



PHILIP S. HENCH WALTER REED YELLOW FEVER COLLECTION

The Story

The Collection

Walter Reed

Carlos J. Finlay

Jesse Lazear

Henry Rose Carter

J. R. Kean

Philip S. Hench

YF Commission

Camp Lazear

Volunteers

Informed Consent

Date

Series

Subject

Search

Books

Highlights

Who's Who

Places

Collection Guide

Site Navigation

IMLS Final Report

Related Sites

Credits

Digitizing History: The Final Report of the IMLS Philip S. Hench Walter Reed and Yellow Fever Collection Digitization Project

Phase I

Prepared by Joan Echtenkamp Klein, Project Director
and Linda M. Lisanti, Project Coordinator,
December 2001

Introduction

Philip S. Hench spent over fifteen years accumulating thousands of documents, photographs, miscellaneous printed materials, and artifacts to decipher the actual events involved in the U.S. Army Yellow Fever Commission work in Cuba at the turn of the 20th century. He eventually intended to write a book about Walter Reed and the Conquest of Yellow Fever that would resolve conflicting memories and controversy surrounding the Commission's work. This collection demonstrates the extremes to which he went to discover the true story -- the exact "who, what, when, where, and why" of it, and the persistence that he maintained in discovering that truth. While accumulating these materials, he was instrumental in memorializing Walter Reed and the Yellow Fever Commission in both Cuba and the United States, he participated in Cuban - American affairs, widely shared medical opinion, served in the Army as a Colonel during WW II, and won a Nobel Prize for his work with cortisone. He never wrote the intended book.

What these documents tell us is a story much larger than the book he would have written. As we examine the record of Hench's tireless inquiry,

we become privy to the sensitive personal and professional motivations behind scientific research, memoir, and cultural drama. The Collection records the stories of many individuals -- their thoughts and feelings, daily labors, controversies, professional activities, cultural perspectives, and personal relationships-- that are seldom written into historic text.

At the close of this two-year project we have successfully met our project goal to digitize a large selected portion of this exhaustive collection, and, we have also exceeded it. We have selected, digitized, transcribed, and analyzed 5,120 handwritten, typewritten, and printed documents including

111 newspapers and 4 maps from the collection totaling 13,007 pages. Included in the digitized documents is the complete 1906 edition of *Walter Reed and Yellow Fever*, by Howard A. Kelly, owned by Philip Hench, which contains Hench's handwritten notes as well as autographs by scientist Carlos J. Finlay's son, Charles E. Finlay, and yellow fever experiment volunteers John Kissinger, and John Moran. In addition we have selected and digitized 314 photographs and 8 artifacts. The material has been preserved in TIFF files on a total of 1,970 CDs for archival and preservation purposes. While XML is now touted as the state-of-the-art mark-up language, when we embarked on this project our use of XML was trend setting. We believe that we are among the first to mark-up this amount of material using XML, with TEI attributes.

Another goal, which was met, was to have the digitized, transcribed, and marked-up primary materials internationally available on the Web, searchable in as many ways as possible, and accessible in a seamless fashion; it was not necessary for visitors to the Web site to see all the behind-the-scenes labor involved, but everything had to work smoothly, invisibly together to make the Web site a success. To address the issue of providing as many ways of accessing the material as possible we created a two-part Web site. "The Collection," which is a searchable World Wide Web database, incorporates a digital image for each page with a corresponding transcription and a summary for each document. We believe that the corresponding "Collection Guide" may be the largest Encoded Archival Description Guide in the world.

In addition to fulfilling this primary project goal we also created a corresponding exhibition referred to as "The Story" -- an exhibition of photographs and text providing background on and context for the key characters and events occurring in the papers contained in the Collection. Also included were master lists of places and people mentioned in the Collection -- a complete "Who's Who Guide," as well as a list of Collection "Highlights" which point out some of the unusual but representative material that our staff found particularly interesting. These two parts of our Web site, "The Collection" and "The Story," were seamlessly joined for easy access to both the Collection itself and our background material. The decision to house the Reed project Web site at the University of Virginia Electronic Text Center was made to allow us to more than meet our future goal of being able to seamlessly search across collections. These major accomplishments were not met without some major challenges.

Our initial analysis of this collection led us to believe that the collection content, both in terms of topic and format, was a tidier one than we eventually discovered. After our initial review we assumed the collection contained primarily the personal correspondence and photos of the principal figures in the yellow fever investigations and additional research materials compiled by Hench, centering primarily on the field of medicine. We estimated the Collection to be half handwritten and half typewritten material. Our original focus was on volume -- our estimated 30,000 pages, and the technology involved. We were fascinated by the cutting-edge technology that we were among the first to use for a project of this kind.

We understood digitization as a mechanical process, but access and transcription were far more complex. Capturing an unfolding complexity from material spanning a time period of close to a century, with the required analysis for metadata and design, was challenging and became far more labor-intensive and time-consuming than we initially expected. Neither had we predicted the many variations in document types that emerged. In addition, letterhead images, logos, pictures, scribbled

notes, and various formats -- usually insignificant matters, took on new meaning. As meeting the project time schedule was a necessity, completing this in-depth analysis within the constraints of a rapid work pace became the overriding challenge of this project. We were forced to determine the importance of information at the very moment we were discovering it.

We had reviewed only a reasonable sampling of the material in this vast collection. As the project progressed, we realized the content -- what it was about and how it appeared -- would shape the project far more than our original idea -- and in several ways. Meeting this initial challenge concerning the importance of content and additional challenges as they arose, however, opened new areas for accomplishment that we had not anticipated.

Year One December 1999 to December 2000

In the first six months of the project, the Project Team consisted of Joan Echtenkamp Klein, Assistant Director of Historical Collections and Services, as Project Director; Nadine Ellero, Head of Intellectual Access for metadata and authority control; Aulia Gies, Associate Director of Information Services for his XML knowledge and programming expertise; Kim Guenther, Internet/Clinical Information Services Coordinator, for Web design and project delineation; Joby Topper, Historical Collections and Services Assistant, for project assistance; and David Seaman, the Director of the University of Virginia's Electronic Text Center (E-text Center), serving as project consultant.

The Team met twice a month and work was primarily involved with reviewing existing Web sites for design, search options, maneuverability, defining the mechanical tools for the project, and selecting a conversion vendor for transcription and markup of the material. The Associate Director of Information Services, Aulia Gies, and Internet/Clinical Information Services Coordinator, Kim Guenther, were instrumental during this initial stage and the Director of the University of Virginia's Electronic Text Center, David Seaman, lent his wise and valuable guidance. Sites affiliated with the University's Electronic Text Center and the University of Virginia Library Special Collections Digital Center were focused on as models. Equipment was purchased and electronic methods were researched. Our decision to use XML with TEI attributes allowed us to create richer, more flexible applications for searching and presentation and enabled us to interface with the E-text Center collections. We purchased three Macintoshes for scanning the materials, and decided to scan TIFF files at 600 dpi for preservation purposes. Servers and server software was also discussed.

