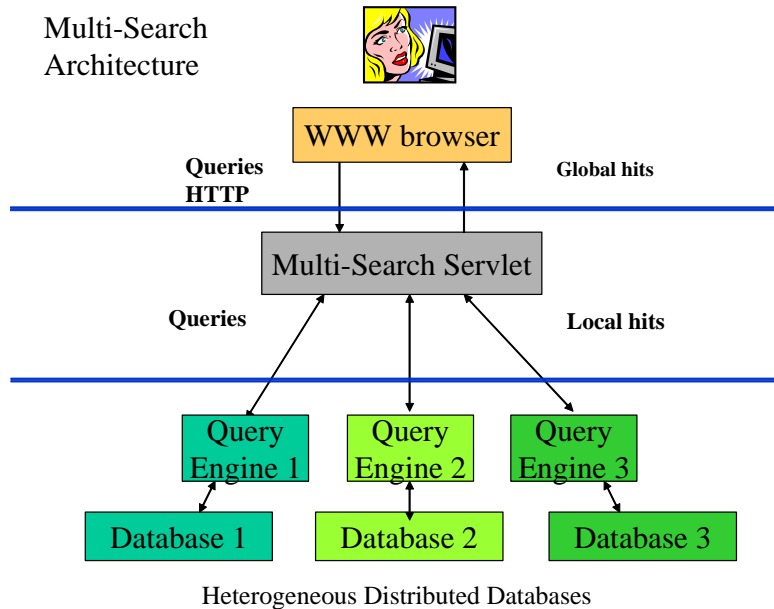


*Figure 7. MuseumFinland search results for a search on 20th century sporting and game items grouped by the Times facet as seen on a Nokia Series 60 browser*

## **Making Museum Collections Interoperable on the Web**

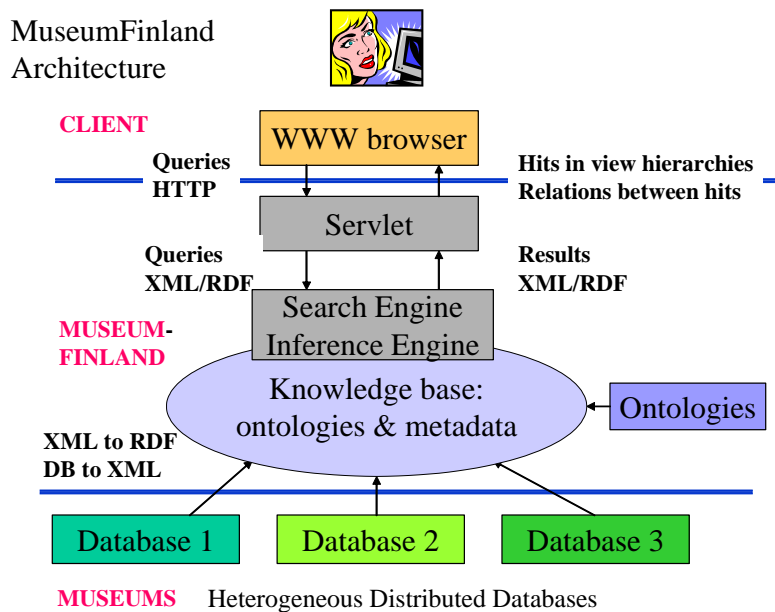
Museums and their collection databases are usually situated at different locations. This creates an obstacle to information retrieval for both the public and for researchers. To address the problem, the web can be used for creating a single interface and access point through which a search query can be sent to distributed local databases and the results combined into a global hit list. This “multi-search” approach, as depicted in figure 8, has

been widely applied, and there are many cultural collection systems on the web based on it, such as the portals Australian Museums Online (<http://www.amonline.net.au/>) and Artefacts Canada (<http://www.chin.gc.ca/>).



