

W2005 – research panel presentation

Accessing Content Data from Multiple Museum Sources

April 2005

Joan C. Nordbotten, Assoc. professor
Dept. of Information and Media Sciences, University of Bergen
<http://www.ifi.uib.no/staff/joan/>
joan@ifi.uib.no

Problem statement

Museums world-wide are deploying collections (databases) of multimedia data on the Internet for use by researchers, educators and the general public. When these users need thematic information, multiple autonomous sites must be located and individually searched for relevant information that must be extracted and stored for 'off-line' development of an integrated presentation. The problem can be summarized as a need for methods and tools to aid users in locating, accessing, extracting and presenting information from multiple, multimedia-database systems developed and maintained by autonomous museums. Or more simply, the problem is to provide seem-less access to the Web-data of multiple museums.

Connecting the user with an information need to heterogeneous data sources is an increasingly urgent requirement as the wealth of material grows on the Internet (see also Addis et al., 2005). Portals, Google and similar search engines provide indexing and searches of Web-sites consisting of html/text pages and return a list of links to independent sites. However, they do not yet provide search of the content databases supporting Web-sites or even available for user search from individual web-sites. Nor do they provide tools for processing the data from the search result.

The goal of the VED – Virtual Exhibits on Demand – project is to develop tools that can support searches in multiple content databases and return an integrated presentation, i.e. to provide a dynamically created information presentation. Figure 1 shows the assumed environment for the user with an information need that can be serviced from multiple museum content databases.

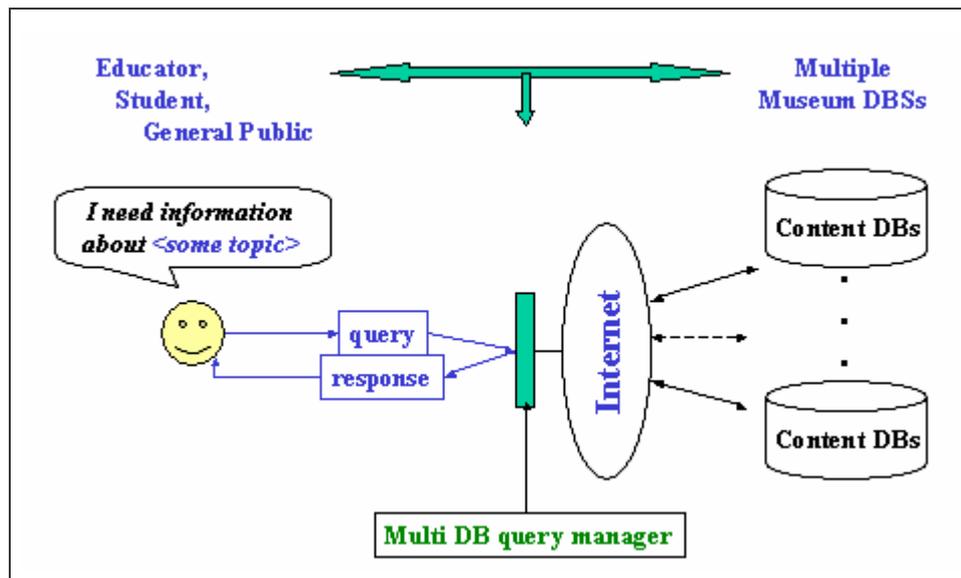


Figure 1. Multi-source support for user queries

Approaches that support access to multiple databases.

There are three general approaches¹ to providing access to multiple databases as illustrated in Figure 2:

- 1) **Centralized DB** to which data is copied from the source DBs for distribution to users. This is the goal architecture/view for DB integration. I.e. the ideal view from the user perspective.
- 2) **Union catalog**, in which source systems maintain metadata and indexes in a common format, f.ex. Dublin Core, and a union of this metadata is maintained at a central location. This architecture gives the users very efficient search and reasonable retrieval times, while content management remains with the local database owner. A thesaurus/taxonomy may be added to facilitate translation between query terms and the stored data.
- 3) **Integrated (global) schema**, in which component DB systems maintain DBs with heterogeneous structures for their metadata. A global schema represents a single integrated description (or global metadata) of the component DB systems. It also includes a thesaurus/ontology² relating the total metadata for the underlying systems, as well as the translation routines necessary for resolution of both the underlying DB structures and index values.

Each of these approaches support metadata/keyword search of typically manually produced text descriptions. A typical query could be to *find (information about the) whales of the north Atlantic*. In this case, the object type and location are given as search criteria to the system.

¹ It is common to utilize a combination of techniques when implementing a specific system/application.

² The term "ontology" has various, overlapping definitions in the ICT field, none of which adhere to the classical definition from philosophy.

