

**Harvard University Library
Library Digital Initiative Grant**

FINAL REPORT

Peabody Museum, Harvard University

Carnegie Institution of Washington Collection of Maya
Archaeological Photographs:
Phase 1 and 2



Submitted by David Schafer, LDI Project Supervisor, Peabody Museum

2/25/2005

Executive Summary:

The Carnegie Institution of Washington (CIW) Collection documents the CIW's archaeological research undertaken in Central America from 1929 to 1957. The Peabody received the CIW photographic collection (approximately 42,000 images) after the institution's department of archaeology closed in 1958. The Peabody never assigned unique ID numbers to the photographs, nor integrated the data into the museum's catalog card indexes or archives, nor created computer records for the images. The Peabody holds title to this collection and has copy rights to all except for a small number of images (ca. 100) which the CIW previously published.

The Peabody Museum received two sequential LDI grants (2000 and 2003) to electronically catalog the data records, digitize the photographic images, deposit the digital images to Harvard University's Digital Repository System (DRS), and allow researcher access to the data through the University's Visual Information Access database (VIA) and the Peabody's Collections Online database. The Peabody achieved and surpassed the Phase 1 goal of depositing an initial 9,850 image and achieved and surpassed the Phase 2 goal of depositing an additional 30,000 images, for a combined final total of 40,304 unique images.



Phase 1:

For the initial phase of the project, LDI granted (April 2000) the Peabody Museum funds to digitize 9850 images from the Carnegie Institution of Washington collection. Based on the paper CIW catalog card, the collection contained an estimated 40,000 images. (NB: there is a 5% disparity of number of images between the catalog records, the photographic prints and the nitrate negatives.) The first task involved using the catalog cards to create individual database records in EmbARK, the museum's collection management database. During this task, each photographic record received a unique Peabody Identification Number, a cross reference to the original catalog record, geographic and cultural data, and key word search terms. While the initial plan was to do this only for the 9850 image specified in the grant, early work indicated that these specific catalog cards were randomly distributed throughout the entire dataset. As work progressed, it became obvious that it was more cost effective to data-enter the entire set of 40,000 catalog records, rather than search ad hoc for the sub-set.

Initial work also led to the recognition that the digitizing process for the images could be effectively automated, as newly-discovered copy negatives in the museum's photographic

archives contained backup images on 1000-foot-long film reels. In close consultation with a local vendor (Boston Photo, Inc.) the museum determined that it was much more cost effective for this work to be subcontracted out rather than done at the Peabody Museum or the Harvard College Library Digital Initiative Group. Since the 9850 images selected for the initial phase of the grant were embedded randomly within the rolls of the 38,000 images, the automation of this process allowed us to digitize every image for the same budgeted cost of the original 9850. Therefore, in working through the first phase of this grant, the Peabody not only managed to complete the catalog record work noted above for the entire collection, but managed to receive digitized TIFF's of all copy-positive-film-based images. An estimated 2500 CIW images exist only on historic photograph boards (H-Boards) and were not digitized during this phase.

In addition to the catalog cards and digitized images, terms of the first grant also contained final steps wherein the museum opened each digital image, prepared it for archiving, renamed the file and linked it to the database record. Additionally, the records were data enhanced through the additional subject key-wording. The Peabody developed and began depositing the images to the DRS and VIA repositories, accompanied by the appropriate metadata and coding. At the conclusion of the first LDI grant, the Peabody completed for the target number 9850 images (the final total was 9934 images).