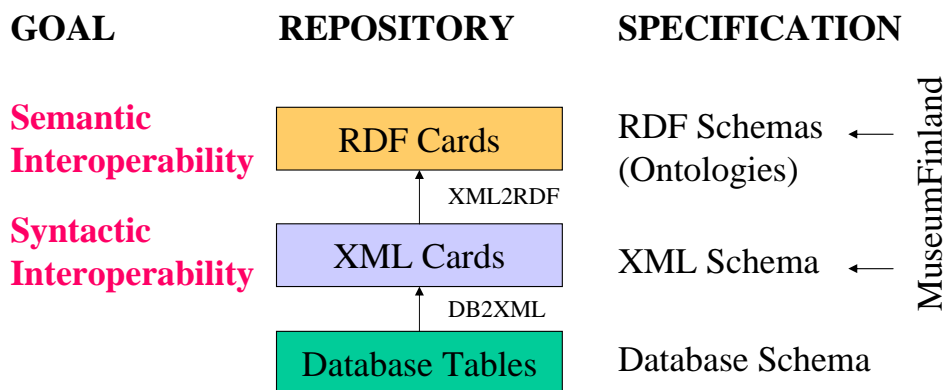
**Figure 8. Multi-search architecture. The global query is answered independently at each local database.**

A problem of multi-search is that by processing the query independently at each *local* database, the *global* dependencies, associations between objects in different collections are difficult to find. Exposing such global semantic associations between collection items is one of the main goals of MuseumFinland. The system cannot therefore be based on the traditional multi-search paradigm. Instead, the local collections are first consolidated into a global repository, and the queries are answered based on it, as illustrated in figure 9.



**Figure 9. Information retrieval in MuseumFinland. Local database contents are first merged and the query is evaluated with respect to the global interrelated data.**

Museums join the system by producing collection metadata in RDF format from their databases. Figure 10 depicts the process. The database contents are first transformed into XML form. Next, the XML is transformed into the final RDF metadata form used by the portal. In below we motivate and describe these transformations briefly; a more detailed description can be found in (Hyvönen et al., 2004).



**Figure 10. Transforming databases into RDF format.**

### Database to XML Transformation

The idea of the database to XML transformation (Raatikka and Hyvönen, 2002) is to re-represent selected database content in XML format defined by an XML Schema. It tells what (meta)data must be provided for describing collection items. The motivation for using an explicit XML level here is to provide a simple, open language by which the

participating museums can agree upon the syntax for representing collection data. Based on the schema, each collection item has an XML description of its own called the *XML card*. For example, the XML card representing a calendar is presented below. (The example is translated and slightly simplified from the original version in Finnish.)

```
<artifactCard created="2003-7-29 10:43:16">
  <artifactId> ECM:22461:1 </artifactId>
  <artifactType> Christmas calendar,
    Finland's Scouters Assoc. </artifactType>
  <museum> Espoo City Museum </museum>
  <material> cardboard </material>
  <keywords>
    <keyword> Christmas </keyword>
    <keyword> calendar </keyword>
    <keyword> scouts </keyword>
  </keywords>
  <placeOfUsage> Tapiola, Espoo </placeOfUsage>
  <creator> Ulla Vaajakoski </creator>
  ...
  <photo> photos/image3451.jpg </photo>
</artifactCard>
```

An XML card presents the main *features* of a collection object by sub-elements. The values of the features, such as the string “Espoo City Museum” in the sub-element <museum>, are read from the underlying database tables.

