Middle East Digital Libraries: Some Principles for Development

My vision is that there will not be a single Middle East Digital Library but many, and that each will contain multiple, high-quality collections of primary and secondary materials. These collections will be from or about the Middle East and aimed at specific types of users either in the Middle East or interested in the region. The material in these collections would be represented as complex digital objects that are persistently identified and contain secondary information about the material, as well as either a datastream of the material itself or a link to its location elsewhere.

Overall, the libraries would be structured architecturally in multiple technical layers. At one layer, the digital objects would exist in a preservation format and be subject to appropriate preservation processes. At another layer, use copies would be accessible with interfaces for direct user interaction as well as various types of machine processing designed to facilitate general activities such as search, discovery, and analysis. These libraries may also have additional layers equipped with tools and facilities specifically designed for certain kinds of specialized uses that involve the library holdings such as teaching, editing, or publishing. Requirements and standards for interoperability among the objects and between objects in different libraries will vary depending on the layer at which it occurs.

Because digital library development is still in its infancy, there is considerable unevenness across existing libraries, archives, museums, and other related institutions in their interests and abilities to participate in digital library development—and in the standards, tools, and processes that operate at each layer. In some cases, basic infrastructure, such as network connectivity or catalog description, may need to be developed as a precondition of participation. Elsewhere, considerable progress has been made and valuable lessons learned that can guide further digital library development. Below I outline some of the digital library developments related to Middle Eastern Studies, in which The Andrew W. Mellon Foundation has been involved. These include work in JSTOR on the development of relevant secondary sources, a series of projects developing collections of primary source materials, and an initiative specifically focused on the use of digital materials for teaching. I then draw on this experience to suggest several key principles that might prove useful to workshop participants in planning further development of Middle East digital libraries.

Selected Mellon-supported projects.

JSTOR. Originally created with Mellon support in 1995, JSTOR now preserves and provides access to a digital archive of full runs of nearly 600 scholarly journals, including almost 20 million pages, in more than 40 different scholarly disciplines. There are 2,650 participating libraries in 98 different countries whose users viewed 128 million pages and printed 28 million articles in 2005 alone. JSTOR includes a small collection of eight journals specifically focused on Middle Eastern Studies, and is in the process of adding another ten. JSTOR journals in other fields, of course, contain many other relevant articles. See www.jstor.com.

ARTstor is a non-profit initiative, founded by the Mellon Foundation, with a mission to use digital technology to enhance scholarship, teaching and learning in the arts and associated fields. It provides a repository of hundreds of thousands of digital images and related data; the tools to actively use those images; and a restricted usage environment that seeks to balance the rights of content providers with the needs and interests of content users. ARTstor is adding approximately 25,000 high quality digital images of the art and architecture of Islam from the personal archives of Professors Jonathan Bloom and Sheila Blair of Boston College, and Professor Walter B. Denny of the University of Massachusetts, Amherst. See www.artstor.org.

Advanced Papyrological Information System (APIS). APIS links together in a single environment various sources of information about texts written on papyrus held by several major research libraries in the United States. APIS contains descriptions of the papyri, digital images of many of these texts, connections to databases with the texts in their original languages, a specialized search engine, and an interface that permits the user to move back and forth among text, translation, bibliography, description, and image. In 2003, Roger Bagnall, who launched APIS, received a Mellon Distinguished Achievement Award and is using the funds to invest in both the technological and papyrological work necessary to enhance and make interoperable APIS and a number of other large digital resources for papyrology created in the past few decades, especially those associated with repositories in Europe. See www.columbia.edu/cu/lweb/projects/digital/apis/.

Duke Data Bank of Documentary Papyri (DDbDP). Created with funds from the Packard Foundation, the DDbDP is an electronic corpus of Greek and Latin texts found on papyri, ostraca, or wooden tablets, which have been published in discrete volumes or in series. The DDbDP contains all texts so published to 30 June 1996, representing over 5,000,000 words and nearly 500 volumes, and has been available through the Perseus Project (www.perseus.tufts.edu/cache/perscoll_DDBDP.html). With a planning grant from the Mellon Foundation, papyrologists, led by Josh Sosin at Duke are developing a vision of what how the DDbDP could serve research and teaching; what technological solutions are needed to implement this vision; and how a significantly enhanced DDbDP could best be embedded in Duke's institutional infrastructure, so that it may be assured of a long and healthy future.

