From Information Dissemination to Information Gathering
Using virtual exhibits and content databases in e-learning centers[1]

Joan C. Nordbotten
Department of Information and Media Science
University of Bergen
P.B.7800, N5020 Bergen, Norway

http://www.ifi.uib.no/staff/joan
Joan.nordbotten@infomedia.uib.no

March 2005

ABSTRACT
Concurrently, large multimedia databases, termed *content databases*, are being created to store and manage digital representations of a museum's physical collections, such as scanned images, text documents and videos. These databases can also provide valuable information and data for use in the development of e-learning centers[i] for tasks ranging from presentation of information about the museum's educational resources to providing a full interactive learning experience for students and casual information seekers. There are at least three ways an e-learning center can support learning: first by providing information on a given set of topics, second by providing educational activities to reinforce learning, and third by supporting information gathering. In this chapter, we will present and discuss how different e-learning center architectures support these different forms of learning.

Keywords: information systems, databases, user interface, museum applications, educational IS, e-learning systems, Web applications, Virtual exhibits

[1] : in *E-Learning and Virtual Science Centers*, edited by Tan & R. Subramaniam. Idea Group Inc. 2005. as chapter 11, pg.220-242.

Background Assumptions and Definitions

Figure 1, shows the 3 basic components of an e-learning center, the user/learner, the learning site and the (optional) content database.

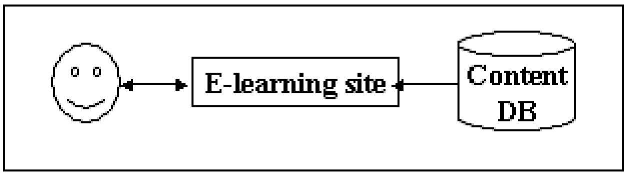


Figure 1: e-learning center components

The users of a museum e-learning center include anyone who 'discovers' the Web-site. The users most commonly thought of during design of an e-learning center include: teachers – who are expected to guide their students in their use of the site – and students – who are expected to use the site as a learning tool. However, since the site will also be available to the general public, it is important to consider the needs of interested information seekers. We can assume that these users are interested in the site topic (most likely they found the site through a search engine), but will not have other guidance in the use of the site than that which is given on the site. In order to assure effectiveness, it is important that the e-learning site is designed for these multiple user-types. In the following, both the student and interested information seeker will be termed a *learner*.

A database can be generally defined as "*a logically coherent collection of related data, representing some aspect of the real world, that is designed, built, and populated for some purpose.*" (Nordbotten, 2000-2004). In a museum[ii] context, a *content database* contains digital representations – scanned images, text documents, videos, etc. – of physical objects in the museum's collections. In the following, we assume that a content database contains such representations and that the e-learning site can utilize them.

An *e-learning site* consists of a set of Web pages that present material intended to educate its users, the learners. A site's architecture can be classified according to the degree of user control over the material that is presented. A basic *virtual exhibit* presents topic material as a hyper-linked story that the viewer can navigate, much as one finds in traditional museum exhibits. An *interactive* site also presents its topic as a hyper-linked story but adds a variety of interactive activities aimed at increasing the learning effect for the learner. An interactive e-learning site is based on the idea of a "hands-on" physical exhibit, in which experiments can be run and questions answered. Typical for both of these architectures is that they are museum controlled in the sense that the user is presented with prepared material that is to be explored within a pre-defined site.

Increasingly, Internet users, both students and casual information seekers, search for information about a specific topic of individual interest. The topics of these searches vary widely and many, though well within the interest areas of the museum community, may not be well covered by existing virtual exhibits. Alternatively, if the topic is popular and/or part of a common interest area, it is likely that relevant information can be found at multiple sources. For example, someone interested in *space exploration* can find relevant information from numerous sources[iii] including science museums, universities, and government research centers. A challenge for the museum community would be to provide multiple-site responses to a user request for information, i.e. to support construction of a virtual 'exhibit' on-demand (Nordbotten, 2001).

Current pedagogical practice emphasizes activating the learner and learning by doing, thus encouraging active exploration of available information and gathering of information to address his/her own interest area. We consider 3 learning models in order of increasing user initiative, which can be supported by different e-learning center architectures:

1) *Information browsing*, in which the learner explores material provided by a museum through a pre-defined virtual exhibit.
2) *Problem solving,* in which the learner interactively works through pre-defined, interactive activities, and
3) *Information gathering*, in which the user searches for information on a self-defined topic, without concern for the source of the response set.

In the following, we will discuss e-learning site architectures and the information technology tools needed to support each of the learning models listed above. Each section will present the primary audience and anticipated use, discuss site structures through the presentation of examples, and finally present the strengths and the principal problems of the architecture type as a learning environment.

Note that it is beyond the scope of this chapter to give detailed information on the implementation tools mentioned. This information can be found in numerous books on web-site implementation, such as those by Lane & Williams (2002), Lowe & Prince, (2003) and Murray & Everett-Church (2003).

Information Dissemination - the Basic Virtual Exhibit

The basic virtual exhibit is one of the most common site architectures used for museums presentations on the Web today and the one most frequently used for e-learning centers. The exhibit presents a topic as a hyper-linked story, patterned on that of a physical museum exhibit or textbook. As an e-learning site, the learner can have more information available and greater choice in its presentation sequence than that which is possible in a physical exhibit.

When using a virtual exhibit as an e-learning environment, the learner is expected to be a student or an interested information gatherer. He/she is assumed to be looking for information on the topic of the exhibit and willing to view, explore and read the material presented. The learner browses through the exhibit by following established links. Ideally, both a navigation bar and site map are included so that the learner can change paths to explore alternative sections of the exhibit.

The basic virtual exhibit can be an excellent method for information dissemination, since the museum/center determines the material presented and guarantees its correctness and educational quality. A potential problem is that the exhibit is *fixed* in the sense that the material and the links used to determine the presentation sequence are pre-defined, thus reducing the possibility for the user to follow his/her own interests.

Architecture and implementation
The typical exhibit architecture consists of a core *home* page with information about the exhibit and links to separate sub-topic sections. Topic units can be developed independently and then attached to the core forming a *starburst* or tree structure. The embedded links enable the choice of topic to be viewed, selection of additional in-depth explanatory material, and/or connection to 'outside' (of the parent museum) information sources. Since the "lost in space" phenomenon is a frequent problem for viewers of even small sites, a navigation bar and/or site map should be included so that the viewer can re-orient him/herself (Conklin, 1987; Shneiderman, 1998).

