

Multimedia CD-ROM as a Medium for Manuscript Preservation and Dissemination: the Design and Development of ‘Treasures of Islam’.

Andrew Large, Jamshid Beheshti and Haidar Moukdad

Abstract

The paper discusses the use of electronic publishing for the preservation and dissemination of rare manuscript material. It is based both upon the authors' earlier work on multimedia integration and upon their experience of designing and developing "Treasures of Islam: Art and Design in Islamic Manuscripts". This trilingual multimedia CD-ROM (with interfaces in English, French and Arabic) contains rare or unique art work, together with a number of complete texts, selected from a collection of Islamic manuscripts held in the Department of Rare Books at McGill University. It combines text, still images, video sequences, music and voices. The paper deals with the following issues: Content selection (selection criteria, authentication); Digitization; Authoring system; Interface design; Information-seeking approaches; Multimedia integration; Multilingualism; Copyright; and Production and marketing. The paper also discusses CD-ROM versus the Web as a publication medium for this type of material.

Manuscript Collections

Manuscripts traditionally have posed major problems for libraries and archives charged with the twin but contradictory tasks of preservation and dissemination. Typically manuscripts are produced in limited numbers, and therefore any particular copy is likely to be rare, and perhaps even unique. This in turn may place a considerable financial value on the manuscript.

Furthermore, manuscripts tend to exist in a more fragile state than, for example, published books. The manuscript may lack any kind of binding or cover, and indeed may comprise a single sheet of paper. If the manuscript is old it may also suffer from the various problems associated with paper and ink degeneration. It is therefore not surprising that many institutions place a very considerable onus on preservation. This can involve such measures as climate control and special storage containers. An additional major means to promote conservation, however, is simply to restrict access to the manuscript; each time a manuscript is handled it risks further physical degeneration. This can encourage policies ranging from refusal to lend the manuscript to other institutions, through restriction to use in a special room (where, for example, pens are banned), to the extreme where the manuscript is only available for display behind glass. Requests to photocopy or even photograph the manuscript may well be denied for fear that the manuscript will be damaged in the process.

Such conservation practices are understandable with rare, fragile and valuable manuscripts. They do conflict, nevertheless, with the expectation that archives and especially libraries will actively

promote the use of their collections. Some manuscripts may be important works of art in their own right, worthy of viewing much as might be a museum or gallery exhibit. But in many cases legitimate requests will be made to handle the manuscript, read it and copy it; refusal to accede to such requests for fear of damage will support the conservation role only at the expense of the dissemination role.

Electronic replication of manuscripts offers one solution to this dilemma. Once the replication has been undertaken, limitless copies can be made with no degradation of quality and no need for further handling of the manuscript (in contrast to photocopying). When the importance of the manuscript lies in its beauty (for example, in its calligraphy or illustrations) the electronic version should be able to capture the full glory of the original. It may also be feasible and beneficial to enhance the original manuscript in the electronic version by, for example, scholarly commentary, indexing or translation into one or more languages. The copies then can easily be disseminated to remote users. Furthermore, the owners of the original may be able to generate income from sales of the electronic versions.

It is not surprising, therefore, that a growing number of institutions should be turning to electronic storage and dissemination of rare manuscript collections. The Memory of the World project undertaken by UNESCO is an especially prominent example of a coordinated global attempt to safeguard and disseminate documentary treasures that are in danger (Courrier & Large, 1997). Such a solution, however, is not without its problems. This paper explores the issues involved in electronic replication of rare manuscripts, drawing especially upon the authors' experience of designing and creating one such product. It also discusses the advantages and disadvantages of the two most popular electronic technologies for such work: CD-ROMs and the World Wide Web.

Treasures of Islam

The Islamic Studies Library, and the Rare and Special Collections Division of the Library at McGill University contain impressive holdings of manuscripts from the Islamic world. These include texts on jurisprudence, philosophy and theology in Arabic, Persian and Ottoman Turkish dating back as far as the thirteenth century. Many are in a fragile condition, inadequately documented and not easily available to the public. Islamic manuscripts are valued not only for their intellectual content but also for their physical beauty. Authors, calligraphers, painters and binders collaborate to produce works of art. As such, they provided a rich source from which to

select the content of what would become the *Treasures of Islam*, a multimedia CD-ROM published in 1999 by McGill University.