InscriptiFact. Developed by Bruce Zuckerman and his colleagues from the University of Southern California with partial support from the Mellon Foundation, InscriptiFact is a data and image base system for the distribution of images of ancient inscriptions that originated in the Near Eastern and the Mediterranean World. Zuckerman's Western Semitic Research Project contains an archive of more than 100,000 photographs of rare and fragile text-fragments that are located in museums and archives around the world and at field sites in the region. These photographs were created often under various sources of illumination, using the latest photographic and imaging techniques to create the most legible images possible. Approximately 20,000 of these images are now available digitally in high resolution in the InscriptiFact system, which makes it possible to bring together, view, and compare images of inscriptions that would normally be difficult or impossible to study in association. Given the analytical tools that are being developed for the online archive, InscriptiFact is an emerging resource that offers substantial value and utility to archaeologists, philologists, linguists and other scholars and students of Middle Eastern culture. See www.inscriptifact.com/.

Giza Archives Project. Between 1905 and 1942, the Boston Museum of Fine Arts and Harvard University jointly sponsored excavations of the Giza Pyramids and necropolis. These excavations resulted in a massive collection of ancient Egyptian objects, ranging from masterpieces of royal sculpture to everyday tools and implements of daily life, accompanied by meticulous documentation in the form of diaries, object registers, site plans and maps, drawings, and photographs. Upon receipt, as was the customary practice in archaeological collections, the Museum of Fine Arts divided and separated the objects by format for storage and access. Just as digital technologies make it possible to reunite collections virtually from different institutions, as in InscriptiFact, they also make it possible to reunite the material from different parts of a single institution. Using digital technologies and with support from the Mellon Foundation, and under the leadership of Peter Manuelian, the Giza Archives Project at the Boston Museum of Fine Arts is advancing international scholarship focused on Giza by virtually reuniting the materials and integrating them around the scholarly unit of study—the tomb or mastaba—and making possible a variety of other combinations for research and teaching. See www.gizapyramids.org/code/emuseum.asp.

Al-Musharaka. NITLE (pronounced "nightly") is a not-for-profit organization created by the Mellon Foundation to promote innovation and collaboration around the effective use of technology for teaching, learning, scholarship, and information management in liberal arts colleges in the U.S. One of the first projects of NITLE was to create Al-Musharaka, a collaborative web-space in which faculty and staff from participating colleges are able to implement a robust approach to teaching and curricular development in Arab, Islamic and Middle Eastern studies. The offerings include a blog, listserv, support for inter-campus teaching, and the Arab Culture and Civilization Web site of on-line multimedia course materials including a variety of original texts, images, video clips, and audio files that have been originally developed or culled from various print, media, and online sources. The course materials are organized thematically into a set of 11 modules covering such topics as Ethnicity and Identity, Literature and Philosophy, Popular Culture and Performing Arts, Family and Society, Art and Architecture, and the Arab Language.

Principles for digital library development.

The experience of the Mellon Foundation is these various projects suggests several key principles that will likely be critical in the development of digital libraries focused on the Middle East.

The first guiding principle is that digital libraries must seek to ***create scholarly value by exploiting the distinctive features of the technology***. Investments in electronic journals, such as the materials available from JSTOR and from most major academic publishers, have proven to be clear winners because of the economies and ease of use afforded by using the technology to aggregate and search text in thousands of articles. Similar benefits are being achieved for reference works such as encyclopedias and dictionaries and, with recent investments of Google, Yahoo, Microsoft, and others, for books and monographs.

The scholarly advantages of applying digital technology to the collection and dissemination of primary sources are no less real but have proven somewhat more elusive. Few institutions, for example, have begun systematically to assemble primary source collections of Web pages, e-mail correspondence, electronic manuscripts, software programs, electronic games, scientific datasets, and other uniquely digital artifacts that would help serve as a record of modern culture for future scholars. However, many have experimented with digitizing existing collections of primary sources in order to make them more accessible. Many of these projects have digitized the material simply in the hope, often unrealized, that an audience will emerge that is willing and able to sustain the collection. Avoiding the risky "field of dreams" approach requires, in part, a careful appraisal of how the technology can be exploited for primary sources to create scholarly value. For example, ARTstor, InscriptiFact, the Giza Pyramid project and the papyri initiatives aggregate related materials, which are otherwise widely dispersed, and do so at a high quality making possible comparison and analysis at levels of sophistication that could not otherwise be attained.