The design of the basic virtual exhibit is analogous to the design of a traditional physical exhibit and is not un-like the structure of a textbook. Exhibit objects are selected, scanned, commented using textual and/or audio-visual material, and finally organized using links for presentation. For large exhibits, material can/should be stored in a content database to facilitate site maintenance. Smaller exhibits may not make use of the museum's content database. Rather, the site will be built using scanned objects that have been stored on separate file areas and given url's that can be directly referenced using the html <img> tag. The current primary implementation language is HTML, frequently enhanced with FLASH™ animation, though other implementation tools may also be used.

There are two main architectures for virtual exhibits:
1. Single-tiered exhibits in which objects and information items are presented completely before continuing with the exhibit presentation, and
2. Multi-tiered exhibits that include links within the main story that allow the viewer to select more detail about individual objects from an underlying information layer, possibly retrieved from the museum's content database.

Single-tiered exhibits consist of a, possibly multi-threaded, story in which most links select the next part of the story. Depending on the topic and available object media, single-tiered exhibits generally use one of two metaphors:
- The traditional *museum exhibit*, which relies on presentation of visual/ image material with accompanying text, or
- The *textbook*, which presents textual material with some illustrations.

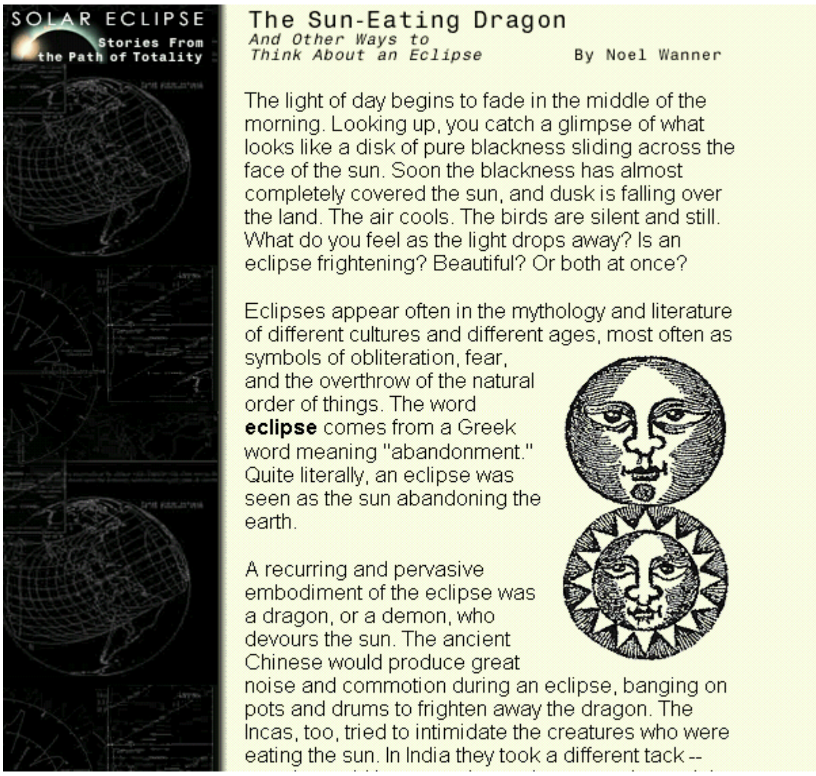
An example of use of the traditional museum exhibit metaphor can be seen in *The Write Experience – preparing for Kitty Hawk* at http://www.wrightexperience.com/. This exhibit first presents the user with brief introductions to a series of videos documenting various activities from the planning, construction and execution of the centennial celebration of the first flight by the Write brothers. Figure 2 shows a screen shot from the exhibit in which a photo is accompanied by a short text that gives the context and introduces the video from which the image was taken. In this example, each video clip contains the compete presentation of its sub-topic.

Figure 2: Museum exhibit metaphor
From The Wright Experience™ ©2004

The San Francisco Exploratorium, at www.exploratorium.org/, has used a *textbook metaphor* for their presentation of a set of stories and myths about solar eclipses in an exhibit titled: *Solar Eclipse* at http://www.exploratorium.org/eclipse/index.html. Figure 3, gives a screen shot of part of a typical page from this exhibit. In the exhibit, each story is presented as a single-tiered exhibit on a series of linked pages.

The single-tiered exhibit architecture provides an excellent environment for information dissemination. For e-learning sites, this helps assure pedagogical and quality control of the story/message being presented and supports focusing of the presentation to a particular age and/or interest group. General familiarity with the presentation metaphor makes these sites 'easy' to design and use. An added advantage is ease of implementation and deployment since well-known tools and technology can be used.

A common problem/mistake made in the implementation of the virtual exhibit is the inclusion of too much explanatory text for each visual object, such as one would expect to find in a traditional textbook. The Web is a very visual medium and studies of casual viewers indicate that they spend less than 30 seconds on a page (Nordbotten, 2000; Nordbotten & Nordbotten, 2002; Shneiderman, 1989; Yamada, 1995). This is insufficient time to read lengthy text sections.



SOLAR ECLIPSE
Stories From the Path of Totality

The Sun-Eating Dragon
And Other Ways to Think About an Eclipse
By Noel Wanner

The light of day begins to fade in the middle of the morning. Looking up, you catch a glimpse of what looks like a disk of pure blackness sliding across the face of the sun. Soon the blackness has almost completely covered the sun, and dusk is falling over the land. The air cools. The birds are silent and still. What do you feel as the light drops away? Is an eclipse frightening? Beautiful? Or both at once?

Eclipses appear often in the mythology and literature of different cultures and different ages, most often as symbols of obliteration, fear, and the overthrow of the natural order of things. The word **eclipse** comes from a Greek word meaning "abandonment." Quite literally, an eclipse was seen as the sun abandoning the earth.

A recurring and pervasive embodiment of the eclipse was a dragon, or a demon, who devours the sun. The ancient Chinese would produce great noise and commotion during an eclipse, banging on pots and drums to frighten away the dragon. The Incas, too, tried to intimidate the creatures who were eating the sun. In India they took a different tack --

Figure 3: Single-tier - Textbook metaphor
Reproduced with permission (c) Exploratorium, www.exploratorium.edu

Multi-tiered exhibits aim to reduce the information overload that can be triggered by text quantity in the primary exhibit, by dividing the exhibit into multiple tiers, a primary tier for an image-based principle narrative with underlying layer(s) of detail that the interested/curious user can select from various points in the main exhibit. This support material can be accessed using different strategies, including through the use of:

- Active zones in the display that are used as links to pages with object enlargement or alternative media presentations.
- Zooming or a "magnifying glass" to highlight detail and/or provide more information.

The Natural history (Natuurhistorisch) Museum in Maastricht has used the multi-tiered approach in their *Virtual tour,* at http://www.nhmmaastricht.nl/engels/index.html, of the museum. The tour consists of a series of images that give the illusion of entering a room or passing into a further section of the current room. Wandering through the exhibit is accomplished by clicking on the highlighted rectangle, shown in the screen shot given in Figure 4a. In this tour, animated object names are used as links to a detail presentation, as illustrated in Figure 4b. The detail page appears 'on top' of the primary page thus supporting orientation within the site.