### XML to RDF Transformation

Each XML card with its string-valued feature values is transformed into an *RDF card* with similar RDF properties, but where the string values are transformed into the Uniform Resource Identifiers (URI) of the corresponding classes and individuals in the ontologies. For example, the XML card above in RDF form is:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:card="http://www.fms.fi/RDFCard#">
  <card:RDFCard
    rdf:about="http://www.fms.fi/rdfCard#card11023">
    card:artifactId="16851"
    card:artifactType-www="calendar"
    card:artifactType="http://www.fms.fi/artifacts#calendar"
    card:museum-www="Espoo City Museum"
    card:museum="http://www.fms.fi/agents#EspooCityMuseum"
    card:material-www="cardboard"
    card:material="http://www.fms.fi/materials#cardboard"
    ...
  </card:RDFCard>
  ...
</rdf:RDF>
```

The features of collection items fall in two categories: *literal* features and *ontological* features. The value  $x$  of each feature  $p$  in the XML card (e.g., material value “cardboard”) is represented by the corresponding literal property  $p\text{-www}=x$  in the RDF card (e.g., material-www=”cardboard”). Literal property values will be shown to the user in the user interface (cf. the metadata values in the middle on top in figure 6). In addition, each ontological feature in the XML card will be represented by an additional *ontological property* with same name in the RDF card. Its value is a URI that relates the card to the ontological RDF resource(s) in the underlying knowledge base. The classes and individuals referred to in the RDF card are defined by the set of RDFS ontologies of table 2.

For example, the feature `artifactId` is literal and is not connected with the ontology resources in the above RDF card. In contrast, the ontological feature `material` is represented with a literal property `www-material` and the ontological property `material` that has an RDF resource (URI) as its value. This URI connects the card resource with the material ontology and through it with other resources.

Two tools have been implemented for facilitating the XML to RDF transformation: Terminator and Annomobile. Terminator is used for creating *term cards*, that essentially define a mapping between words and expressions used at the XML level and the corresponding ontological concepts (URIs). Term cards make MuseumFinland flexible with respect to variance in terminologies used at different museums and by different catalogers. The museums can keep their local terminological conventions as long as they tell the meaning (URI) of their own terms by term cards.

Given a set of term cards and ontologies, Annomobile performs XML to RDF transformation. This cannot be done fully automatically due to unknown terms and complicated descriptions encountered in the databases and homonymous terms. Annomobile can, however, identify such situations and point them out to a human editor that has to make the right decisions and corrections by hand. In our work, we have used the Protégé-2000 ontology editor (<http://protege.stanford.edu>) for editing the ontologies, term cards, and RDF cards. Its user interface is simple enough to be used by museum personnel that usually do not have programming skills.

An important side effect of the XML to RDF transformation is *semantic enrichment* where new meaning is automatically added to the collection data in three ways. Firstly, semantic associations between related collection item instances emerge automatically by shared resources (URIs). For example, a particular bench from a museum A may have the same manufacturer as a footstool in another museum B, which may be an important piece of information to the user. Secondly, generic ontological relations defined for the classes are automatically inherited by instance data. For example, the class `Ylioppilaslakit` (a special student’s white cap used in Finland) has a property that relates the concept to the class `Ylioppilasjuhlat` (a graduation ceremony event). This means that all individuals of the class `Ylioppilaslakit` will have this property as well. As a result, for each student cap, recommendation links to other objects related with graduation ceremonies can be created. The museum cataloger does not have to provide this information as a piece of metadata for each individual cap. Thirdly, the knowledge base can be enriched semantically by

logical rules defining expert domain knowledge concerning the exhibition. In our case, such rules have been used as a basis for the semantic recommendations.

## Discussion

### Contributions

This paper presented an overview of MuseumFinland from the end-user's and museum's viewpoints. In our work, the use of ontologies and semantic web technologies turned out to be useful in many ways:

- **Exact definitions.** By using ontologies, museums can define the concepts used in cataloging in a precise, machine understandable way.
- **Terminological interoperability.** The terms used in different institutions can be made mutually interoperable by mapping them onto common shared ontologies.