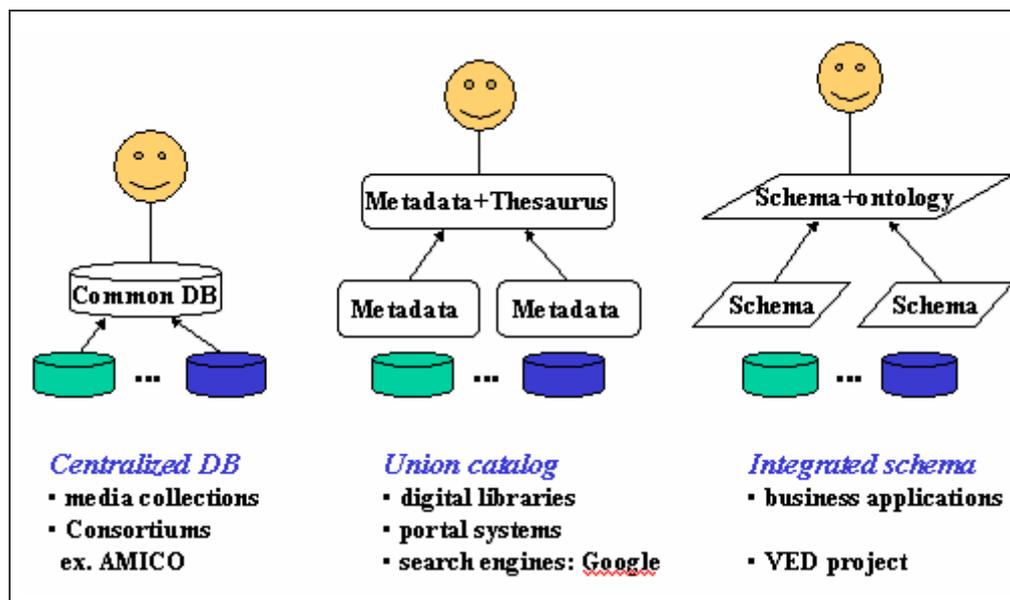


Figure 2. Database integration approaches

Each of these strategies has strengths and weaknesses:

- Though the *centralized DB* provides the best response times it can become out of date with respect to the content of its member databases. It also requires maintenance in addition to that given the local/member source systems.
- The *union catalog* supports user queries and addition of new source systems, but requires that the local systems adhere to the metadata model chosen for its catalog. Less data needs to be copied to the central systems, thus reducing total system maintenance.
- The *integrated schema* approach allows the most architectural freedom at the local level, but is difficult to achieve due to the semantic heterogeneity within the metadata system.

A research approach

The aim of the “*Virtual Exhibits on Demand*” project – VED is to *integrate and develop IT methods, and techniques to support development of 'on-demand' virtual exhibits from multiple multimedia DB sources for educational applications* (Nordbotten, 2001).

We have developed a set of databases with core data from and describing Bergen Museum’s whale collection (Hufthammer and Nordbotten, 2003). The multi-DB test-bed consists of:

- A traditional, relational databases containing catalog/metadata,
- 3 (4) variations of image databases focused on marine animals,
- 2 text/html databases with articles concerning marine animals and
- A video database containing video clips from a research excursion in the North Atlantic.

The image databases will be combined and extended into a test-bed for further research in image retrieval.

There are currently 6 working prototypes demonstrating elements of DB integration, text and visual query formulation and result presentation and processing, which are accessible from the project web-site at <http://nordbotten.ifi.uib.no/VirtualMuseum/VMwebSite/VEDweb-site.htm>. The prototypes

Accessing content data

explore different layouts for presentation of data from multiple sources as illustrated in Figure 3, from <http://nordbotten.ifi.uib.no/VirtualMuseum/Prototypes/Osdal2>. In this prototype, query result data is presented in 4 quadrants: the upper left for text, upper right for image, lower left for fact data (from a relational DB) and the lower right for user interaction and external/search engine lists. The user can step through the results in each quadrant and save relevant results by clicking the '+' sign. Note that individual elements are appended with their source to provide an element of copy-write protection and enable the user to evaluate the validity of the data. An editing function for placement and further integration of the results is accessible from the LR quadrant. Details of this system can be found in Osdal (2004).

The screenshot displays a web interface with a header for 'Bergen Museum' and 'UNIVERSITET I BERGEN'. The navigation menu includes 'Search / Main', 'Search results', 'Exhibits / Account', and 'Log out'. The user is logged in as 'JOAN' with the email 'JOAN@INFOMEDIA.UIB.NO'. The interface is divided into four quadrants:

- Text results:** A text block describing killer whales as effective and awe-inspiring predators, living in a matriarchal society. It includes a '+' sign for interaction.
- Image results:** A photograph of a killer whale leaping from the water. Below the image is a copyright notice for GORM KALLESTAD and a '+' sign. A description and title are provided: 'Description: The picture shows a killer whale jumping up from the water.' and 'Title: Killer whale'.
- Fact results:** A section with a '+' sign and a list of numbers (1, 2, 3, 4, 5) for navigation.
- Search results from Google:** A section with a '+' sign and a list of numbers (1, 2, 3, 4, 5) for navigation.

Figure 3. Quadrant presentation of multisource data

Acknowledgements:

This presentation has been based on work being done on these problems at the Department of Information and Media Science at the University of Bergen in Norway within a project titled: *Virtual Exhibits on Demand*, funded by the Norwegian research council under grant #148827/530.

References:

Addis,M., Martinez,K., Lewis,P., Stevenson,J. and Giorgini,F. (2005) *New Ways to Search, Navigate and Use Multimedia Museum Collections over the Web* <http://www.archimuse.com/mw2005/papers/addis/addis.html> Int'l conference on Museums and the Web, Vancouver, BC. April 2005.

Nordbotten,J. (2001) *Virtual exhibits: Theory, methods, and tools for development of virtual exhibits on demand.* http://www.ifi.uib.no/staff/joan/VM-project/project_description.htm

Osdal,Ø. (Nov. 2004) *Developing the SquareVed Prototype* . TR: IFI/VED-ØO 11-04. Dept. of Information and Media Sciences, University of Bergen

Web-site and Prototypes:

<http://nordbotten.ifi.uib.no/VirtualMuseum/VMwebSite/VEDweb-site.htm>

<http://nordbotten.ifi.uib.no/VirtualMuseum/Prototypes/Osdal2>

<http://nordbotten.ifi.uib.no/VirtualMuseum/Prototypes/Nesland>