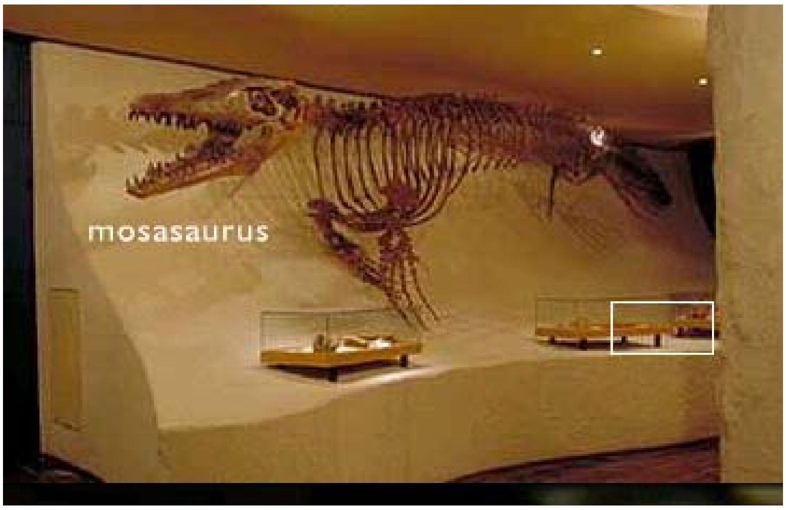


Figure 4a: Hot spots for detail and navigation

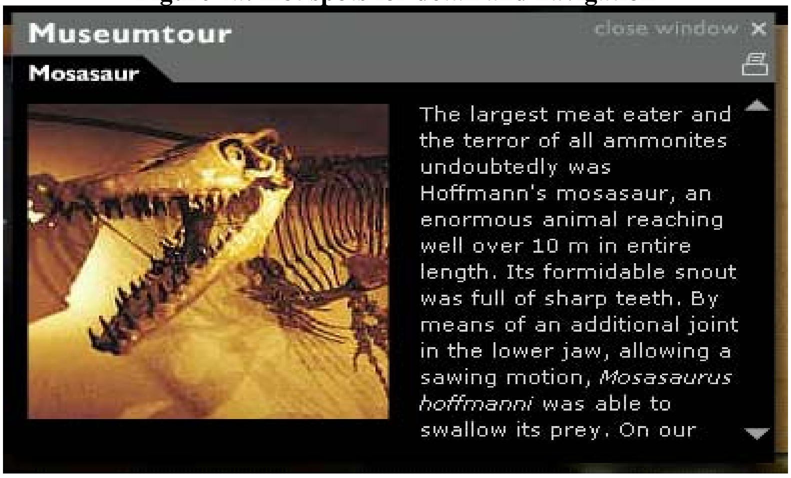


Figure 4b: detail presentation selected from Fig.4a
Figure 4: A multi-tiered presentation of the Mosasaur dinosaur
Reproduced with permission © Natural History Museum at Maastricht

Supportive detail can also be retrieved using "*magnifying-glass*" technology, where a window is passed over an image and an enlarged version of the window is shown elsewhere on the user's screen. This technique is especially effective when the original image is content rich, as is the case in the screen shot from the *Ology/Marine biology* exhibit at http://www.ology.amnh.org/marinebiology/workthesystem/mangrove.html, shown in Figure 5. In this case the primary image contains many creatures that are barely visible. The magnified version also gives the common and scientific name for the central object within the magnifying glass.

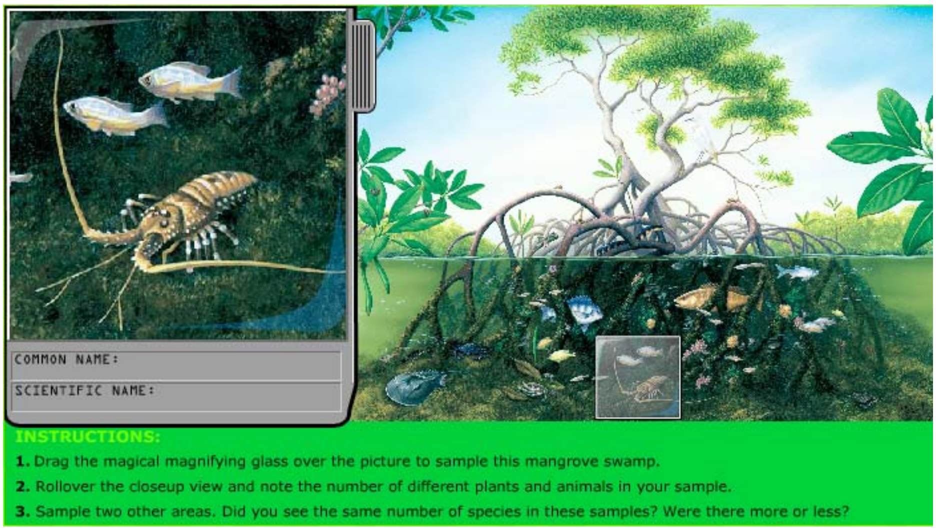


Figure 5: Spiny lobster under a Mangrove tree
From the American Natural History Museum©

Potentials with Multi-tiered exhibits

As learning tools, multi-tiered virtual exhibits transfer some control to the user so that he/she can choose to select added detail for those objects that are most interesting for him/her, thus helping to tailor the information presented to the information requirement of the viewer/learner. Multi-tiered exhibits support inclusion of a wealth of information to that can support interested information seekers, while avoiding some of the information overload problems for the learner. This is a useful strategy for addressing multiple audiences from school to university students to interested, casual information seekers.

The main problems of multi-tiered exhibits include the temptation to make increasing use of multi-media, such as Flash animations and 'real' video, which require high bandwidth for seamless presentation systems. Though presentation software is often free, they still may not be available to all users. A growing problem is that as the quantity of secondary support material grows, users can get lost within the site, in the sense that they loose the flow of the overall story.

Both single and multi-tiered virtual exhibits can provide very good support for the learner and information gatherer needing an introduction and/or overview to a topic area as well as supporting those who wish more depth of information. Content and quality of the information provided are assured and maintained by the information provider. Increasingly, these sites are well advertised and easy to locate. And finally, the technology required to design and launch interesting and attractive exhibits, as well as to download them is readily available to both the provider and information user.

As learning-centers, virtual exhibits function well as a source of information but frequently do not include user activity aimed at reinforcing knowledge about the exhibit topic. Though suggestions for associated user activity may be included in a teacher area of the site, these are often off-line activities and may be difficult to find for the learner, particularly the casual information seeker.