Upon the recommendation of David Seaman, and, because of the cost implications of such a large volume of material, we decided to use an overseas conversion vendor. Transcribing such a large amount of handwritten text had never been taken on before to our knowledge by such vendors. Innodata's division in the Philippines was procured for the project, and a work contract was developed outlining delivery of the materials to and from the Philippines within a tight project schedule. It was decided to transport the material to them by burning the images documents onto CDs and mailing them, while return delivery of the marked-up material would be made to us by electronic transmission. Work was scheduled to begin in August of 2000 and end in September of 2001 with a work flow of between 1,000 and 2,500 pages transported to Innodata every two weeks and returned to us in the amount of 313 pages to us on a weekly basis.

For the grant proposal, we had estimated the entire volume of material at 30,000 pages by first briefly reviewing 20 of the 147 boxes in the Collection to determine original versus printed material. After arriving at specific counts in the first 7 boxes, we then arrived at the estimate of approximately 30,000 pages of 5,000-10,000 documents. All items, including envelopes and copies were included in the count.

In April, Scanning Technician Amy Pannell was hired to join the Historical Collections Assistant Joby Topper. The Project Coordinator was not yet hired but the role was being defined as contributing to production and quality control of materials returned from the conversion vendor while managing the project daily operations, supervising and training student and temporary employees, coordinating project operations with members of the Project Team, producing reports, and assisting in the implementation of online delivery of the project images and text.

Metadata and Authority Work

The Head of Intellectual Access and Project Metadata Specialist, Nadine Ellero, was involved with the selection and application of various types of metadata and their coding in the marked up documents from the onset. It was agreed that the Project Team as a whole would make final decisions. The primary goals of the Project Team were to ensure as many points of entry to the Collection as possible to provide maximum assistance to the user. Personal and geographic names, subject terms, free text or keyword searching, and specialized searching were discussed. The Team was aware of the extreme labor-intensive and time-consuming nature of the cataloging and authority work, particularly with such a large collection. Concerns for time were expressed, as meeting the project schedule was primary; however maximum facility for the scholar and researcher was desired.

It was decided that project staff would assign a metadata header and write a brief summary for each document. After lengthy discussion, the Team determined a list of approximately twenty keywords consisting of a standardized vocabulary including subject terms selected by the Team, as well as Medical Subject Headings and Library of Congress Subject Headings. The value to a researcher of being able to search the Web site by subject, and having a summary for each item, was considered of primary importance to the usefulness of the Web site.

The Team created a template or matrix in chart form to indicate the subject headings, major names and places, and a summary for each document to be filled in by project staff. This information would then be supplied to Innodata for the front-matter markup. Determining the appropriate titles and dates for the documents would be included as part of the transcription and mark-up procedure.

Lists of names and places entered into the matrixes would be compiled into working master authority lists, which would be continually reviewed and updated by the Head of Intellectual Access and referenced by staff as they completed -- what we referred to as -- the "matrix work". Various methods of name entry were debated to ensure as much standardization as possible; however, with the volume and time constraints, the simplest methods using the least amount of time were desired. With standardization foreseen as a possible challenge in this context of volume and time, it was predicted that the use of computerized programs might assist with the cleanup at a later time.

A document identification number scheme was also determined as three digits for the box number, two for the folder, and three for the pages of each document. The documents were numbered consecutively by pages within the folder with the document identification number being the page sequence in which it was placed in the folder.

Scanning and matrix work was undertaken by the project staff; however, from May to August the project shifted emphasis because of the permanent absence of four members of the project: the Associate Director of Information Services, the Internet/Clinical Information Services Coordinator, the Historical Collections and Services Assistant, and the Scanning Technician -- all who left the project because of employment and education opportunities outside of the Health Sciences Library. In July 2000 Hal Sharp filled the Historical Collections Assistant position, and Mollie Donohue replaced our Scanning Technician, Amy Pannell. In August a part-time Historical Collections Specialist, Ina Hofland, joined the office team. (We were unaware at that time, that it would not be until November of 2000 that the Internet/Clinical Information Services Coordinator position would be filled by Bart Ragon who joined the project in February 2001, and that the Associate Director of Information Services position, renamed the Information Technology Services Manager, would be filled in April of 2001 by Anthony Head who would join the project shortly after.)

With the daily project work beginning, the challenge of preparing, analyzing, and digitizing the estimated 30,000 pages within our work schedule grew larger. The first shipment of scanned images and corresponding matrixes to Innodata was scheduled for August, but was not mailed out until the last day of the month. It was clear that the analysis for the metadata -- reading and determining the subject headings, key names and places, writing a summary for each document, and entering this information in the matrix template-- was even more labor-intensive and time-consuming than anticipated. The Historical Collections Assistant Hal Sharp had solely and persistently completed the matrix work for the initial 995 documents, consisting of Walter Reed's letters, and it was predicted that even with the assistance of the newly hired part-time Collection Specialist (Ina Hofland) more staff might be necessary.

In September of 2000, we hired a Project Coordinator, Linda Lisanti, and shortly after, added two additional in-kind, part-time Historical Collections staff as Collection Specialists, Janet Pearson and Susan Swasta, for the matrix work. By December, after reexamining the project with the Project Team, the work was redefined into three basic overlapping phases: Pre-transcription, Transcription, and Post-Transcription, with the emerging challenges that had evolved in each. While these phases implied sequence in terms of workflow, they were dealt with concurrently as decisions in one phase affected the others.

Pre-transcription

In the initial months of the project, the pre-transcription work of scanning, filling out matrixes, and mailing the CDs to Innodata had been well defined by the Project Team. With our reanalysis we broke it down even more:

Each manuscript was reviewed for quality, legibility, relevancy, visual interest, and selected for inclusion. Both the Historical Collections Assistant and the Scanning Technician were involved with

the unique requirements of selection.

The Historical Collections Assistant corrected Collection processing errors discovered along the way. The Collection, originally processed in 1964, contained a large number of incorrectly filed documents that we discovered as we selected the material and assigned identification numbers. Reprocessing the Collection prior to review and selection was a lengthy and at times frustrating process; however, it led us to examine the material more closely and, as result, the Historical Collections Assistant eventually created an additional series (the Kean Series) for the online collection.

Photocopies, carbon copies, photographs, printed, oversized, and miscellaneous material were noted on a "skipped list," maintained by the Scanning Technician and an identification number was assigned to each selected document and entered on a corresponding empty matrix.

After reading and analyzing each manuscript, the staff filled in the matrixes, noting the key subject terms, major names and places, and wrote a brief-phrased summary for each document. Online authority lists of names and places were maintained by the Head of Intellectual Access and were referenced for consistency.

By December 2000, we had generated over 600 names in our working list. When we created the initial subject terms and the beginning authority file of names and places, we initially thought that the subjects, names, and places found in the letters of Walter Reed would be primary to the Collection as a whole; however, while that original assumption was partially true, we found that the material was far more complex. We were generating more names far faster than the planned 10% project time available to the Head of Intellectual Access that would allow for the authority list to be "ahead" of the analysts. In addition, challenges emerged regarding methods of standardized name entry.