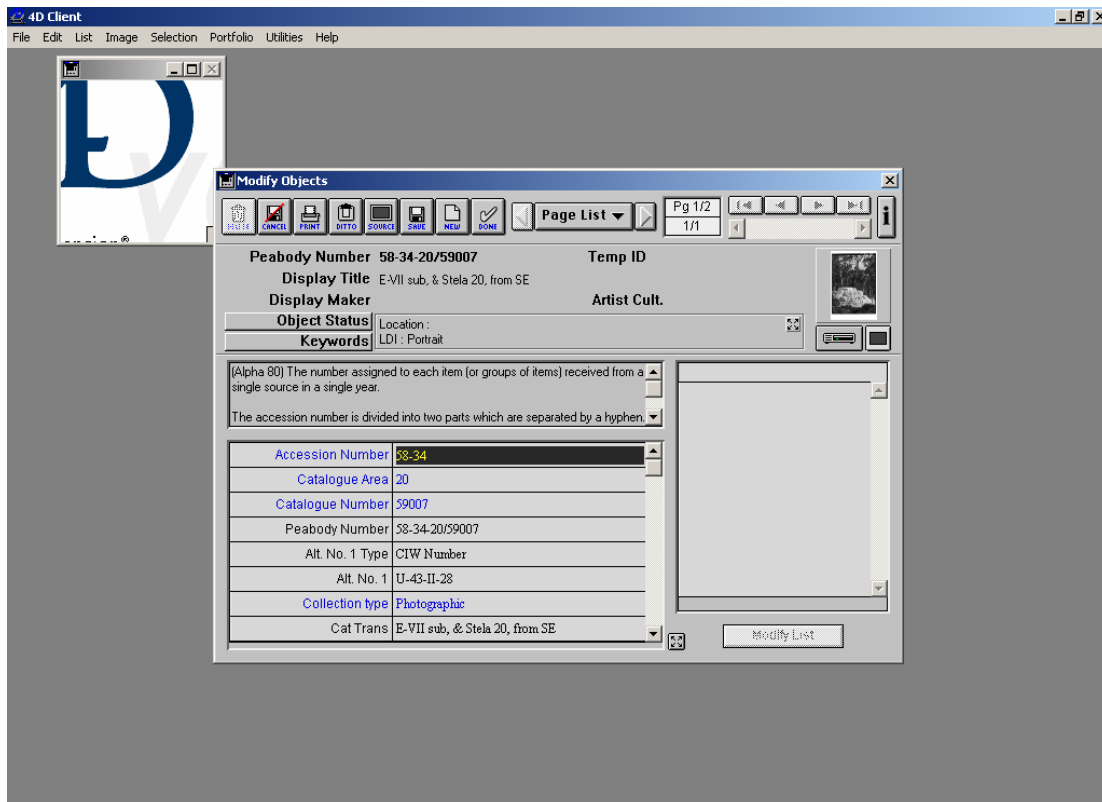
Work on Phase I allowed the Peabody to develop consistent and coherent means to proceed with this project. In addition, it trained and amplified the skills of Peabody staff associated with all aspects of the project. In developing the code and metadata for the DRS and VIA deposits, the museum created the entire procedures needed to digitize the remaining 30,000 images. Especially helpful was an automated EmbARK tool for text replacement that staff developed in-house and utilized in the generation of coded XML for deposits to the DRS. The Peabody also determined and implemented issues concerning the creation of administrative, descriptive and structural metadata with regards to this collection. The project also allowed staff to complete steps with the registration and utilization of DRS and NRS for storage and access.

The deposits made to the DRS consisted of three digital object versions: an archival TIFF master file (restricted from public access) and two JPEGs; one a VGA-sized image (usually between 30 and 75 KB) and another smaller thumbnail. Both of these JPEG

images are publicly accessible and the museum utilized owner-supplied names to allow standardized use in VIA. After deposits to the DRS, the database administrator worked with the SuperReports module of EmbARK to use these names in the generation of the coded XML for subsequent VIA deposits. After finalizing deposits to the DRS and VIA, these digital images are also available through the Peabody Museum’s Online Collection Database.

Sample pages from EmbARK, the museum’s website and VIA are shown here to demonstrate how the digital objects generated in Phase I are used:

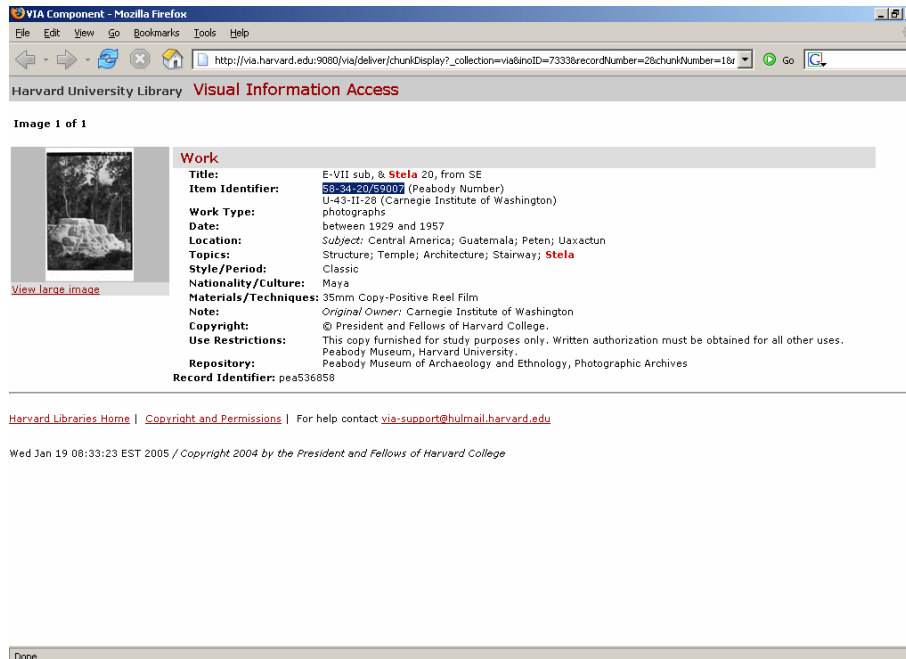
EmbARK Collections database:



Peabody Museum Collections Online (<http://www.peabody.harvard.edu/col/default.cfm>)



Sample VIA Record:



At the completion of Phase 1 of the LDI – Carnegie Institution of Washington grant, the Peabody Museum deposited 9934 archival TIFF images, a VGA-sized JPEG and a thumbnail JPEG (totally 29,802 individual image files) to the DRS. In addition, the museum contributed 9934 data records to VIA, and created over 40,000 data records in EmbARK.