At a certain level, fixed exhibits are a passive information source, well suited to information dissemination, but where user interaction is limited to selection of pre-determined topic presentations. There is little support for reinforcement of the information content, other than to repeat a sub-topic presentation, often in its entirety. A technical problem is that the information is commonly kept on a growing number of Web pages/presentations that may overlap and will become difficult to manage – update, extend and reuse.

Interactive Exhibits for Active Learning

The current leading pedagogical paradigm is "learning-by-doing" in which the student builds knowledge and expertise through active participation in development projects. Techniques to support learner activities abound and range from paper-and-pencil workbooks, through hands-on experimentation, to Web games and simulations.

An *interactive exhibit* is one that can adapt its presentations to user input and thereby support active learning through such interactive activities as *games* or *experiments* that are based on the thematic material of the parent exhibit. Interactive exhibits are similar in concept to the "hands-on" exhibits found in museums.

Architecture and implementation
Dynamic Web programming tools[iv] provide the implementation environment for construction of dynamic exhibits. Figure 6 below, gives an overview of the system components necessary to support these exhibits. When the user activates an interactive exhibit, his/her browser sends a request and receives the exhibit through a web server from the application server. The browser presents the exhibit with its audio-visual material and interactive scripts. During use, the client-side application controls the user interaction and for some exhibits the whole presentation, while the host machine's application server controls such server-side activities as access to a content database.

Exhibit scripts can be developed in Flash™, Java or some other programming language. Communication with the user is through his/her browser. When a client-side application is used, it is executed on the client machine, thus saving interaction time between the client and host machines at the cost of the initial load time. However, when material stored in a content database is required, a request must be sent from the client system to the host system's application server, which converts the request into one or more DB queries to the available databases, marked C_DB in figure 6.

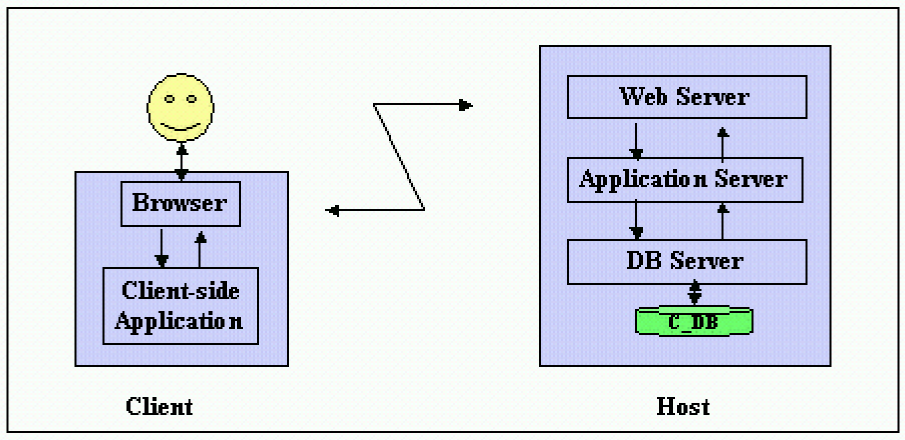


Figure 6: Interactive exhibit support

Games

Perhaps the easiest learning tool to add to an e-learning site is inclusion of a quiz on the material presented in the virtual exhibit. An interesting example can be found on the Ology-site at http://www.ology.amnh.org/index.html of the American Museum of Natural History. Here both facts and quiz questions are placed on cards that can be 'collected' and saved on the Museum site, i.e. in a learner DB for future reference.

With some added effort from the developer, the quiz can be presented in the form of a game, encouraging learners of all ages to use the information acquired from the exhibit so that it can become useful knowledge. Typically the game is formed as a set of questions about topics that have been presented in the virtual exhibit. The system evaluates the answer(s) and forms a response commonly consisting of point assignment, a reinforcement answer, and a new question. This cycle of questions and answers repeats until the game completes. If questions are selected randomly, the game can be replayed with new questions repeatedly. The success of a game, as a learning tool, lies in its ability to engage the learner(s).

A good example (for all age groups) 'teaches' the discovery and use of rubber for making balls through a Mesoamerican ball game, at http://www.ballgame.org/[v]. The Mesoamericans were the first to use rubber to make balls and to use a team ballgame for conflict resolution. This game uses animation and sound effects, as well as period costumes and background. The ball is first 'batted' between the teams in a manner similar to volley ball. A question appears with 3 answer options, 2 of which are shown in Figure 7. Answering correctly gives the users' team a point. Sound effects clearly signal the quality of the answer, which is enhanced with a textual explanation. (To play, go to http://www.ballgame.org/sub_section.asp?section=3&sub_section=2)

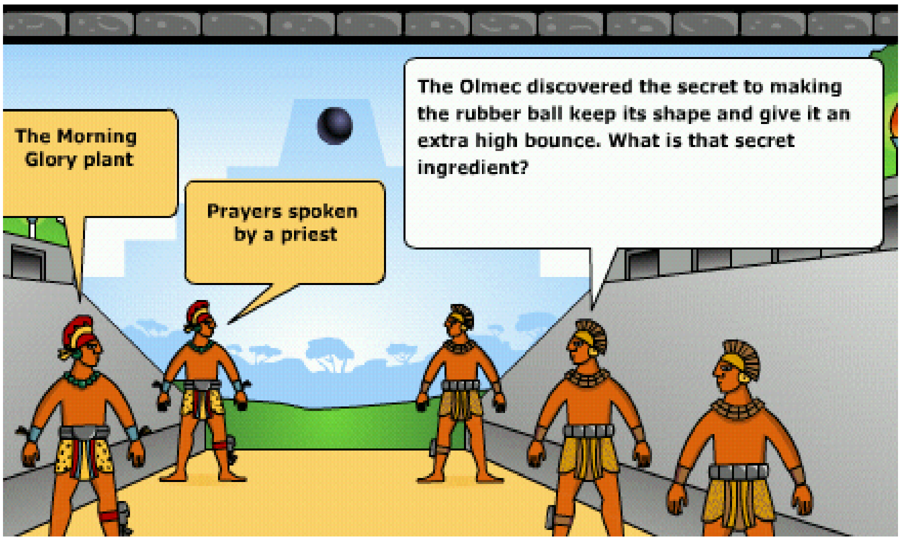


Figure 7: Quiz in a game setting
Example from the Mint Museum ©

All of the material in the above ballgame has been packaged in Flash™ and sent to the client as a client-side application. The trade-off is between download time (long) or a high frequency of interaction between the client and server machines for the question and answer sets.