The *Treasures of Islam* CD-ROM comprises four sections: 40 examples of Islamic calligraphy in several different styles including Kufi, Thuluth and Naskh; 42 miniature paintings representing a variety of styles and sources; 16 richly decorated lacquer bindings; and eight complete manuscripts dating from the thirteenth to the nineteenth centuries. In addition, a short video clip and spoken summaries of each section are included. Three separate language versions of the interface are provided on the disc: Arabic, French and English. Not only the interface itself, but all supplementary texts (introductions to each section and captions for each image) and speech segments are available in the three languages.

During the design and development of *Treasures of Islam* many decisions had to be taken concerning content, digitization, authorware, the interface, information seeking approaches, multimedia integration, multilingualism, copyright, and production and marketing. The solutions to these problems had to relate to the specific product under development, but the issues raised are generic and likely to be encountered by many multimedia developers.

Content

Although electronic versions of manuscripts can be used for preservation and dissemination, a considerable amount of time and effort must be expended in order to achieve this end. The content must justify the expense, and this is especially the case if it is intended to market commercially the product to cover costs or to generate income. Several selection criteria can be employed, including the rarity of the manuscript, its aesthetic interest, its financial value, and the level of demand for access to it. Decisions on content normally will be taken by subject experts rather than the technical designers. In the case of *Treasures of Islam*, for example, the Librarian of the Islamic Studies Library and the Curator of the Rare and Special Collections Division provided their expertise for content selection.

Many of the manuscripts had not been well documented, and in order to present them on the CD-ROM in a meaningful fashion it was necessary for further research. Such an eventuality had not been anticipated by the developers, and indeed funds were unavailable within the digitization project for this kind of work. Fortunately, the Islamic Studies Librarian (Adam Gacek) generously was willing to undertake the necessary scholarship regardless. The lesson to be drawn

here is that when planning such a project it should not be assumed that the manuscripts can be transferred to a new medium without provision for scholarly as well as technical assistance.

Another consideration for content is the copyright status of the material selected for inclusion. The developers must be certain that they have the right to use the material without infringing copyright. Obviously this is especially important if it is intended to produce a commercial product for sale. Even if copyright is not a problem, the library or archive custodians may be reluctant to subject the manuscript to the rigors of scanning, a necessary step in digitization. Assurances may be demanded that every care will be taken with the manuscript, and it may be necessary to arrange that scanning be undertaken in the manuscript department, thereby obviating the need to remove it to the digitization project's premises.

The sound content of *Treasures of Islam* caused a number of problems. A decision was made to include background music, which naturally had to have links with the Islamic world. It proved difficult to find suitable music, even though only short segments of a few minutes' duration were required for looping. Copyrighted music could not be used (unless copyright permission was obtained and any royalty fees accepted), and finally a local Islamic musician agreed to record for the CD-ROM (using the recording facilities of McGill's Faculty of Music). Voice commentaries were required in the three languages of the CD-ROM, and after much experimentation with various voices and primitive amateur recording equipment, the Canadian Broadcasting Corporation (CBC) was hired to provide both professional readers and professional recording facilities.

Digitization

The manuscripts must be converted from paper to a digital version by scanning each page. Text can be stored as ASCII code very economically (a CD-ROM will hold around 250,000 manuscript pages). The situation is very different for still images, and especially for sound or video. The short high quality video clip of one minute on *Treasures of Islam* consumes 40 megabytes of disc space!

Digitization is not only a slow process that must be undertaken carefully if handling rare and valuable manuscripts, but also one that calls for several decisions that will prove critical for the final product. In particular, a decision must be taken concerning the level of resolution to be employed. The higher the resolution, the better the quality of the resulting image. This is likely to

be especially important if the manuscripts contain art work (as with *Treasures of Islam*). It should be noted, however, that image quality will also be affected by the kind of monitor and video cards to be used when viewing the product. Although high resolution will give a sharper image than low resolution, it can only be achieved at a price – the amount of disc storage space needed for the manuscripts. In other words, there is a trade off between resolution and space. An increase in resolution from 640 x 480 to 1024 x 768, for example, increases file size by more than 2.5 times. Although a CD-ROM will hold more than 650 megabytes, space can quickly be at a premium (this problem is alleviated with DVD technology that increases storage capacity seven fold). If high resolution is chosen then it may be necessary to compromise over the size of the manuscript collection or the way in which it is displayed (see below).