Phase 1 Project Summary:

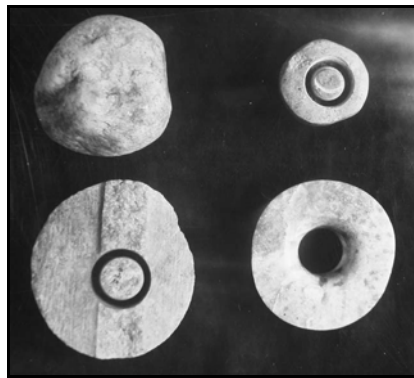
Stage 1 - Data entry: COMPLETE

Stage 2 - Digitizing of 35 mm copy positive reel film: COMPLETE

Stage 3 - Data enhancement/quality control: COMPLETE

Stage 4 - Peabody Museum Deposit Image to DRS: COMPLETE

Stage 5 - Peabody Export Data to VIA Test Server: COMPLETE



Phase 2:

In May 2003, LDI granted the Peabody Museum additional funds to process, deposit to DRS and VIA, and make available on-line the remaining 30,000 CIW images. Since the initial phase of the grant completed data entry of 100% of the images and digitized the 38,000 copy-positive film images, the second phase consisted solely of processing the images, digitally photographing the remaining sub-set of approximately 2500 H-Board images, and depositing the resulting data to DRS and VIA.

The Peabody followed the standards and practices developed during the initial LDI grant. Museum staff linked each digital image to an existing EmbARK record, verified the image, and created keyword search terms. Each image also received a unique Peabody Digital File Name. The staff also processed each image to current museum standards, which allowed the museum's Office of Information Technologies to batch-process the images into EmbARK and prepare the data for batch VIA exports and DRS deposits.

The first portion of the Phase 2 grant consisted of simply processing the remaining digital images created by Boston Photo, Inc from the copy-positive reel films. This portion of the grant was very effective and was completed quickly. During this portion, however, Peabody staff discovered that sections of one of the copy-positive reels contained

duplicate images of an earlier reel. As a result, approximately 6000 existing digital images turned out to be duplicates. In order to account for this disparity in numbers, a total of 8500 H-boards needed to be photographed instead of 2500.

Although digitally photographing this larger number was more time-consuming than simply processing extant digital images, the LDI team was able (with slight assistance from the collections management department) to achieve the target number on time and on budget.



Similar to Phase 1 of the grant, deposits made to the DRS consisted of three digital object versions: an archival TIFF master file and two JPEGs; one a VGA-sized image and another smaller thumbnail. Both of these JPEG images are publicly accessible.

During this deposit portion of the grant, Peabody staff slightly revised the EmbARK SuperReport that generates XML coding. These revisions incorporated recent changes to the data structure in VIA. Also, by working closely with DRS staff, the Peabody's Digital Image Administrator was able to prototype and design code for possible future deposits of JPEG 2000 images. The Peabody will likely use both of these functions for future projects.

At the completion of Phase 2 of the LDI – Carnegie Institution of Washington grant, the Peabody Museum deposited an additional 30,370 archival TIFF images, a VGA-sized JPEG and a thumbnail JPEG (totally 91,110 individual image files) to the DRS. Combined with the Phase 1 results, the projects produced a final total of 40,304 VIA records and archival images (totaling 120,912 individual image files).

Phase 2 Project Summary:

- Stage 1 - Catalogue card data entry: completed during Phase 1 grant.
- Stage 2 - Digitizing of 35 mm copy positive reel film: completed during Phase 1 grant.
- Stage 3 - Hire new Peabody Museum project: COMPLETED.
- Stage 4 - Image processing/quality control/data: COMPLETED.
- Stage 5 - Create deliverable: COMPLETED.

Stage 6 - Peabody Museum deposit image to DRS: COMPLETED.

Stage 7 - Peabody export data to VIA: COMPLETED.

Stage 8 - User evaluation: Initial evaluation COMPLETED; continued testing and system improvements, ON-GOING. (Evaluation to be submitted separately)



Project Budget Summary:

In Phase 1 of the project, LDI spent \$91,019 and Peabody spent \$20,439

In Phase 2 of the project, the program required 1:1 matching funds, so LDI and Peabody both spent \$67,581.

LDI Expenditures

Total for Phases 1 and 2: \$158,600.

Phase 1

\$39,019 imaging vendor

\$52,000 staffing including project management and temporary project staff

\$91,019

Phase 2

\$67,581 temporary project staff

Peabody Expenditures

Total for Phases 1 and 2: \$ 88,020

Phase 1

\$14,739 Staff

\$3,700 Equipment

\$2,000 Preservation rehousing

Phase 2

\$67,581 temporary project staff