Virtual experiments
Games can also be used to frame simulations that allow the user to explore some physical phenomena that might otherwise be inaccessible. The idea is to give the user an understanding through experimentation. An example is the 'particle accelerator simulation developed by the education department of European Organization for Nuclear Research, CERN. The goal of this simulation is to get the particle to accelerate by passing it through magnetic fields that are 'charged' by flipping the batteries. Figure 8 shows a screen shot[vi] as the positively charged particle is entering the 2nd negatively charged loop from the right. Note: A new version of this simulation is at http://public.web.cern.ch/public/Content/Chapters/Education/OnlineResources/LHCGame/LHCGame.html.

A simulation game can be packaged as a client-side application, as the one presented, or it can utilize data that has been stored in the museum databases. Data generated from the simulation can be saved onto the user's machine for local use. In more extensive simulations, both the status for the simulation and data generated 'to-date' can be stored on the museum site. This would support long term and/or multiple player games or experiments.

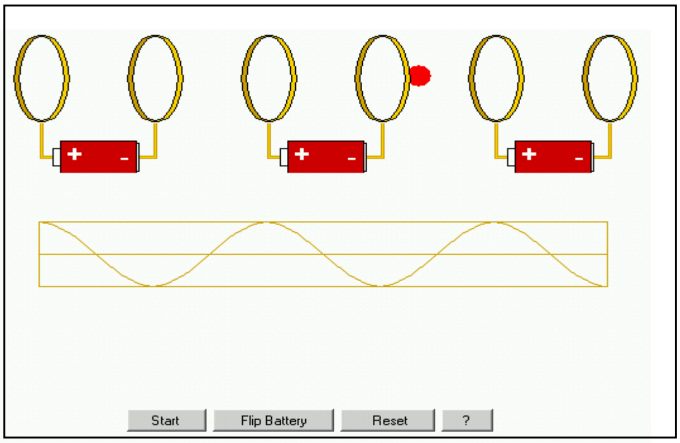


Figure 8: Particle acceleration - simulation game
Reproduced with permission © CERN

Virtual projects

Virtual projects are similar in concept to the physical "hands-on" activity that allows the museum visitor to develop/build something that can be taken home. In the virtual environment, the user is encouraged to solve a problem by using/collecting objects from the museum's content database(s). Frequently, support is provided for adding new objects, which could include notes taken and/or experiences, created during the project. The resulting material can then be stored on the museum site for future use and discussion, for example in the classroom or with on-line friends or colleagues.

Summary

Interactive exhibits, when used as part of e-learning sites, can enhance learning by encouraging the learner and interested information seeker to become engaged in interactive activities that have been selected by his/her own interest, thereby reinforcing the information acquired from the exhibit. In providing interactive activities, the museum /science center can guarantee the correctness of the material presented, as well as its relevance to the theme of the parent exhibit. The technology and expertise for development of interactive exhibits are available and ability to access and use this type of site is increasing, in part through the introduction of IT into the school systems.

A problem with interactive exhibits is that it can be difficult to adjust the system questions, activities, and responses to the learner's level of expertise, potentially making the interaction too difficult or too trivial. To alleviate this, professional educators should be involved in the design of these sites.

While both the basic and interactive virtual exhibits can provide good support for learning, they are still fixed sites, in the sense that the learner is 'bound' to the material provided in a single exhibit and to the sequences that have been implemented for topic presentation. The learner is invited to explore and use the site, but is only infrequently given the possibility to search for particular information on the site or in the museum's content databases. This limits the possibility for the learner to gather or focus on information about a topic of personal interest (or an assignment given by a teacher).

Information Gathering
Objective: "On-demand" exhibits

Increasingly, students and general Internet users search for information about current and/or personal interests by using one or more of the Internet search engines, much as one would use an encyclopedia. Often, relevant information exists in one or several museum exhibits and/or content databases. The problem is to find and retrieve it.

To locate information on the Internet, the information seeker must:

1. Choose one or more search engines.
2. Formulate and submit a query statement of his/her information need.
3. Select potentially relevant sites from a large result list
4. Evaluate selected sites for 'real' relevance
5. Review and extract (copy) information from the different sources, and finally
6. Compile and integrate the information.

This can be a daunting task unless a real need to learn exists. Happily, a number of consortiums and research programs have been formed to find ways to ease information retrieval from the multitude of Internet sites.

Preparing a virtual exhibit for location and retrieval by the Internet search engines can be accomplished by including html <meta> tags in the site header that specify both the thematic content description and a set of search keywords. These will be found by the Internet search engines and used for indexing and retrieval. The advantage is that the learner/information gatherer will then be more likely to locate museum exhibit material relevant to his/her information need.

Internet search engines use 'crawlers' that retrieve text from html pages to index a site. Currently, crawlers are not capable of extracting text or features from Flash™ or visual/audio exhibits for indexing. In addition, the crawlers cannot access the indexes of content databases. The result is that search engines are unable to locate video-based sites or information from content databases. The current work-a-round is to construct an html 'envelop' containing a good description of the multimedia presentation to be used for indexing the site. In addition, a specific search facility for the database must be provided.

The above strategy only addresses the first 2 steps on the above search list. Accessing content databases through their home pages and their specific database search techniques still puts a high burden on the information gatherer in the form of multiple site and DB accesses with follow-up manual/local information consolidation. It must be a goal for the information provider community, in this case the museum community, to provide seamless interfaces to their multiple information resources.

Web databases – Architecture and implementation

A Web database is simply one that is accessible from the Internet. In a museum context, the content database must be assumed to contain multiple media objects, ranging from text documents through scanned images of physical objects to full multimedia presentations. Ideally, one would want to support user access to the media objects through a description of the type of information required without concern for the media type (text, scanned image, streamed media) that was used to represent the information.

Figure 9 presents an extension of the architectural environment given in Figure 6 that includes the components needed for multiple multimedia database access. The figure can be explained by following a request for information about a particular topic, for example *North Atlantic whales*. (A prototype site illustrating this technology, at http://nordbotten.ifi.uib.no/VirtualMuseum/Prototypes/Osdal2/, has been developed in the Virtual exhibits project (2004).)

A request is initiated when a user submits a query containing the 3 keywords, in the same way as using a Google™ search. The user's Browser sends the request to the Web server for processing by the Application server, which translates the input into one or more DB queries that are sent to relevant DB servers. The DB servers search for and retrieve relevant data, for example multiple texts and videos of (some of) the 18 species of North Atlantic whales from each DB. Each DB server returns relevant content data to the Application server, which packages it for return by the Web server and display by the Browser to the user. If the information includes multimedia (Flash, audio, video, …) presentations, appropriate media player(s) must be retrieved as part of the client-side application for presentation of the media data.