Decisions over resolution are also influenced by the kind of hardware that users are likely to employ. If the CD-ROM is to be accessed only in-house then this information is available to the designers; otherwise they are compelled to take the best decision they can, balancing image quality against the size of the potential user community. For example, selection of low resolution (640 x 480 pixels) will ensure that the CD-ROM can be played on any of today's computers. Although higher screen resolutions typically are now found (especially 800 x 600) many portable (laptop) computers still operate at lower resolution. It also ensures the best results regardless of how the resolution is currently set on the computer. On the other hand, if it is known that the CD-ROM will be used only on high-end desktop computers then a high resolution of 1024 x 768 will generate the sharpest images. It might seem a good idea to leave resolution choice with the user by providing more than one version on the disc. Unfortunately, this would require not only much more disc space but also the development of multiple interfaces, each one with its own screen design.

Image quality is also affected by color depth. Sharper colors are obtained if the user's computer is set on high color (16 bits or 24 bits per pixel) rather than eight bits.

Authoring Software

Software must be developed or more likely purchased to create the multimedia product. A number of suitable commercial packages are available, but the most popular (and the one used to develop *Treasures of Islam*) is Macromedia's Authorware. This icon-based authoring program is used extensively in academic development environments, in large part because of its versatility in integrating multimedia components.

Interface Design

Interface design critically affects the usability of the program. Many authors have identified criteria for successful interface design. These include considerations such as the time it takes a user to learn the interface, the ease of retaining this knowledge once acquired, the time taken to issue an instruction and to receive a response, and so on. Interfaces should also be consistent, predictable and responsive to the user. (see Shneiderman, 1998, Galitz, 1997).

It is both exciting and challenging to be given an opportunity to design an interface from scratch. The first step is to decide upon an overall design concept (or metaphor) for the interface. In the case of Treasures of Islam, a library building was initially selected as the metaphor; each of the four components comprising the content would be represented by a room within the library. Users could then “enter” a particular room and begin to select material from the shelves. This metaphor was later abandoned for the metaphor of a book (since the CD-ROM contained information about various aspects of manuscript books). Once the metaphor had been selected, the interface could take shape – the content was represented as four chapters, a title page and a contents page led into the contents themselves, the screens were designed as “pages”, the colors employed were chosen to give the impression of a manuscript book, and so on.

Careful attention must be paid to screen layout. Cluttered screens should be avoided, and design motifs should be employed consistently throughout the interface. A lot of time can be spent on color and font selection. Color schemes are meant to enhance clarity in the interface rather than to construct a psychedelic display. Certain colors work better for backgrounds, others for lettering. Italic fonts are less effective on screens than on paper, and sharply defined fonts are to be preferred. Colors and fonts should be used sparingly and to draw user attention to specific activities on the screen. Above all, they should be used consistently throughout the entire interface.

A critical element within the interface is the dialog transaction mode – the means by which the user can issue instructions to and receive responses from the computer. Traditionally this was accomplished via a command language, but more typical today are various kinds of menus, windows, buttons and icons, as well as various direct manipulation techniques.

Information Seeking

How does the user locate information on the CD-ROM? It would be possible to provide nothing more in the way of navigational tools than the capability to progress through the manuscripts sequentially one screen at a time. Such a minimalist approach would neither equal the possibilities offered by traditional print format (flicking through pages, opening at a random page, locating specific pages through a list of contents or a back-of-book index, etc.) nor exploit the many sophisticated possibilities available in a computerized version. Fundamentally, two navigational approaches are available (both can be employed, of course, with any one product): searching and browsing.