- **Ontology sharing.** Ontologies provide means for making exact references to the external world. For example, the Locations ontology (villages, cities, countries, etc.) and the Actors ontology (persons, companies, etc.) is shared by the museums in order to make the right and interoperable references.
- **Automatic content enrichment.** Ontological class definitions, rules, and consolidated metadata enrich collection data semantically.
- **Intelligent services.** Ontologies could be used as a basis for intelligent services to the end-user, in our case for the semantic search engine and the recommendation system.

The first pilot version of MuseumFinland shows that underlying ideas presented in this paper are feasible and that the technology scales up at least to the order 10,000 of cards and view categories. The response times for search queries have typically been under 2 seconds on an ordinary PC server.

### Related Work

The idea of the view-based multi-facet search has been developed, e.g., in (Pollitt, 1998; Hearst et al., 2003). The novelty of MuseumFinland lies in its capability of using RDF(S) ontologies and inference rules as the basis of search. The idea is to combine virtues of the view- and ontology-based search paradigms (Hyvönen et al., 2003). The multi-facet search algorithm of Ontogator itself is independent from the underlying ontologies and their semantics. The "semantic" flavor of the system is based on the Ontodella Prolog-server and its knowledge base in two ways. Firstly, the facet hierarchies for Ontogator are created by a set of logical rules that depend on the ontologies used. For example, a facet can be based on a subclass-of or a part-of property. Secondly, the museum objects are associated with the facet categories by using the underlying ontological relations and inference rules.

This idea of linking collection items with semantic associations is related to Topic Maps (Pepper, 2000). However, in our case the links are not given by a topic map but are determined by logical inference using the underlying RDFS ontology and RDF metadata. Another application of this idea to generating semantically linked static HTML sites from



RDF(S) repositories is presented in (Hyvönen et al., 2003b). In the HyperMuseum (Stuer et al., 2001), collection items are also semantically linked with each other. Here linking is based on shared words in the metadata and their linguistic relations, such as synonymy and antonymy. In contrast, our system is not based on words but on ontological references in the underlying RDF(S) knowledge base. As a result, the links can be defined freely in terms of logical rules. The idea of annotating cultural artifacts in terms of multiple ontologies has been explored, e.g., in (Hollink et al., 2003). Other ontology related approaches used for indexing cultural content include ICONCLASS (<http://www.iconclass.nl>) (van den Berg, 1995) and Art and Architecture Thesaurus (AAT) ([http://www.getty.edu/research/conducting\\_research/vocabularies/aat/](http://www.getty.edu/research/conducting_research/vocabularies/aat/)) (Peterson, 1994), and CIDOC CRM (<http://cidoc.ics.forth.gr/>) (Doerr, 2003).

### **Further work**

Several practical problems were encountered in transforming the database contents into RDF. Even if the XML card is syntactically well-formed, several semantic interpretation problems have to be addressed during the XML to RDF transformation. The values of the features in XML cards may be complicated expressions and come from various data field in the database. For example, value “Christmas calendar, Finland's Scouters' assoc.” is not a term but a complex phrase. The same concept may be referred to with different syntactic expressions (e.g., “Scouters' Christmas calendar”) depending on the cataloger and notational conventions used. Using standard terminology in cataloging would help in solving this problem but in practice this is impossible, and there will be variation in descriptions.

The XML to RDF transformation cannot be fully automated due to problems of homonymy and emergence of new terms and concepts with new collection items. To solve the problem, the cataloging systems should be enhanced with ontology support. Ways of collaboration between museum content providers and portal maintenance people need to be developed in order to develop MuseumFinland from an application into a continuous publication process for the participating museums. For example, protocols for adding, modifying, and retracting RDF cards and ontology resources according to the wishes of the museums need to be developed.

More content analysis work is needed in developing a set of recommendation predicates that would be of most interest to the users. It is possible that their implementation would require changes in the ontologies and better annotated content.

In the near future we plan to extend the collections of the system with paintings and graphics from the Finnish National Gallery. We also plan to incorporate in the system a database from the National Museum describing the most valuable cultural sites in Finland. Our goal is to show how RDF can be used as the basis for making very different kind of contents semantically interoperable.