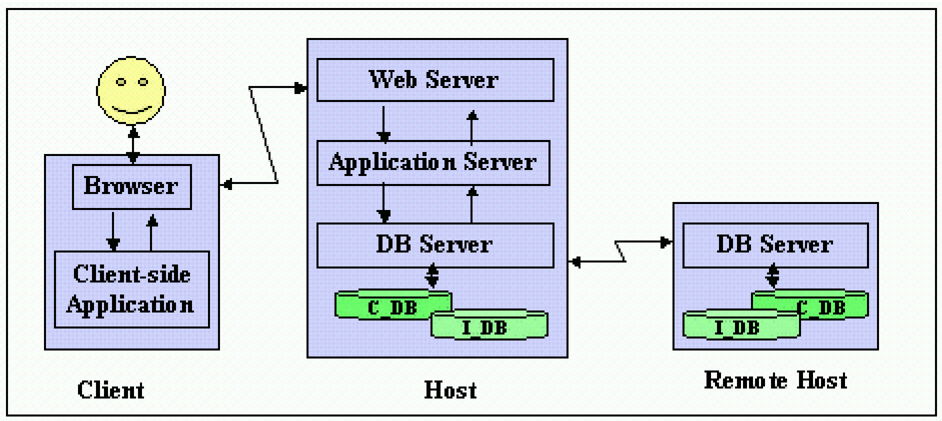


Figure 9: Accessing Web databases

The DB server uses information from the DB indexes, labeled I_DB in the figure, to locate the actual data within the content database, labeled C_DB. If the search is to be preformed on multiple databases, then the indexes for all databases within the scope of the query need to be searched in order to locate those content databases that contain data relevant to the user query.

The data within an index database is derived from two primary sources:

- The metadata used to describe each object upon entry into the database and
- The terms and features in the object itself.

Table 1 gives an example of some of the index data that could be assigned to the story about Eclipses of the sun, illustrated in Figure 3. Note that the metadata elements used in this example have been selected from the 15 core elements in the Dublin Core Standard (Hillmann, 2003).

DC Metadata element	***value***	***Comment***
Title	The Sun-Eating Dragon and other ways to think about an eclipse	
Creator	Noel Wanner	
Publisher	Exploratorium	
Identifier (location)	http://www.exploratorium.org/eclipse/dragon.html	
Source	Solar Eclipse http://www.exploratorium.org/eclipse/index.html	
Description	Myths and stories from Greek, Chinese and Indian cultures describing eclipses of the sun	Given in free/full text
Subject (keywords)	Eclipse, sun, dragon, myth, mythology, ….	From a controlled vocabulary or formal classification scheme
…		
Object features	***Location***	
Abandon	§2	These search terms may be weighted according to descriptive importance
Dragon	Title, §3, …	
Eclipse	Sub-title, Source, §2, 3, …	
Sun	Title, §1, 2, 3, …	
Sun-Eating	Title	
…		
Earth image	Lower right quadrant (LRQ) above Sun image	As seen in Fig.2
Sun image	LRQ, below Earth image	
…		

Table 1: example Index data for a specific text object

Object features are selected from the object itself. For text objects, the feature set consists of the most descriptive terms contained in the document. For visual objects – images, video – automatically extracted features are currently limited to color and texture distribution as well as the shape of primary objects in the image. Semantic description of image objects is still a predominantly manual task (Hove, 2004).

There are at least three problems in using the above index to search for multimedia data:

1. The search terms given by the user in his/her request for information may not match exactly the index terms used to describe the DB content.
2. The terms in the controlled vocabulary may not match exactly the terms used by the creator of the text objects or the descriptive text for image objects.
3. The DB descriptive terms used for one content database are unlikely to match exactly those used for other content databases covering the same topic.

Therefore, in order to locate the most relevant information, some form of semantic matching must be made between the terms that the searcher uses to characterize his/her information requirement and the terms used by the cataloger to index the items in the database as well as the terms used between the cataloger and the creator of the media objects in the database. Information Retrieval Systems have long used thesauri as an aid for interpretation of user queries, thus addressing the 1st of the above problems.

Current best practice proposals from research and the Semantic Web community (Sowa, 2000) advise creation of a set of *domain ontologies* that all content database providers use to describe the objects in their database. This addresses the 3rd problem in the above list by standardizing the values chosen for the metadata elements. However, as seen in the example in Table 1 above, the metadata values may not match completely the index terms selected from the document, the 2nd problem in the above list.

Optimally, the content of the ontology must combine the information in the hierarchic structure and detail of a domain taxonomy, with the term relationships and descriptions in a thesaurus, and with the historical development of the common natural language terminology that information searchers and object creators can be expected use. This is a tall order that has not yet been fulfilled, even though many years of research and development effort have been addressed to the synonym resolution problem in the context of both database integration (Elmagarmid, et al., 1999) and information retrieval from both text and image databases (Baeza-Yates & Ribeiro-Neto, 1999).

Current practice for Information Retrieval from content databases

A recent study of 100 museum Web sites, nominated for awards at the 2003 Museum and Web conference (http://www.archimuse.com/mw2003/) found that only 37 had included any form for database search facility on their sites (Peacock, 2004). Of these, 20 sites supported only keyword searches, while 17 provided some supplementary form for search support. The search support included:

- Keyword search, possibly with keyword selection from a given list of terms
- Browsing/selection from an index of terms or set of thumbnail images
- Selection via a category structure
- Linking to related and/or popular items.

Keyword search is known to be difficult for the information gatherer. Without a term list to choose from, the user is faced with an empty box into which search terms need to be entered. This assumes that the user has rather detailed knowledge about the database content and structure, knows which topics and themes are included in the database as well as the index terms used to describe the data. Often a labeled multiple-box form is given to the user who is then expected to be able to submit a title, creator name(s), and dates in addition to one or more descriptive keywords. Adding drop-down term selection lists can greatly ease the selection of appropriate key words and fill-in options.

Searching an image database can be eased by the presentation of an index of thumbnail images, as shown in Figure 10 below. Here, each image is a link to an enlargement and possible completion of the thumbnail image with descriptive data about the object presented.

Figure 10: a multi DB index – hand constructed
Reproduced with permission © The National Museum of Denmark

A problem with the *thumbnail index* is that the number of elements that can be presented (and thus indexed) are limited by the screen area. This makes it necessary to provide a *category index* for the collection. The user must first select a category for his/her area of interest and can then browse the thumbnails picking items of interest.

The example in figure 10 shows the index for the "mathematical chamber" of the Kings Kunstkammer, at http://www.kunstkammer.dk/MathematischeGB/gemach_mathematischeGB.shtml, which held examples of instruments and inventions. There are actually 9 chambers in the Kings Kunstkammer, at http://www.kunstkammer.dk, each with its own thumbnail index. Even so, it is a problem to display all of the 250 items in this relatively small content database. One way to expand the index scope would be to implement it as a hierarchy of related items or to link to 'hidden' (not in the index) items. Items related by theme, materials, time, creator or any combinations of perspectives can be displayed with the selected item. An example can be found at the virtual Walker Art Center at http://collections.walkerart.org/item/object/7440.