Optimum form in name entry was changeable with various permutations of the material. After much discussion on the part of the Project Team as a whole, it was decided that meeting our time schedule with Innodata was primary for the project's timely completion, and that full name refinement could be accomplished later. We wanted to apply, when possible, library standards to our online collection, but with the concern for meeting the project deadline, the Project Director decided that we would attempt to limit the authority file to around 100 to 200 major names in Library of Congress form. We anticipated, however, that as we proceeded deeper into the project we might revise our plan. (A possible two-month extension of the project was discussed at this time because we were aware of the time that could be involved with name refinement as a result of our decisions.)

After a full collection box of manuscripts was analyzed and corresponding matrixes completed, each page was scanned. Between 6 and 7 TIFF images of scanned manuscripts were burned onto a CD, labeled with appropriate identification and a mosquito logo, and stored in binders for archival and preservation purposes. When a full shipment amount was completed, JPEGs were then burned along with the corresponding matrixes onto a single CD and mailed to Innodata. Our Scanning Technician Mollie Donohue operated three scanners concurrently, batched and burned the images onto CDs for both archival and transcription purposes, maintained the "skipped list," and tracked the documents.

Transcription

To our knowledge this quantity of handwritten, nineteenth-century material had never been transcribed and marked up by anyone else to date. In developing our original work statement with Innodata, the Project Team outlined basic guidelines to be followed by Innodata's keyboarders with the expectation that as work was underway, additional guidelines would be added. As we proceeded, we examined with Innodata the nuances of Walter Reed's hand, in addition to unusual military and government documentation and telegrams. Under the guidance of David Seaman, a specialized set of general rules was developed to guide Innodata transcribers in both transcription and markup. Transcriptions were thought of as a guide to reading the documents rather than a replica.

Responding to unique daily queries from Innodata on how to capture the material with the appropriate tag sets was challenging for us and took more time than we originally anticipated. We did not originally anticipate that the Project Team as a whole would require coding expertise, because it was anticipated that the Library computer and systems staff would have this knowledge and that training would be provided to the Project Team as the project progressed. Our consuming workload left us little time for our self-training and David Seaman took on a greater role in advising us. This process educated all of us, requiring that we examine the material in more depth, and fostered a closer working relationship with our overseas partner that was a delight to all participants.

Many of Innodata's initial daily queries were concerned with deciphering Walter Reed's handwriting. Specific items such as casual capitalization, double dashes, crossed out words, various forms of underlining, superscripts, and unusual postscripts written on the edges of pages or at the beginning of the letters were questioned for the transcription. Long pen strokes at the end of words made it difficult to determine spacing and dashes. Other questions concerned capturing symbols such as \$, #, or @, geographic locations, and nineteenth-century abbreviations for military ranks. Made up words and endearments that Reed used when writing to his wife required particular focus. Official documentation presented other distinctive elements. Letterheads, rubber-stamps, form layouts, multiple signatures, blank lines, and charts all required special attention.

After two-and-a-half months of daily queries, analysis of two multi-manuscript samples, and a teleconference, in mid-November Innodata transmitted the final mark-up of our initial shipment in late August. We were impressed with their work and their genuine concerned attention to accuracy. Letters that were difficult for us to decipher were transcribed well, with only the expected one to two error average per page and the appropriate words tagged as unclear. Rarely, but at times, cultural difference in language, grammar, and word association appeared to influence deciphering words.

Having passed this critical stage, we expected to receive a faster return on our shipments; however, the unique challenges involved in accurate deciphering of the handwritten material, along with the unanticipated time involved in the pre-transcription workload on our end, caused us to rethink our future quality control needs and selection priorities.

Innodata drew up a revised delivery schedule ending in October of 2001, a month later than originally planned, which we readily agreed to. While at this juncture we could not predict with exact certainty the actual number of total pages to be transcribed, we expected that our total might be under the 30,000 initially contracted for, but as we moved forward into the Collection, it also appeared that there would be several more handwritten authors than Walter Reed and varied formats for typed and

printed materials we had not predicted. Continual attentiveness to these crucial aspects was employed, and documents were selected with Innodata's challenge of deciphering difficult handwriting in mind. The time involved for both our pre-transcription work and their challenging task of transcription and mark-up was a key factor for consideration.

Post-transcription

While it was originally anticipated that the Project Coordinator would be solely responsible for quality control of materials returned from Innodata, it soon became evident that this responsibility was more labor-intensive than anticipated both because of the sheer volume and the unique nature of the work. Additional staffing would be required. Both markup and transcription required careful proofreading and editing because of the challenges involved with deciphering handwritten text. Reed's own made-up words, retaining original spelling errors, and the necessity of deciphering words by reading in context, as well as the emerging variety of documents, all presented reasons for a more in-depth accuracy check. In addition, it was being anticipated that the handwriting of approximately twenty to thirty people -- each with unique handwriting and grammar styles -- would be deciphered.

By our first returned shipment in mid-November, we established through the E-text Center a working online site with over 700 transcribed and marked-up documents to be viewed and ready for editing. We initiated a general quality control review and the following were our initial observations of corrections that would need to be made outside of the overall text editing, and checking the basic classes of tags:

Joined Documents or "Multiples":

Letters with enclosures, and in some cases, multiple form letters were included in a single file separated by divisions to be divided and determined for linking later. This was due to our grouping like or related documents in a single matrix in the first CD. We decided to put these on hold until a later time when we could review them again to determine appropriate linking.

Words designated as unclear:

As previously stated, we were impressed with Innodata's skill in deciphering handwriting, particularly Reed's, but certain words were genuinely unclear to read. As advised by David Seaman, a liberal use of unclear tagging had been encouraged to speed up the return. We were impressed with Innodata's "guesses" for the unclear words; however, "unclears" would be checked by us later, and we expected a natural amount of spelling error. It was predicted that unclear and misspelled words could be identified through the use of computer programs; however, spelling errors original to the documents were to be retained and these would be need to be checked. We knew more authors would provide greater challenges.

Titles and Dates:

While the Team had originally determined that Innodata transcribers would enter titles and dates from letter salutations and signatures with editing by us later, we suspected that, particularly without

supplied information and the change in document types later on, this might prove too challenging and was an area on which to keep our attention.

Misinterpretations:

Infrequently, an occasional word was not unclear it was spelled correctly, but incorrectly deciphered. For example in a phrase such as "he wanted to live," the word "live" would be transcribed as "line." In checking the original, the word would, in fact, look like "line." We understood that the customary nature of Innodata's keyboarding work did not usually require reading in context, as most of the material they dealt with was in type. We also did not want to compromise the speed of return. While only a few of these misinterpretations were in Reed's letters, we suspected they might increase with the changes in authors, and would require our attention.