Many designers will opt for a searching mechanism so as to allow specific screens to be retrieved quickly. This will involve the construction of an index through which individual words or phrases can be located in the text. Two approaches to index construction can be taken. The content can be examined by a human indexer to identify concepts that are considered potentially searchable by users. Alternatively, software can be used that will generate a keyword index of all the terms from the text, or perhaps only from certain segments of the text such as titles (typically excluding frequently employed “stop words” that do not convey subject information). The resulting index (or indexes) can then be presented as a browsable list from which terms can be highlighted by a user for selection. Alternatively, a search engine might be included that will allow users to input a term into a search box and then will search the index for occurrences of that term. The search engine may permit several terms to be combined using Boolean operators as well as offer features like truncation and phrase searching.

Searching is effective to locate specific information to meet a well-defined user information need. In other cases, however, users have a less clearly formed idea of what they are looking for; they would prefer to browse through the information store, perhaps to get an overview of its contents, to see what they might discover by chance, or simply to read screens sequentially. Browsing can be facilitated by a number of devices. As mentioned under Interfaces (see above), menus, buttons and icons can be provided that allow users to move rapidly around the interface. Typically every screen will provide the means (a button or an icon) to move forward one screen, back one screen, to return to the opening screen or to exit the program altogether. In *Treasures of Islam* users can move quickly and easily to any particular screen within a chapter by clicking on the screen number pad displayed on the right-hand side of the screen (see Figure 1). An effective way to browse images (which are difficult to search except using an accompanying textual caption) is to

display an entire set of images, reduced to thumbnail size, on one screen; the user can then click on any thumbnail image to select it in full size. Hypertext links offer a popular and highly flexible browsing mechanism, though in a large data store ancillary aids like a list of screens already visited, may be necessary to avoid disorientation.



Figure 1.

Multimedia Integration

When media such as still images, video and sound as well as text are included on the CD-ROM, thought must be given as to how they might most effectively be integrated and made accessible. Images, for example, can be displayed on the screen along with text; alternatively, either the image or the text can be hidden unless activated by a button or menu. In the Miniatures chapter of the *Treasures of Islam* the image is always present on the right-hand half of the screen, but the caption is hidden; in order to activate it a button must be clicked (another click will then close the caption window). As the image reproductions of calligraphic examples, miniature paintings and lacquered bindings constitute an important part of *Treasures*, attention was paid at the design stage as to how best they might be displayed. Several techniques are used in the various chapters to display enlargements of all or a part of the images. One is to allow the user, with a button click,

to enlarge the entire image so that it fills the entire screen, eliminating the half normally used to contain the various navigational buttons. Another is to allow the user to enlarge to full screen either the top half or the bottom half of the image. A third is to enable the user to enlarge just one portion of the image (this portion is pre-determined by the designers; it would have been good to allow the user to choose which part of the image to enlarge, but this would have meant storing multiple representations of each image, thereby greatly increasing demands on disc storage space).

Sound segments may also have to be accommodated. *Treasures of Islam* includes both the spoken word and music. In both cases the sound can be toggled on or off by the user, using buttons. It is important always to allow users to switch off sounds should they wish.

Multilingualism

The impressive storage capacity of a single disc can make the idea of multiple language versions attractive to designers – why leave empty space on the disc when it can be used effectively? In the case of *Treasures of Islam*, it seemed particularly suitable for multilingual treatment, being produced at McGill (an English-speaking institution) in Quebec (a French-speaking province) for the largely Arabic-speaking Islamic world. This decision, lightly taken, proved the most troublesome of any made. It was straightforward to develop three versions of the interface, one for each language, accessible from the opening screen (see Figure 2). Translation of all the text on the CD-ROM (except the actual eight manuscripts themselves that are only to be found in the original Arabic), the voice commentaries, and the CD-ROM's printed insert booklet and printed covers proved more complicated. Difficulties were focused on two areas. First, the specialized vocabulary used in the introductory texts to each chapter and the image captions was not always easy to translate from the original English version into French and Arabic to the complete satisfaction of all the subject and language experts consulted. Second, each time that a change, major or minor, was made to any part of the English original, that change then had to be replicated accurately in the other two language versions. The lesson to be learned here is that translation should not begin until the original text is finalized.