Second, assuming an investment in the technology is warranted, a second guiding principle for digital libraries is to ***build collections of coherence and integrity***. Many of the early efforts to digitize primary sources have placed insufficient emphasis on intellectual integrity and coherence as criteria for selection. In some cases, digitizing projects have settled for a highly, selective "greatest hits" approach, which illustrates and interprets a collection rather than making it available. As a rule, the more a collection is bound to a specific interpretive agenda, the less useful it is to a general audience. In other cases, databases of images, texts or other materials may draw on contributions from organizations that are collaborating simply for the sake of the collaboration without concentrating on generating a resource that is genuinely useful to scholars, teachers or other audiences. As an alternative to these paths, coherence and integrity can be achieved by being comprehensive, digitizing all or nearly all of a strong existing collection. If selectivity is required, experts who would themselves use the resource that is being developed must assist in deciding which materials to include and how they should be directed toward specific pedagogic, research or other need without sacrificing the core qualities of the collection.

Even in cases where coherence and integrity of the collections are a prime objective, the high quality that an audience might demand can be significantly impeded by a failure to deal directly with intellectual property issues. Many digital library projects aim too low, settling for poor-quality access—in the form of thumbnails of visual materials, for example—as a way of skirting these critical issues. The legitimate rights of content owners must of course be protected, but the third guiding principle is for digital libraries to do so in ways **that *protect and foster an intellectual commons for scholarly and educational uses***. Many of the advantages of the technology are gained only by aggregating materials from various sources and making them, as one observer has said, "processable" or amenable to computational uses. These activities often require permissions because the uses may violate copyright, the moral rights of owners, or both, and the diplomatic, legal, and other skills needed to negotiate and obtain access for scholarly purposes to various resources are typically in very short supply among digital librarians and their technical teams. The Mellon Foundation's experience in developing JSTOR, ARTstor and the Middle Eastern projects suggests that several distinctions may be useful in such negotiations. For example, to the extent possible, commercial uses of copyrighted materials should be rigorously distinguished from noncommercial, educational uses. Negotiations are especially difficult if a

digital library project is perceived to compete with or undermine an owner's interest in commercial exploitation. In the event of such a perception, it is often useful for a digital library to provide a well-regulated environment for noncommercial, educational uses, in which content is available not to all comers but only to authorized users of subscribing institutions that are subject to a strictly enforced user license.

The fourth guiding principle for the development of digital libraries is to ***be realistic about costs, especially the costs of distributing content and sustaining ongoing operations***. Building digital libraries is expensive; the costs are not just technical but, as we have seen, involve aiming the technology at specific goals, carefully selecting content, and managing intellectual property. There are other significant costs, including those of cataloging and network infrastructure. But perhaps the most important and most overlooked costs are those associated with distribution and the organization of ongoing support. Many individuals and institutions are seeking to digitize important materials as primary and secondary sources for scholarship, but with rare exceptions the projects are relatively small-scale, are isolated in data structure, and face enormous challenges in finding an appropriate means and scale of distribution. Entry costs may be low, providing an illusion that cottage-style industry is viable over the long term, but digital library enthusiasts are often better at thinking of exciting things to try out than they are at undertaking the more mundane work needed to demonstrate that a project will have a suitable administrative home with able leadership, enjoy the infrastructure support that it will require, and develop an intelligible business plan that includes at least the potential ability to generate the resources that will be needed for the project to have a lasting impact. Without full attention to the organizational and other costs, digital libraries will be doomed to a Hobbesian life: nasty, brutish, and short.

When digital library development succeeds, the process and the results can prove to be hugely liberating and democratizing, opening new realms of intellectual inquiry and new levels of educational attainment. But to unleash these forces in any systematic way, the need is huge for collaboration across traditional organizational boundaries of various kinds. Collective organization is needed involving shared financing and responsive governance in ways that are probably unprecedented and this need raises a grand challenge in the form of a set of policy questions with which we must ultimately grapple in our workshop. The key policy questions include: What are the right models and incentives for intra- and inter-institutional collaboration in a rapidly changing environment? How can cost effectiveness and economies of scale best be achieved in the creation, use, and dissemination of scholarly resources? What is necessary to ensure the long-term sustainability of the resources and their stewardship? How can the intellectual commons of intellectual property best be nurtured and protected to promote the public good of open intellectual inquiry?

Donald Waters
Program Officer for Scholarly Communications
The Andrew W. Mellon Foundation.