We decided to hire a graduate student in January who would be experienced in both text-editing and markup and trained at the E-text Center for extensive and continual quality control.

Rethinking Collection Priorities

By the end of October, we had completed pre-transcription work for 17 boxes and scanned an additional 15 boxes from the Reed/ Hench Series, but we had not yet received the full return from Innodata and were concerned about the overall project schedule. We mailed a second CD to Innodata with 1,294 images taking us past Walter Reed's death. The documents had shifted from Reed's handwritten letters to a variety of typed and printed documents in the later Reed and Hench Series. We had gained a greater familiarity with the material from analyzing the manuscripts for the metadata and authority work. While before, along with a general content breakdown, our concentration was focused primarily on page quantity, handwritten, typewritten, and graphic materials, we now thought more about the people and events represented in the Collection. We decided to redefine project parameters in terms of content. If we had to reduce the volume of digitized material, what parts of the Collection would be central? The guidance of Project Director, Joan Echtenkamp Klein, proved especially helpful during this stage.

Walter Reed died in 1902. His final letters were in the second CD mailed to Innodata. While we scanned material in the Collection that follows Reed's death, we chose not to send it to Innodata. The items scanned, but not mailed, were material that Philip Hench collected for a potential biography of Reed. They account for the events up to and past the memorializing of Reed and his colleagues at Camp Lazear in Cuba in 1952. We decided that Philip Hench's overall story as a biographer served to frame the primary manuscripts of Walter Reed and his colleagues Jesse Lazear, Henry Rose Carter, Jefferson Randolph Kean, and others who were part of the yellow fever story, and therefore the handwritten material would be primary. We decided to supply Innodata with the primary handwritten materials first, because it was clear that these materials would present the greater challenge. Large quantities of Hench's handwritten notes and particularly difficult-to-decipher handwriting would be eliminated. After the primary manuscripts were transcribed, we would return to Hench's biographical collection up through the dedication at Camp Lazear.

We also decided that letters of immediate family members would be included because of their

influence, and to shed light on the story of the families' struggle to obtain pension funds. Attentiveness to the Cuban perspective of the yellow fever story and the related controversy was also agreed upon as an important element of the Collection.

As a result of these decisions in October, despite the fact that we had scanned a large amount of the Hench material, we switched to the Jesse Lazear Series. We – as well as Innodata -- were delighted to discover that a complete typewritten transcription of his handwritten letters was included in the Collection. After completing that series we moved on to the Henry Rose Carter and Jefferson Randolph Kean Series. While we did not anticipate having to eliminate any significant material from the Collection, creating these priorities helped to shape and guide our work.

Year Two

January to November 2001

In January of 2001, we had a significant amount of material – 1,928 transcribed and marked-up pages -- on our working online site. While we expected to have a trained graduate student assist us with the editing and quality control, the appropriate person with the desired knowledge was not found. While we continued our search, the Project Coordinator continued an analysis of the returned material, while continuing with Innodata's daily queries on specialized treatment for the various documents types, supervising the matrix work and scanning, monitoring the shipment schedule to Innodata, and meeting with Project Team members for key decisions.

Scheduling with Innodata was reoriented. While we had a new schedule moving everything forward to end a month later than originally planned, neither Innodata nor we could keep a strict adherence. Each series had unique qualities that needed to be accounted for. The schedule was used as a general framework, but the primary goal was to be certain we always had enough prepared and shipped to them so that there would be no gaps in the work, and there were none. Their transmissions to us were more frequent but not as regular as originally planned. Both Innodata and we began thinking of the work in terms of boxes that were completed in each series rather than pages, and shipments to them were smaller (from approximately 300 to 900 pages) and averaged twice a month, but not always spaced at regular intervals. It also appeared that they were reading the hand written materials more in context, as there were proportionally less misinterpretations. Henry Rose Carter's papers were a challenge to both of us, and we applauded their work with the difficult handwriting it contained.

As a result of shifting our pre-transcription work from the Hench Series to the Lazear Series, we shifted our workflow sequencing. Whereas earlier, scanning was done prior to the matrix work, allowing the Scanning Technician to be involved in the selection review and matrix identification process, now the matrixes were prepared by the Historical Collections Assistant while he was reviewing and reprocessing the Collection prior to the scanning. This work shift required some adapting. Whereas before, we could view the actual JPEGs as a check while we were doing the matrix work, now we no longer had that reference to double check JPEG and matrix identification numbers.

In early February, the Project Director, the Project Coordinator, and the Head of Intellectual Access met to address some of challenges that had been identified. We had decided to begin composing the document titles and dates and entering them in the matrixes to lift that responsibility from Innodata.

The Head of Intellectual Access identified the Alexander Graham Bell Papers at the Library of Congress American Memory Web site as a model for titles. We also decided to maintain lists of different document formats that were emerging such as reports, notes, articles, questionnaires, certificates, programs, Congressional bills, etc., so that conventions for these types could be refined in the editing process. We also decided to track letters with enclosures for linking.

In the matrix work, determining subject headings was becoming more challenging as the document types changed and summaries were becoming increasingly complex. For the Reed letters, a brief-phrased abstract was sufficient. As the documents grew more complex, particularly in the Carter Series, each summary became a complicated list of phrases divided by semi-colons that were difficult to understand out of context. The Project Coordinator identified the online Florence Nightingale Papers from the Clendenen Library at the University of Kansas as a model for the summaries, and we shifted to writing the summaries in complete sentences, rather than phrases.

Matrix entry guidelines were drawn up for titles, dates, abstracts, and subject headings. We had also begun to encounter documents in Spanish and French, both typed and handwritten. Collection analysts were asked to list foreign-language documents and letters with enclosures, in addition to the names and places lists, and, another list that had naturally evolved -- "particularly interesting material."

The Collection Specialists were analyzing between 12 and 90 documents per day. The unknown amount and length of the documents and flexible part-time work status of the Specialists made the total time involved completing this work difficult to predict. By February we had completed analysis of the Reed and Lazear Series, and were completing the Carter Series. The names list was approaching over 800 names. A strong, passionate, deeply committed office atmosphere was developing. Not only were we all consumed with the task of keeping a timely workflow in coordination with Innodata, we were becoming increasingly involved with what we had been reading, and spontaneously shared our reactions to the more compelling material as it emerged. Much of it sparked lively discussion, and it was decided to note this material as a resource for the Web site's "Highlights."

Impact and Design

Despite the Collection Specialists' varying schedules, in February the Project Team met with them to share what they were discovering as they were analyzing the documents, and to brainstorm ideas and treatment for the Web site design that may have not been previously considered. The information obtained from these initial meetings stimulated and focused our design choices, and assisted us in establishing our evolving criteria for selection. We also reviewed instructions for matrix entry and tried to work out the challenges of determining the appropriate subject headings, names entry, and writing concise summaries. We reviewed major themes and sub-themes that were emerging in each series.