Providing access to multiple data collections
Often information relevant for a user request can be located in the content databases of multiple museums. Given the widespread use of Google, the user may be unaware of (and basically uninterested in) the physical locations of the information. Creating seamless interfaces to multiple information sources is a current, highly active research and development area. There are a number of approaches:

1. Hand crafting a 'traditional' exhibit of objects from multiple sources.
2. Constructing an integration system or layer 'on-top' of a specific collection of participating databases, creating a federated database system.
3. Constructing a multi-database language that can access multiple independent databases, similar to the Google approach.

Hand crafted integrated databases are only realistic when the size of the data collections, in number of items to be integrated, is relatively small, for example less than 1000. The strategy is to collect links to the items that one wants to include in the virtual, integrated collection and build an index with this information. Figure 10 shows an example of this strategy in which the 250 digital copies of items have been collected (manually) from object descriptions from multiple museums in Denmark (Gundestrup & Wanning, 2004).
In the current implementation, each image links to an object presentation – enlarged image with historical data description – that has been stored on the exhibit site. The same index presentation technique could be used to link directly to the multiple source content databases, thereby supporting a seamless (for the viewer) interface to the current locations of the original collection. Unfortunately, this strategy for database integration, though conceptually easy, does not scale since it is limited by available manual labor.

A federated database approach – Portal integration – requires creation of a generic description – metadata – that can describe all of the databases in the federation, as well as translations between the generic and local terminology as well as links to the participating database elements. Referring to the architecture given in Figure 9, the generic description would cover the data in each local I_DB.

It has been suggested that development and acceptance of metadata and ontological standards will allow creation of Web portals to multiple independent databases. (Sowa, 2000; Hyvönen, et al., 2004). User queries for information would be directed to the portal, which would then have the linkage and translation data to direct the query on to relevant database systems. Two elements are required:

1. An agreement on the metadata that must be supplied to describe the content of the local databases, and
2. An agreement on the controlled vocabulary /ontology to be used for selection of metadata values.

In the museum world, there are at least 3 'standard' metadata structures in use: Dublin Core www.dublincore.org , Mpeg-7 http://archive.dstc.edu.au/mpeg7-ddl/, and CIDOC CRM http://cidoc.ics.forth.gr/. Each of these standards has been created/proposed for the needs of a specific community: library, moving pictures, and cultural heritage institutions, respectively. Each defines a framework for the metadata to be collected when describing objects. The standards overlap, but are not identical. Their use creates a problem of searching metadata specified according to different standards.

Defining a general ontology, or controlled vocabulary, has proved very difficult, not least due to evolving language and cataloging cultures. To reduce the complexity and size of the task, current proposals are focused on development domain specific ontologies.

Assuming that agreement can be achieved among a set of information providers as to the metadata framework and the ontologies to be used, a generic description or *integration schema* for the members of the federation can be constructed. Becoming a member of the federation then requires that the participating museum describe its content database using the metadata structure and ontology of the integration schema. A prototype for a portal or integrated schema for several of Finland's museum databases, MuseumFinland, is available at http://museosuomi.cs.helsinki.fi/ and described by Hyvönen, et al., (2004). Once constructed, the generic integration schema can be used to form the base for development of a single interface for the users to provide a seamless search to multiple content databases, and thus reduce the information retrieval search task noted earlier.

The idea of constructing a *multi-database language*, similar to the Google approach, that can access multiple independent databases based on a user query, possibly with synonym definitions was first published almost 20 years ago (Litwin, 1986). The idea is that the user would express his/her information requirement and give relevant synonyms to the query search terms. The processing system would then expand the query and broadcast it to participating database systems. Since little integration work is required, the number of systems that could be accessed is much higher than in the federated approach, which is limited to the number of systems willing and able to join the federation.

A slight modification of this idea would be to give the query language processor access to a generic thesaurus that could map user terminology to the ontologies used for describing content databases. At this writing, this author is unaware of any prototypes attempting implementation of this idea.

Information presentation

For each of the integration strategies mentioned, there is the problem of information presentation. Most prototypes and systems present query results in a list format of text and/or image links. The user is then required to link to each referenced system to determine its real relevance to his/her information need and then to collect and integrate the retrieved data.

Ideally, the information retrieved should be presented in context with the descriptive information presented as an exhibit story. For this to be possible, more research on story construction will be needed. A problem is that much of the metadata recorded today about museum objects is 'administrative', in the sense that it describes the object context (creator, location, materials, etc.) rather than the semantic interpretation of the object and its relationships to other objects. If an *exhibit* presentation format is to be automatically developed, more semantic information about the objects in the content databases will have to be given. First then will it be possible to construct 'real' *exhibits-on-demand*.

Summary and Research Issues

Museums have a vast store of material that can be made available on the Internet. Some is available through virtual exhibits and more is stored in content databases. Virtual exhibits are good vehicles for information dissemination, and when they contain interactive components, can be a good foundation for e-learning systems. However, if an information gatherer/learner is looking for information that is not adequately covered by existing exhibits, but is in existing content databases, it can be difficult to find.

A system to support information gathering should have the following components:

1. An interface structure similar to that found in current, familiar search engines, such as Google.com™. Query terms used to describe the information requirement should be translated into and /or supplemented by synonymous terms that have been used to describe the data. Current proposals to facilitate this translation include the construction of domain ontologies that combine the characteristics and data of taxonomies and thesauri.
2. The system should be able to search at 3 levels for information that match the relevance criteria, as determined from the interpretation of the user request:
   a. Through the museum's virtual exhibits,
   b. Within the museum's database(s), and/or
   c. Through the combined material from multiple museums.
3. Relevant material should then be returned to the user/requester, either as a set of links to the information or, preferably, as a system constructed exhibit, i.e. an exhibit-on-demand.

The advantages of such a system would be to give the user/learner control over selection of information about a topic of interest, expressed in his/her own terms. The system would then be able to interpret the query and utilize known information retrieval technology. However, before this type of system becomes a reality,

- Museums must either agree on the structure of the metadata and the ontologies that they will use to describe their data, or
- Integration tools must be developed to mediate the differences between the set of metadata and ontologies in use.
- It will be necessary to extend the facilities of current indexing techniques to include feature extraction from multimedia databases.
- Methods for automating database integration and/or multi-database querying processors must be developed. And finally,
- Presentation techniques must be developed to provide more meaningful result presentation than the current list structures.

Acknowledgements

This work has been funded in part by the Norwegian National Research Council through project #148827/530. Sincere thanks are extended to the many students and colleagues who have participated in discussions and development of the ideas presented in this chapter while working within the research context of the Virtual Exhibit on Demand project, http://nordbotten.ifi.uib.no/VirtualMuseum/VMwebSite/VEDweb-site.htm.