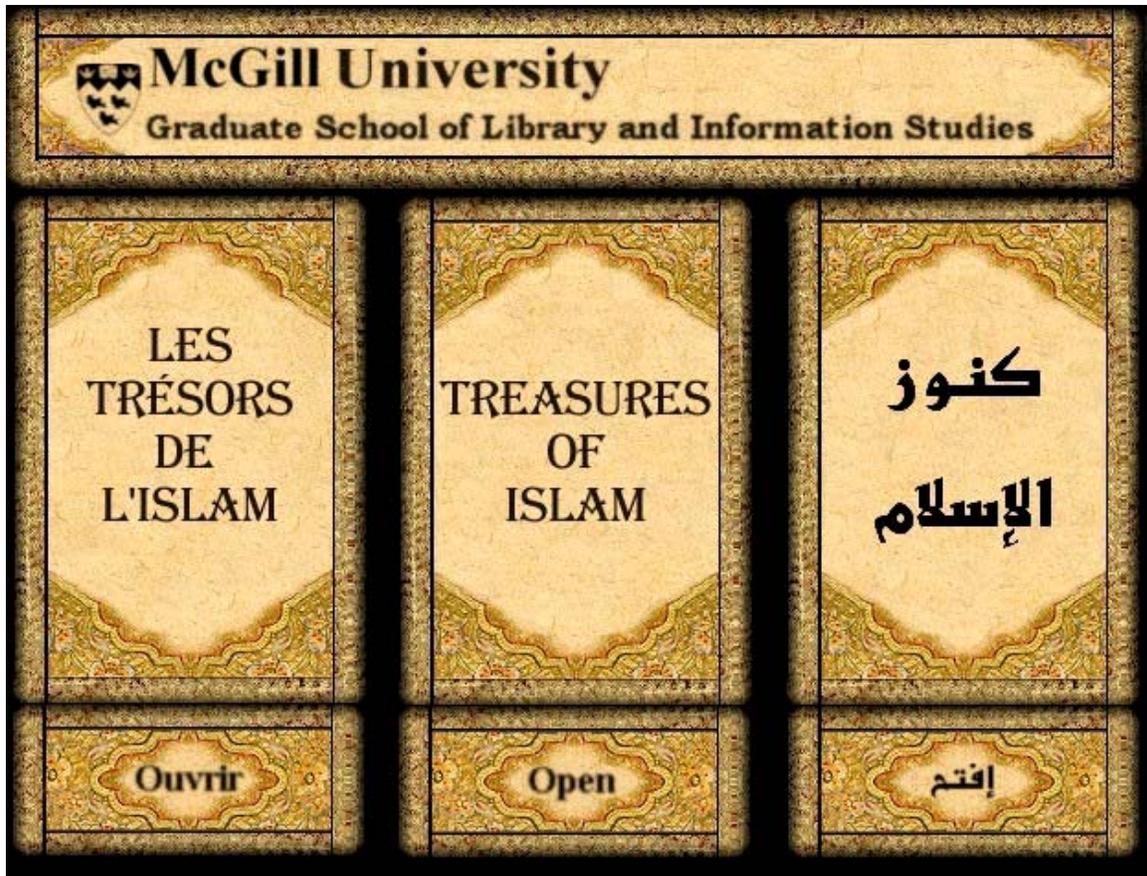


Figure 2.

Copyright

The copyright issue concerning material to be included on the CD-ROM was alluded to briefly above. This section rather deals with the copyrighting of the CD-ROM and its contents, a step that may be considered necessary if the disc is to be marketed commercially.

Copyrighting the content of a CD-ROM might be likened to thief-proofing an automobile – it is nigh impossible to prevent the determined thief, but a few sensible measures can deter the majority of potential law breakers. McGill University, the publisher of Treasures of Islam, was concerned that its copyright over the CD and its content be clearly stated. At the same time, a major attraction of the disc was the beauty of the many images it contained. This precluded the addition of “watermarks” to all the images – the application of a background copyright statement on which the images themselves would be superimposed. It was decided manually to add a copyright statement in small print at the foot of each image on the disc (even though these could be deleted by those so inclined) together with copyright statements on the disc itself as well as the

back cover of the jewel box and the accompanying booklet. It can also be noted here that an ISBN was obtained (and printed on the disc, cover and booklet) and a copy of the CD-ROM deposited with the National Library of Canada).