We tried to identify the major figures, but all agreed that this task was particularly challenging, especially because of the varied content of each series and the time span. Someone who would be determined as insignificant early on might become a major figure later. Reed might mention friends or workers so frequently in his letters that we might assume their importance to the Collection only to

determine that they were not, indeed, primary figures because they lost relevance to the larger story. How were we to determine which people mentioned in the Collection would be important to give our attention to, assuming that some in our Web site audience may happen to be family descendants of people mentioned in the Collection, or casual browsers from a wide array of backgrounds, or our predicted researchers and scholars? Everyone mentioned in this complex woven story was in some way important, and this had been made clear as we analyzed the material

We reviewed our reactions to events made more immediate and arresting by the Collection. The death of Jesse Lazear proved particularly poignant. When examining his fever chart, the Collection Specialist working with it began to shed tears as she suddenly realized that he had died. We admired Henry Rose Carter's humanitarian and international perspective and his emergence in our eyes as a true Renaissance man. Although his papers were the most challenging to analyze, he emerged as an office hero. The women in the Collection -- the wives and daughters -- some of who would later become scholars, fascinated us, and we sympathized with their family trials. A matter of continued interest was the Cuban controversy surrounding the memorialization of Walter Reed. Was Carlos Finlay duly recognized? The first yellow fever research volunteers were also discussed as deserving great honor and respect, though not all were frequently mentioned in the Collection. We all wondered about the present-day status of Camp Lazear in Cuba. The relevance to current events could not be ignored as we referred to recent issues of epidemic disease and bioterrorism in different parts of the world as well as here in the United States -- West Nile fever, HIV, the Ebola virus, the international research of American pharmaceutical companies conducted in developing countries, anthrax, and smallpox.

At later Project Team meetings, Team members decided to divide the site into two distinct but related divisions: "The Collection," a massive database with images of the original documents and their transcriptions, and "The Story," a Web exhibit with historical background narratives linked to relevant materials and providing context for the original materials. Having this two-part division would also permit visitors to the site more options in how they wished to use it at various times; they could go into great depth searching the primary documents or opt for less original material and more interpretive data.

We decided in the discussions concerning Web design for the site to use a color oil sketch done for the painting "The Conquest of Yellow Fever" by Dean Cornwell from the Collection. Cornwell's sketch was chosen as a primary visual for the home page both for its artistic impact and because it had served as a catalyst for discussion in the office and among Project Team members of larger ideas raised by the Collection. Several versions of this sketch were done: one had Carlos Finlay prominently featured, the one selected has Walter Reed center stage. A palette for color choices in the design was agreed upon. Walter Reed's Congressional Medal of Honor and representative fever charts were also chosen for both their relevance to the Collection and for their aesthetic possibilities.

David Seaman provided suggestions in the working online site for layout, design, and search capacities, which were evaluated by the Project Team. Items such as searches by name, date, and topic, JPEG default size, lineation of text, highlighted words, unclear words, added and deleted words in manuscript text, and index display were discussed. It was also decided to maintain the Collection series and accession numbers as part of the Web site design to assist with browsing, research, and finding aids. Anthony Head, Information Technology Systems Manager, helped to clarify our ideas

concerning Web design and was instrumental in bringing them to fruition.

Cross-reference linking was agreed upon for letters with enclosures and foreign-language documents that had English translations.

Library Webmaster, Bart Ragon, who had recently joined the project, designed a splash page and background images for the site; created a site logo, an image map of the Dean Cornwell painting used on the home page, and rollover buttons; modified and enhanced photographs used in the design; and created a Web site banner. Later he worked with David Seaman to develop a "look": complementary, seamless page styles for both "The Collection" and "The Story."

Metadata and authority control

The Head of Intellectual Access, continued compiling and researching a master list of personal names which was increasing to well over 900 personal names. The list had been researched using the Library of Congress names authority file, OCLC, and information obtained from searching the Web.

While it had previously been decided that only 100-200 personal names of major figures would be identified and given full attention in Library of Congress form, with work on additional personal names to be continued as an addendum, it was difficult to determine the 100- 200 major figures. Looking at the growing list was such an interest and delight that the Project Team concluded that the full listing would be displayed on the Web site. Although this part of the Project was not envisioned in the original proposal, we considered it an exciting and valuable enhancement that reflected our own reactions and involvement with the people in the Collection. The list would be translated into a "Who's Who Guide" to names on the site, as well as provide standardization for names in titles and later metadata refinement

We decided to create a Sender/Recipient search, which we anticipated would be added to the Web site after further refinement of the names authority list. The Project Coordinator began to work with the Collection Specialists to create a list of senders and recipients with their corresponding correspondence to assist David Seaman in creating the Sender/Recipient search. Concurrently, editing was conducted to standardize name forms in authors and recipients using the most recently updated personal names list. The Sender/Recipient search was the final search element added to the Web site.

We were delighted to discover that, as a result of our decision to include a summary for each of the documents, when a "Collection Guide" was compiled using Encoded Archival Description (EAD) guidelines, our site yields what may be the largest EAD guide worldwide.

By June of 2001 and 17 CDs later, we successfully completed the pre-transcription work for the Web site manuscripts and printed materials. We selected, analyzed, and scanned handwritten and typed correspondence of the key figures in the Yellow Fever Commission and related correspondents; miscellaneous printed materials of a wide variety of formats; out-of-copyright newspaper clippings; foreign-language documents; maps; and the Howard A. Kelly biography of Walter Reed. Library Circulation Assistant, Will Brierre, assisted our Scanning Technician and completed scanning for the

full Kelly biography. Collection Specialist Ina Hofland was particularly helpful in completing all the newspaper and photo matrices.

The Kelly biography presented some challenges. Initially we had wanted to transcribe Hench's handwritten notes in the margins of the pages; however, they were particularly difficult for us to decipher. In addition, placement of these notes in the transcription was difficult to determine, as transcription formatting had purposely never been emphasized as a priority in the online Collection. Transcriptions were thought of as a guide to reading the documents rather than a replica. The Project Team decided not to have these notes transcribed, particularly in light of time constraints. Instead emphasis was put on linking the page numbers in the contents and index to their appropriate pages.

We began selection analysis and scanning of photographs, and identified artifacts and oversized documents to be digitally photographed. Collection Specialists Janet Pearson and Susan Swasta, in conjunction with the Project Coordinator, began analyzing the letters with enclosures and we developed specialized instructions to be added to the document matrixes for cross-reference linking. This was detailed and intensive work, as we divided the letters into various categories of enclosures and entered instructions for the cross-reference linking to the appropriate matrixes.

In July, the Project Coordinator and the Head of Intellectual Access attended David Seaman's course in Electronic Text at the Rare Book School at the University of Virginia. This training was key as we moved into the final stages of the project.

Editing and Quality Control

In March, we had begun to familiarize ourselves with the Notetab editing program, and had begun some basic editing in the initially marked-up material. We had close to 1800 XML files of the original material sent in the first CD with insufficient titles, dates, and the brief-phrased summaries we had originally decided on.