### **Acknowledgements**

The Espoo City Museum, National Museum, Lahti City Museum, and Finnish National Gallery made their collections available for our research. Our work is funded mainly by the National Technology Agency Tekes, Nokia, TietoEnator, the Espoo City Museum,

the Foundation of the Helsinki University Museum, the National Board of Antiquities, and the Antikvaria Group consisting of some 20 Finnish museums.

## References

- Berners-Lee, T., J. Hendler & O. Lassila (2001). The semantic web. *Scientific American*, 284(5):34–43, May.
- Brickley, D. & R. V. Guha (2000). Resource Description Framework (RDF) Schema Specification 1.0, W3C Candidate Recommendation 2000-03-27, February. <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>.
- Doerr, M. (2003). The CIDOC CRM—An ontological approach to semantic interoperability of metadata. *AI Magazine* 24(3):75–92.
- English, J., M. Hearst, R. Sinha, K. Swearingen & K.-P. Lee (2003). Flexible search and navigation using faceted metadata. Technical report, University of Berkeley, School of Information Management and Systems.
- Fensel, D., J. Hendler, H. Lieberman & W. Wahlster (Eds.) (2002). *Weaving the Semantic Web*. The MIT Press.
- Foskett, D. (1980). Thesaurus. In *Encyclopaedia of Library and Information Science*, Volume 30, 416–462. Marcel Dekker, New York.
- Hearst, M., A. Elliott, J. English, R. Sinha, K. Swearingen & K.-P. Lee (2002). Finding the flow in web site search. *CACM*, 45(9):42–49.
- Hollink, L., A. Th. Schreiber, J. Wielemaker & B.J. Wielinga (2003). Semantic annotations of image collections. In *Proceedings of KCAP'03*, Florida, October.
- Hyvönen, E., S. Kettula, V. Raatikka, S. Saarela & K. Viljanen (2002). Semantic interoperability on the web. Case Finnish Museums Online. In (Hyvönen & Klemettinen), 41–53. <http://www.hiit.fi/publications/>.
- Hyvönen, E., & M. Klemettinen (Eds.) (2002). Towards the semantic web and web services. *Proceedings of the XML Finland 2002 Conference*. HIIT Publications 2002-03. Helsinki Institute for Information Technology (HIIT), Helsinki, Finland. <http://www.hiit.fi/publications/>.
- Hyvönen, E., A. Styrman, and S. Saarela (2002). Ontology-based image retrieval. In (Hyvönen & Klemettinen, 2002), 15–27. <http://www.hiit.fi/publications/>.
- Hyvönen, E., S. Saarela & K. Viljanen (2003a). Ontogator: combining view- and ontology-based search with semantic browsing. In *Proceedings of the XML Finland 2003 Conference*. <http://www.cs.helsinki.fi/u/eahyvone/publications/xmlfinland2003/yomXMLFinland2003.pdf>.
- Hyvönen, E., A. Valo, K. Viljanen & M. Holi (2003b). Publishing semantic web content as semantically linked HTML pages. In *Proceedings of XML Finland 2003 Conference*. [http://www.cs.helsinki.fi/u/eahyvone/publications/xmlfinland2003/swehg\\_article\\_xmlfi2003.pdf](http://www.cs.helsinki.fi/u/eahyvone/publications/xmlfinland2003/swehg_article_xmlfi2003.pdf).
- Hyvönen, E., M. Salminen, S. Kettula & M. Junnila (2004). A Content Creation Process for the Semantic Web. Forthcoming paper.
- Lassila, O. & R. R. Swick (Eds.) (1999). Resource description framework (RDF): Model and syntax specification. Technical report, W3C, February. W3C Recommendation 1999-02-22, <http://www.w3.org/TR/REC-rdf-syntax/>.
- Lee, K.-P., K. Swearingen, K. Li & M. Hearst (2003). Faceted metadata for image search and browsing. *Proceedings of CHI 2003*, April 5-10, Fort Lauderdale, USA. Association for Computing Machinery (ACM), USA.

- Leskinen, R. L. (Ed.) (1997). *Museoalan asiasanasto*. Museovirasto, Helsinki, Finland.
- Pepper, S. (2000). The TAO of Topic Maps. In *Proceedings of XML Europe 2000*, Paris, France. <http://www.ontopia.net/topicmaps/materials/rdf.html>.
- Peterson, T. (1994). Introduction to the Art and Architecture Thesaurus. <http://shiva.pub.getty.edu>.
- Pollitt, A. S. (1998). The key role of classification and indexing in view-based searching. Technical report, University of Huddersfield, UK. <http://www.ifla.org/IV/ifla63/63polst.pdf>.
- Raatikka, V. & E. Hyvönen (2002). Ontology-based semantic metadata validation. In (Hyvönen & Klemettinen, 2002), 28–40, 2002.
- Sihvo, P. (Ed.) (1996). *Kulttuuraineiston luokitus. Outline of cultural materials*. Museovirasto, Helsinki, Finland.
- Sowa, J. (2000). *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Brooks/Cole.
- Stuer, P., R. Meersman & S. De Bruyne (2001). The HyperMuseum theme generator system: Ontology-based internet support for active use of digital museum data for teaching and presentations. In Bearman, D. & J. Trant (Eds.), *Museums and the Web 2001: Selected Papers*. Archives & Museum Informatics. <http://www.archimuse.com/mw2001/papers/stuer/stuer.html>.
- van den Berg, J. (1995). Subject retrieval in pictorial information systems. In *Proceedings of the 18th International Congress of Historical Sciences*, Montreal, Canada, 21–29. <http://www.iconclass.nl/texts/history05.html>.
- Wielemaker, J., A. Th. Schreiber & B. J. Wielinga (2003). Prolog-based infrastructure for RDF: performance and scalability. In *Proceedings ISWC'03*, Florida. Springer-Verlag, Berlin, October.