Digital Libraries, Cultural Heritage and Interpol

Contribution to `Post Digital Library Futures'
Bob Englund, Professor of Assyriology, UCLA
1 June 2003

Recent events in Iraq have reminded us of one of the goals of IT research and applications, that of cultural heritage capture and preservation. In the US and Europe, public institutions have begun to invest some effort in the digital documentation of often fragile archaeological assemblages. The immediate beneficiaries of such emerging digital libraries are of course the communities of museum personnel and of researchers in a number of related fields, including archaeology, art history, palaeobotany, history of science and technology, but also in the case of dead languages such disciplines as linguistics and semiotics.

Cuneiform collections in such disparate sites as St. Petersburg, Ankara, Rome, Chicago and Banning, California, have been targeted by the Cuneiform Digital Library Initiative for inclusion in centrally managed data sets using

- standardized, and widely available methods of the electronic capture and permanent data archiving;
- a flexible and interactive access system containing tools for a networked presentation of early writing both at the level of individual text collections and of virtually reconstituted ancient archives now spread across the globe;
- XML and data-mining tools facilitating the access of a broad public to writing systems and dead languages of ancient Babylonia, with a built-in scalability for use with other writing systems of the ancient world; and
- future-oriented data infrastructures for the continued internet projection of cultural information and heritage.

Cultural heritage partners propose combining the educational potential of networked virtual museums to project into a world community the historical significance of artifacts seldom on public display; to capture in digital inventory large text corpora of dead languages and place this documentation in the public domain; and in so doing to illuminate the character of very foreign civilizations. By creating a unified access to representative corpora of early writing and using computing technology to facilitate their semantic interpretation, IT projects intend to foster scholarly cooperation between researchers and cultural institutions. At the same time, "lines of communication" to the heritage of pre-scientific civilizations dead many millennia will be opened to a networked public.

These are directions that digital library implementation in the field of cultural heritage might take if we are to effectively exploit in future many of the methods of digital capture and internet distribution of knowledge developed in phases one and two of the Digital Library Initiative. However, the current situation in Iraq places the IT community before immediate, pressing needs. If through domestic turmoil, or the invasion by a nation ill-prepared to safeguard cultural heritage as required of the signatories of the UNESCO Convention for the Protection of Cultural Property in the Event of Armed Conflict, antiquities directorates in Afghanistan and Iraq have seen themselves stripped of tens of thousands of historically unique artifacts.

Addressing the dilemma posed by an often intricate system of black market distribution of stolen antiquities reminds IT researchers of questions met in the past. FBI and military intelligence teams on the ground in Iraq, but also Interpol and the culture sector at UNESCO under Assistant-Director General M. Bouchenaki, have called for the development of central and user-friendly data bank applications to be used in the interdiction and confiscation of artifacts now reported to be crossing international borders. How shall we implement user interfaces that assist in the identification, for instance, of 3rd millennium cuneiform accounts, at distant customs houses and border crossings, and by officials unable to distinguish a
Babylonian tablet from a decorated stone, let alone to identify the inscription of a text of known museum provenience? Since we are very far from an optical recognition capability that would automate such a process, we must consider developing IT tools that will allow policing agencies to capture and transfer to specialists data sufficient for human support in computer-assisted object recognition.

Considered as a problem facing IT engineers, the capability of a relatively lay public, represented by policing agents, to interact with data banks written by and for academic specialists shares much in common with data structure and user interfaces being written for visitors of virtual museums generally. Since the modern virtual museum will contain and present an overabundance of documentation on all artifacts in the possession of the physical collection, its user interface must care for a presentation layout and access system that reduces the cost of knowledge acquisition to users, while retaining the strength of data base content. Translated to an Interpol collaboration with managers of digital libraries with antiquities content, such an access system must build in a level of direct access to networked specialists whose training in both archaeological artifact recognition and in data mining make them suitable for quick searches of existing data. For instance, legible sign strings in digital images of clay tablets uploaded to a networked professional would be read and collated against catalogued cuneiform texts, with eventual matches and their respective digital documentation returned to the querying agency for confiscation and possible prosecution. The report of the UNESCO meeting of 17 April held in Paris has made clear that for the foreseeable future this

policing agency within Iraq is the Office of Reconstruction and Humanitarian Assistance for postwar Iraq
(ORHA), created on 20 January by presidential directive, organized within the US Defense Department and put under military command of the US Central Command. Antiquities that have in the meantime entered Western markets fall under the purview of national and international law enforcement agencies.

Both FBI and Interpol will work closely with experts from the field of Ancient Near Eastern studies to build a capability to recognize and typologize seized and unprovenienced antiquities. Streamed tutorials written by these experts along with rough image search capability will attempt to simplify the identification process, however unless for instance HR 2009, which proposes to forbid the importation of all Babylonian antiquities into the US, is passed, customs and policing officials will be hard pressed to isolate for prosecution those artifacts that only trained experts can be expected to easily identify as either looted from known museum collections, or stolen from excavation sites now being systematically plundered in southern Iraq. This is particularly true in the case of cuneiform documents with physical and text content.

In a certain sense a policy of assigning humans as living assistants to IT projects is not unlike developments we have seen in the ongoing battle against spambots that harvest email addresses for their annoying business interests, in which now humans assist computers in object recognition in order to have such addresses assigned automatically by Yahoo! or Earthlink servers. It is a highly efficient means of solving a problem that would otherwise require substantial computing effort, certainly more than is likely to be marshaled for a project with applications restricted to a relatively narrow audience. It represents, moreover, a quick fix that would not deflect from the goal of creating, at the collection level in cultural heritage, virtual museums with all of the technical applications that have resulted from recent IT research, including a full electronic catalogue of the museums' holdings, nor from the inter-collection level of recreating ancient contexts necessarily disturbed by both licit and illicit excavations and artifact distribution.