In April we hired a graduate student in the History Department, Michael Alexander, as a full-time Project Editor through the summer. Because of his fluency in several languages we quickly put him to work analyzing and completing matrixes for the Spanish and French documents. The Historical Collections Assistant had already begun rewriting summaries into complete sentences for the 995 that he originally wrote. The Project Editor assisted with these and then proceeded to diligently edit the remaining 4,000 XML files before he returned to his studies in late August. Having a single full-time editor was an essential factor in standardizing titles, summaries, and dates both in the text and mark-up. Summaries were particularly challenging. Many were detailed and lengthy which provided us with an opportunity to create thoughtful refinement.

Editing such a large volume of material was a daunting task particularly in consideration of our small staff and number of workstations. The transcriptions and index needed to be checked as well as the JPEG order and match up. We began with the transcriptions of the handwritten material. In July, Library Circulation Manager, Susan Yowell, joined the Reed project to assist in editing the misinterpretations and unclear words primarily on the handwritten letters, comparing the transcription to the originals, noting misinterpreted words, and changing or removing unclear guesses

and tags. We hired an additional part-time graduate student, Stephen Bell, as an Assistant Editor who joined with the Collection Specialists in completing the task in August.

We then began an overall site review. Part-time Collection Assistant, Sara Huyser, and Library School Intern, Alison White, joined the office team, as well as Library Circulation Assistant, Jay Nottingham. The site was reviewed by six to nine reviewers for basic copy-editing, standardization in titles and dates, missing and out-of-order JPEGs, joined documents, remaining unclear words, and any additional errors that were previously missed. As transmissions were still coming in, this was done in various stages, often repeating the process in a sifting manner as missed errors were discovered in previously checked files. Using Notetab, corrections were made in the XML files; the files were parsed and returned to the site in stages coordinated with David Seaman. The Project Coordinator monitored the editing process, checked corrected files, and worked with special challenges as they arose, while continuing work with Innodata for the material that was still in the transcription process, and supervised the remaining scanning and analysis.

Special problems in the XML files were referred to our Scanning Technician Mollie Donohue for corrections; Mollie was also tracking the returned material from Innodata, and scanning some of the larger items in sections. The Historical Collections Assistant Hal Sharp also scanned some of the larger items and, with the Library Webmaster, Bart Ragon, photographed those too large to fit on the scanner. Each time the material was loaded onto the E-text Center server parsing errors and some code needed to be changed, particularly in the date fields.

While at the beginning of the project timely return of the documents was a concern, Innodata's work had sped up to such a degree, that we actually thought we might have everything returned in September. We had anticipated at the beginning of the project that more time would be needed after the final transmission for the remaining editing, and particularly coordination in the design between the Library and the E-text Center. We received the final transmission in October on schedule as planned. We were able to work together in such a way that every requirement we requested was met. The work done by Innodata's project team in the Philippines cannot be commended enough. We particularly appreciated the genuine interest and support of Liza Velasco, Innodata's liaison with us for this project.

The Project Director, Joan Echtenkamp Klein, had been working closely with the Webmaster, Bart Ragon, and the Historical Collections Assistant, Hal Sharp, since the summer to create the exhibition for "The Story" part of the site. The Historical Collections Assistant wrote scholarly, elegant, insightful narratives for the main people, events, and places in the yellow fever story. He selected illustrations and primary documents that complemented the narratives. "The Story" part of the Web site is a format we were used to creating in Historical Collections and Services, as we have many Web exhibits already up on the Web; these online exhibits are heavily used, as statistics from the University of Virginia Health System Web Center attest. The Library Webmaster provided invaluable assistance with the aesthetic presentation of "The Story." He also worked closely with the Historical Collections Assistant to identify and establish hyperlinks that provide ease of navigation between "The Collection" and "The Story."

While editing and special problems -- such as missing and oversized JPEGs -- were being addressed and the narratives for "The Story" were being written, the Library Webmaster (Bart Ragon) and David

Seaman worked together to join the two sites, "The Collection" and "The Story," for overall continuity. The Library Webmaster had designed the basic template for the site and created the graphics for navigation. The exhibition text written by the Historical Collections Assistant, graphics, and photographs for "The Story" were assembled and provided to David Seaman at the E-text Center to incorporate into "The Collection." Continuing coordination with the E-text Center worked out various adjustments to the final design for the entire Web site.

By the end of November the work for the grant was successfully completed. The site was opened to the Library and selected reviewers in December. A "Grand Opening" for the site and a presentation about the project was planned as part of the 2001/2002 History of the Health Sciences Lecture Series in January.

We have already begun publicizing the IMLS Reed project and the Web site at events such as Walter Reed's 150th birthday celebration in Gloucester, Virginia and a talk given for the Deans of the University of Virginia School of Medicine. A press release will be sent formally announcing the availability of the site. Proposals for papers and sessions at conferences such as the Society of American Archivists, the Medical Library Association, and the Mid-Atlantic Regional Archivists Conference have been submitted. Papers to be submitted for publication will be written, as editors of journals such as the *Journal of Archival Organization* have expressed interest. We will also disseminate the information to appropriate listservs and newsletters.

Lessons Learned and Recommendations for Similar Projects

Our staff and Project Team learned a tremendous amount as we thoroughly scrutinized, analyzed, and selected the more than 5,500 original documents included in the Philip S. Hench Walter Reed Yellow Fever Collection Web site. The primary lesson we learned is that in preparing electronic data, we could not bypass our own in-depth analysis of the material. Time and labor involved for a project of this magnitude cannot be underestimated. While in the project grant we anticipated the number of total digitized pages to be 30,000, they were actually between 12,000 and 13,000, and the total number of documents was over 5,500. While photographs and maps were anticipated to be 1,000, they were significantly less at just over 300. Only 9 artifacts were used rather than the 100 estimated.

While the total volume was significantly less than our original estimate, work and staffing was stepped up in all areas. By the end of the project office staffing had increased from the originally estimated three people to a total of 14 people over time. Original work-time percentage estimates required for Project Team members increased from 10% to as much as 75%, and Innodata added 12 people to their original 15 member team in the Philippines. As a result, it is clear to us that while we originally overestimated the amount of material included in this collection, we dramatically underestimated the amount of labor and time needed to treat each individual document. This, as has been stated elsewhere in this final report, also led to a reprioritization of our emphases to focus on including extensive metadata at the item level.

We would suggest allowing three to six months solely for reviewing the material for document analysis that would later be applicable for the appropriate metadata entries, conventions, mark-up, and design.

We would strongly suggest providing for permanent, full-time staffing for the project and planning on a greater amount of in-kind time -- from 50%-75% -- for key Team members. We would suggest hiring the Project Coordinator during the first year to assist in planning work and overall project cohesiveness.

We would provide training in XML mark-up and creating electronic text for the full team and staff.