References

Archives & Museum Informatics. *The international conference on Museums and Web.* (2003, last viewed March 2005). http://www.archimuse.com/mw2003/.

Baeza-Yates & Ribeiro-Neto (1999). *Modern Information Retrieval.* Addison Wesley.

Conklin,J. (1987). *Hypertext: An introduction and survey*. IEEE Computer, 20 (1),17-41.

Elmagarmid, A, Rusinkiewicz,M., and Sheth,A. (eds). (1999). *Management of Heterogeneous and Autonomous Database Systems.* Morgan Kaufmann Publ.

Gundestrup, B. and Wanning, T. (2004). *The King's Kunstkammer: Presenting Systems of Knowledge on the Web.* In Bearman & Trant (eds.). *Museums and the Web 2004: Proceedings*. Toronto: Archives & Museum Informatics. P79-87 also at: http://archimuse.com/mw2004/papers/gundestrup/gundestrup.html

Hillmann, D. (2003). *Using Dublin Core* http://dublincore.org/documents/usageguide/

Hove, L-J. (2004). *Improving Image Retrieval with a Thesaurus for Shapes.* Masters thesis, Dept. of Information and Media Science, Univ. of Bergen, Norway. Available at ***http://nordbotten.ifi.uib.no/VirtualMuseum/Publications/LarsJacob/Thesis.pdf***

Hyvönen, E., Junnila, M., Kettaula, S., Mäkelä,E., Samppa,S., Salminen,M., Syreeni,A., Valo,A., and Viljanen,K. (2004) *Finnish Museums on the Semantic Web: The User's Perspective on MuseumFinland.* In Bearman & Trant (eds.). *Museums and the Web 2004: Proceedings*. Toronto: Archives & Museum Informatics. P21-32 also at: http://archimuse.com/mw2004/papers/hyvonen/hyvonen.html.

Lane,D. and Williams,H.E. (2002). *Web Database Applications with PHP & MySQL.* O'Reilly & Associates

Litwin,W. and Abdellatif,A. (1986). *Multidatabase interoperability.* IEEE Computer 19:12, p10-18.

Lowe,D. and Prince,A. (2003). *Murach's ASP.NET Web Programming with VB.NET.* Mike Murach & Associates

Murray,C. and Everett-Church (2003). *Macromedia Flash Mx 2004 Game Programming.* Premier Press.

Nordbotten, J. (2000). *Entering Through the Side Door - a Usage Analysis of a Web Presentation.* In Bearman & Trant (eds). *Museums and the Web 2000: Proceedings.* Archives & Museum Informatics. p.145-151. Also at http://www.archimuse.com/mw2000/papers/nordbotten/nordbotten.html

Nordbotten, J. (2000-2004). *ADM – Advanced Data Management.* http://nordbotten.ifi.uib.no/ADM/ADM_text/ADM-frame.htm

Nordbotten, J. (2001). *Virtual Exhibits on Demand.* ***http://nordbotten.ifi.uib.no/VirtualMuseum/projectDescription.htm***

Nordbotten,J. & Nordbotten,S. (2002). *Evaluation of User Search in a Web-database.* Proc. Hawaii International Conference on Systems Sciences, HICSS-35, IEEE Computer Society. CD. ISBN 0-7695-1435-9. Aslo at http://www.hicss.hawaii.edu/HICSS_35/HICSSpapers/PDFdocuments/ETNON08.pdf

Peacock,P, Ellis,D. and Doolan,J. (2004) *Searching for Meaning: Not Just Records.* In Bearman & Trant (eds.). *Museums and the Web 2004: Proceedings*. Toronto: Archives & Museum Informatics. p11-20 also at: http://archimuse.com/mw2004/papers/peacock/peacock.html.

Shneiderman, B., et.al. (1989). Evaluating Three Museum Installations of a Hypertext System. *Journal of the American Society for Information Science,* **40**(3), 172-182.

Shneiderman, B. (1998). *Designing the User Interface - Strategies for effective Human-Computer Interaction* (3rd ed). Addison-Wesley.

Sowa,J (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations.* Books Cole Publ. Co.

Yamada, S., et.al. (1995). Development and evaluation of hypermedia for museum education: validation of metrics. *ACM Trans. of Computer-Human Interaction,* **2**(4), 284-307.

Exhibit sites referenced:

American Natural History Museum (last viewed March 2005). *Ology/Marine Biology Exhibit* at http://www.ology.amnh.org/marinebiology/workthesystem/mangrove.html, created by National Center for Science Literacy, Education & Technology at the American Museum of Natural History.

CERN (2004, last viewed March 2005): *The LHC Game.* at http://public.web.cern.ch/public/Content/Chapters/Education/OnlineResources/LHCGame/LHCGame.html

Exploratorium (1998-99, last viewed March 2005). *Solar Eclipse- Stories for the path of totality.* Noel Wanner. *The Sun-Eating Dragon.* at http://www.exploratorium.org/eclipse/dragon.html. Produced in conjunction with AboveNet Communicatios Inc.

Mint Museum of Art (last viewed March 2005). *The Mesoamerican ball game.* at http://www.ballgame.org/main.asp. Produced in conjunction with Interactive Knowledge Inc.

Museum Finland (2004, Retrieved March 2005) from http://museosuomi.cs.helsinki.fi

National Mueseum of Denmark. (last viewed March 2005). *The King's Kunstkammer.* at http://www.kunstkammer.dk/GBindex.shtml Produced with the support of CulturNet Denmark.

Natuurhistorisch Museum Maastricht (1995-2004, last viewed March 2005). *The virtual tour of the museum.* at http://www.nhmmaastricht.nl/engels/index.html

The Wright Experience™ (2004, last viewed March 2005). *The Write Experience in Flight.* at http://www.wrightexperience.com/edu/12_17_03/html/index.htm. Produced in conjunction with Cognitive Applications Inc.

---

[i] The term *e-learning center* is used here to refer to a Web site developed with the intention to provide educational material to students as well as to information gathers from the general public.

[ii] In this chapter, museums are organizations that collect objects and present them in context as information to their public. From a data management and information dissemination point of view, there is no difference between a science and an art museum/center.

[iii] Google.com found 32,500 links for the query "space exploration in science museums". The list included links to museums, as well as to research institutes, business and newspaper articles (Feb.2004).

[iv] Programming tools that enable functions to be embedded in a Web exhibit to support interaction with the user, receiving user input and possibly accessing one or more databases for construction of new response pages.

[v] The ball game was developed by the Mint Museum of Art and Interactive Knowledge Inc. and placed in an exhibit on Mesoamerican cultures.

[vi] CERN developed the version of the simulation shown in Fig.8 in 2000. The simulation presentation was updated and extended in 2004.