Production and Marketing

Prototypes were burned in-house, but this is still a relatively slow process if production runs are to be in the hundreds or thousands rather than the tens. Production was therefore contracted to a commercial pressing plant which handled not only manufacture and labeling of the discs themselves, but printing the covers and booklet, and packaging the finished product within a cellophane seal.

Marketing is a more difficult issue. The designers themselves had been more interested in the challenge of producing a finished product than in identifying niche markets for the product. A CD-ROM including rare Islamic manuscripts from a university collection might seem to have an easily identifiable, if limited, market – the libraries of other universities with an interest in Islamic studies, and especially Islamic art. This is certainly the prime target. But other potential markets can be envisaged. Educational institutions with an interest in the art rather than Islamic aspect of the disc might be targeted, a much wider group of institutions. Various multimedia marketing experts who were consulted suggested more ambitious strategies; one, for example, suggested it as a suitable departure gift for travelers leaving airports in the Islamic world! Interesting as such general markets might be, the problems of marketing in them for a university should not be under-estimated.

CD-ROM vs the Web

Since the CD-ROM's debut in 1985 many thousands of titles have been published, attracted by low production costs, high storage capabilities and widespread availability of CD-ROM drives. It has proven a particularly attractive means for distributing multimedia content of all kinds. Since the early 1990s, the growth in content and connectivity on the World Wide Web has been dramatic, and it also is associated with multimedia content. Which technology is most suitable for the preservation and dissemination of manuscripts?

The Web certainly offers a number of attractions. Once the content has been placed on a Web server it can be seen by anyone with Web access. Although this eliminates production issues, it does not entirely eliminate marketing considerations. No Web search engine indexes the entire

Web content, and a site is only likely to be accessed by users if it is ranked high in any retrieved hit list. Both problems can be addressed at least partially by the producers informing the search engines of their site (see Beheshti, in press, for a study of the extent to which various search engines give access to marketing pages on *Treasures of Islam*). It is also relatively straightforward to create web pages with multimedia content. Finally, the Web is seen by many as the only tool now worth using for any kind of dissemination, and alternatives, whether print or CD-ROM, as heading rapidly for extinction.

The CD-ROM nevertheless offers several important advantages as a means of disseminating multimedia information. First, the quality of its images and sound are still hard to equal on the Web, even with the fastest bandwidth and best hardware. Second, limited bandwidth on the Web can make transmission of images and sound, and especially video, unreliable and unsatisfactory. Third, despite the expanding tentacles of the Web, some parts of the world still encounter problems in accessing networked information. Fourth, if it is intended to sell the information content rather than freely make it available, it is more straightforward to sell it as a product on CD-ROM rather than to try to get payment for providing web access; the Web remains strongly associated in many minds with free information. Fifth, CD-ROM may still raise some doubts about long-term durability and hardware obsolescence, but it is likely to be a more suitable medium than a web site for the preservation of fragile paper manuscripts. Finally, the CD-ROM can be acquired and added to the collection so that it is always available for consultation; web sites are much more ephemeral in nature.

East and Leach (1998) conclude from a study of CD-ROM use by academic libraries that “despite challenges from alternative modes of delivery, CD-ROM retains its popularity as a format for the storage and retrieval of electronic data.” If a library or archive wishes to make a manuscript collection widely available in a high quality version while safeguarding the original from constant handling, and simultaneously generate a little income for further collection development, then the CD-ROM remains a very attractive proposition despite the virtues offered by the Web.

Acknowledgements

Treasures of Islam was designed and developed in part with support from the Canadian International Development Agency (CIDA).

References

Beheshti, Jamshid. Looking for the eye of the needle in a haystack: searching the Web. *International Conference on Advances in Infrastructure for Electronic Business, Science and Education on the Internet*. SSGR 2000, l'Aquila, Italy, July-August 2000.

Courrier, Yves and Large, Andrew. *World Information Report*. Paris, Unesco, 1997, p. 336.

East, H. and Leach, K. The continuing prominence of CD-ROMs in academic libraries: the findings and evaluation of a survey. *Aslib Proceedings* 50 (8) 1998, 208-214.

Galitz, Wilber O. *Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. Nw York: Wiley, 1997.

Shneiderman, Ben. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 3rd ed. Reading: Addison-Wesley, 1998.