We would encourage the assumption that the intended audience could very well be the curious layperson from any background as well as the serious scholar and researcher. Many of our decisions were influenced by this consideration as we went along, based in large part on how involved our office staff became with the material, the story, and the people as they worked with the documents. They became enthralled by all the details, including the scientific aspects of the experiments, and the complexity of people's lives revealed in the Collection.

One of the best lessons we learned -- and which we would highly recommend to others undertaking a similar project -- was to take full advantage of our Library's Intranet, KnowledgeWeb (Kweb), as a vehicle for sharing information about the project at every stage. We started by putting a copy of our grant proposal and letters of support on Kweb, together with some sample documents from the Collection. We put up all the minutes of our various meetings concerning the project, memos written to project staff by the Project Coordinator, copies of our Interim Reports to IMLS, and copies of talks given by the Project Director about the project. The "Who's Who" name authority and Places authority lists were works-in-progress, kept up to date on Kweb. The Intranet linked all of us together and kept us informed of where we had been, where we were, and where we were going. This synergy was especially important with so many part-time staff members working on the Project.

Outcome and Evaluation

We will be able to track hits on the IMLS Reed Project Web site. We have experience in tracking hits on the Health Sciences Library Historical Collections and Services Web sites and have statistical evidence to prove that Historical Collections and Services online exhibits are consistently the most accessed of all pages available in the entire University of Virginia Health System Web. We expect this site will be actively hit-upon as well. This site has the added advantage of being available as part of the University of Virginia Electronic Text Center as well, which will increase its accessibility and visibility.

We will incorporate suggestions made by Library staff and selected other reviewers during the "soft" opening phase of the site.

We will invite comments from visitors to the site.

Conclusion

As we worked with the material it became increasingly clear that Hensch's efforts to research the life of Walter Reed and discover the true story of the Yellow Fever Commission's work in turn-of-the-twentieth-century Cuba, touched on events that impact American and global culture today. We are

proud to present this online Collection to the public in hopes that the many others will discover its hidden treasures with a fascination similar to ours and stimulate the lively discussion so essential to our contemporary world.

Statistics

of TIFFs and JPEGs created (including documents – 12,286; Kelly biography – 318; Wood manuscript – 279; artifacts – 8; newspapers – 124; and photos -- 314): 13,329

of Documents Digitized (including documents – 5,120; Kelly biography – 1; Wood manuscript – 1; artifacts – 8; newspapers – 111; photos – 316; maps - 4): 5,562

of CDs with TIFF images: 1,970

Technology Data

Scanning and burning CDs

Software used:

- Photoshop
- Toast
- Sony Discribe PPC

Hardware used:

- Macs (3)
- Epson Scanner (3)
- HP Legal size scanner (1)
- PlexWriters internal 12x CD RW (2)
- Sony external 10x CD RW (1)
- CD Label Printer (1)

Documents were scanned at:

- Resolution: 600 dpi
- Document type: Color Photo
- Unsharp mask: on
- Scale: 100%

CDs were burned as ISO-9660 so as to be readable on both PC and MAC.

Images were batched to create JPEGs to be used on the Internet.

Open first image

Set image size to:

Width: 650 pixels

Resolution: 150 per inch

With constrain Proportion

Interpolation: bicubic

Save image:

As: JPEG

Quality: 8

Matte: none

In: "specify file to save in"

With Lower Case

Close

Images were moved to server from the Macs using Fetch.

We used a CD printer to print labels on the CD to identify what was on each.

Images too large to fit on scanner were done in pieces and reconnected in Photoshop.

Photographs that were too dark were brightened in Photoshop

"Who's Who" Name Authority File Information

The names on the rough lists given to the Head of Intellectual Access were researched using the original documents, the Library of Congress Name Authority File, and the World Wide Web to obtain the full name whenever possible. The World Wide Web became an invaluable resource for finding people and events. The mostly commonly used Web sites were:

For People:

Surgeon Generals of the United States Army: <http://www.armymedicine.army.mil/history/tsgs/default.htm>

Surgeon Generals of the Public Health Service: <http://www.nih.gov/about/almanac/historical-data/surgeons.html>

United States - State Senators: <http://www.senate.gov/search/index.html>

United States Congressmen: <http://bioguide.congress.gov/biosearch/biosearch.asp>

United States Secretaries of the Treasury: <http://www.treas.gov/Architext/AT-allquery.html>

United States Department of Interior's Secretary of the Interior: <http://www.doi.gov/anniversary/secretaries.html>

U.S. Secretaries of States: http://www.state.gov/www/about_state/history/sectravels2.html#tenure

U.S. Party Leaders in Congress: <http://www.house.gov/rules/97-136.htm>

U.S. Presidents: <http://www.americanpresidents.org/> and <http://www.whitehouse.gov/WH/glimpse/presidents/html/presidents.html> and <http://lcweb2.loc.gov/ammem/ndlpedu/features/pres/preslist.html>

Principle Officers of the Department of State: http://www.state.gov/www/about_state/history/officers.html

Famous West Point Graduates: <http://www.dmi.usma.edu/Milresources/Generals/famgrads.htm>

"Political Graveyard": <http://politicalgraveyard.com/index.html>

Biographies of the U.S. Chiefs of the Army Corps of Engineers: <http://www.hq.usace.army.mil/history/coe.htm>

Yellow Fever Experimentations Congressional Gold Medal Awardees: <http://dallaslibrary.org/CGI/>

goldmedals/yellowfever.html, and <http://clerkweb.house.gov/histrecs/househis/lists/medal.htm>

United States Secretaries of War and Secretaries of the Army: <http://www.army.mil/cmh-pg/books/sw-sa/SWSA-Fm.htm>

For Events:

Public Health in Cuba 1865-1917: <http://www.armymedicine.army.mil/history/booksdocs/spanam/gillett3/ch9.htm>

Records of the Military Government Government of Cuba (Major General Ludlow): <http://www.nara.gov/guide/rg140.html>

Philip S. Hench Walter Reed Yellow Fever Online Collection

<http://yellowfever.lib.virginia.edu>

Phase II

Final Report

February 2004

The Philip S. Hench Walter Reed Yellow Fever Collection Digitization Project Web Site received the **2003 Waldo Gifford Leland Award** for writing of superior excellence and usefulness in the field of archival history, theory, or practice from the Society of American Archivists.

I. Organization

A. Review of Archival Materials and Collections

1. Conducted a preliminary review of supplemental collections to estimate the number of additional documents to include in this final phase of the project. Collections and materials reviewed included the Philip S. Hench Walter Reed Yellow Fever Collection, the Henry Rose Carter Papers, the William Bennett Bean Papers, and four books now out of copyright in the Claude Moore Health Sciences Library; the Jefferson Randolph Kean Papers, a James Clayton Reed letter, and the Dorsey Mahon McPherson Papers in the Albert and Shirley Small Special Collections Library of the University of Virginia Library; and six letters by Walter Reed in the Library of Virginia Archives and Manuscripts Department. We anticipated adding approximately 2,500 additional pages to the 5,500 already available on the Web site. The new additions, in keeping with the standards set for materials in the first phase of this grant project, were digitized, identified, described, transcribed, marked up using XML, and

provided worldwide access via the Web.

2. Conducted a full review of the Philip S. Hench Walter Reed Yellow Fever Collection, including photographs, to select additional items for scanning. As a corollary to the photograph review process, we noted corrections needed in the existing photograph files on the site, and identified and corrected processing errors in the original photograph collection that were incorporated in the online version.

3. Conducted a full review of the Henry Rose Carter Papers (seven manuscript boxes) and selected items for scanning.

4. Conducted a full review of the Jefferson Randolph Kean Papers (all accession numbers: 25 manuscript boxes) and selected items for scanning.

B. Secured Copyright Permissions

1. Obtained permission from the Albert and Shirley Small Special Collections Library of the University of Virginia Library to post their images on the Web.

2. Obtained permission from the Library of Virginia to post their images on the Web.

3. Clarified the copyright policy of government documents in the National Archives and Records Administration (NARA), and obtained permission to post items from NARA on the Web.

C. Undertook Technological and Equipment Review

1. Collected information on the capabilities and pricing of additional project hardware to handle large-format, fragile-book, and artifact imaging.

2. Selected and purchased a digital camera and lens.

3. Selected and purchased a large-format scanner.

4. Arranged for and installed software upgrades for project scanners and computers.

5. Arranged for use of the library's Media Studio equipment to supplement our own department's capabilities.

6. Created space on the library's server for uploads of Innodata completed XML files and tracking documents.

D. Established Staffing Priorities and Agreements

1. Renewed contract for data conversion and markup with Innodata, Inc.

2. Assigned project responsibilities to existing staff:

- i. Joan Echtenkamp Klein, Assistant Director for Historical Collections and Services, Project Director;
 - ii. Hal Sharp, Historical Collections Assistant, Project Assistant and Exhibition Text Writer;
 - iii. Nadine Ellero, Head of Intellectual Access, Project Metadata and Authority Control Specialist;
 - iv. Bart Ragon, Library Webmaster, Project Web Designer and Technical Consultant;
 - v. Ina Hofland, Sara Huyser, Janet Pearson, and Alison White, Historical Collections Part-time Staff, Document Selection, Analysis and Editing, Name and Place Authority. Ophelia Payne, Library Cataloguing Specialist, also provided authority assistance.
 - vi. Library Circulation and Administrative Staff also provided assistance: Susan Yowell, document editing and authority work; Liz Ford, Bukuriye Maqani, and Wendy Rosson additional scanning.
3. Interviewed for additional project staffing, and hired three temporary project staff members:
- i. Tim Noakes worked from January until mid-May 2003, and assisted with document selection and editing;
 - ii. Mollie Donohue worked from mid-March until mid-June 2003 scanning documents.
 - iii. Jennifer Hogg provided assistance with Spanish language editing.
4. Established a Memorandum of Understanding with the University of Virginia's Electronic Text Center, which houses the site and serves as technical consultants.
5. Matriculated a project staff member into two Rare Book School courses to provide in-house support for certain aspects of the current phase of the project:
- i. "Implementing Encoded Archival Description" taught by Daniel Pitti (January 2003);
 - ii. "Electronic Texts and Images" taught by David Seaman (March 2003).

II. Implementation

A. Scanning and Processing Documents

1. Created archival digital files (600 dpi TIFF files, burned on CD) of 5392 scanned pages.
2. Batch-processed image files for Innodata and Website use (400 dpi, 150 dpi, 72 dpi JPEGs, and 72 dpi GIFs), for a grand total of 26,960 image files.

3. Created metadata for all 1950 documents, including:

- i. writing an abstract or summary;**
- ii. assigning each document to one or more subject areas;**
- iii. adding names and places to metadata headers as well as to master names and places lists.**

4. Prepared and mailed 18 image file and metadata CDs to Innodata India.

5. Created back-up image file CDs with index and back-up of metadata files.

B. Review and Editing

1. Received periodic FTP transfers from Innodata for all 1950 documents, transcribed and marked-up as XML files.

2. Resolved technical questions from Innodata concerning images and interpretation of documents.

3. Edited all documents for accuracy of transcription, name and place authority, metadata, and XML coding.

4. With the University of Virginia Electronic Text Center, created a "dummy" site to link preliminary edited documents and images in a web environment.

5. Undertook second and third edits of the XML documents.

6. Edited and corrected the master names and places authority lists.

7. Corrected cataloguing errors in the original manuscript collections.

8. Corrected errors in Phase I XML documents and images.

C. Web Site Preparation

1. Made back-up of existing (Phase I) site.

2. Assembled corrected XML documents and image types from Phases I and II:

i. made cumulated master DVDs for the Electronic Text Center site;

ii. made cumulated master DVD back-ups for our records;

iii. made additional master back-up CDs for our records.

3. Rewrote existing Web site pages (HTML documents) and added new pages as necessary.
4. Had the Electronic Text Center regenerate preprogrammed search and index pages for Web site access.
5. Replaced Web pages on the appropriate server and recreated links to HTML documents as needed.
6. Opened the expanded site to the public on 27 February 2004.

III. Publications and Professional Activities by Project Staff (Phases I and II)

1. "Grand Opening: The Philip S. Hench Walter Reed Yellow Fever Collection Digitization Project Web Site," presentation for the 2001/2002 History of the Health Sciences Lecture Series, the Claude Moore Health Sciences Library, 30 January 2002.
2. "The Philip S. Hench Walter Reed Yellow Fever Collection Digitization Project Website," presented for luncheon workshop, "Unlocking the Promise of the Internet: A Sampling of the History of Medicine Websites," 2002 American Association for the History of Medicine Annual Meeting, 26 April 2002, Kansas City, MO.
3. "Rediscovering yellow fever: the Philip S. Hench Walter Reed Yellow Fever Collection Digitization Project," in the "Educational Media and Technologies, History of the Health Sciences Section, Library Digitization Projects" Medical Library Association Annual Meeting, 19-22 May 2002.
4. Nadine P. Ellero, "Panning for Gold: Utility of the World Wide Web for Metadata and Authority Control in Special Collections," *Library Resources & Technical Services* 46 (July 2002): 79-91.
5. "The Philip S. Hench Walter Reed Yellow Fever Collection Digitization Project: Two Years and Over 5,000 Documents Later," Science, Technology, and Healthcare Roundtable of the Society of American Archivists Meeting, 24 August 2002, Birmingham, AL.
6. "Administering a Grant, or What WERE we Thinking?" Mid-Atlantic Regional Archives Conference Fall 2002 Meeting, 25 October 2002, Poughkeepsie, NY.
7. Joan Echtenkamp Klein, "The Philip S. Hench Walter Reed Yellow Fever Collection Digitization Project: Two Years and Over 5,000 Documents Later," *Journal of Archival Organization*, 2002 1(3): 5-34.
8. Phase I project demonstration, Institute for Museum and Library Services annual Web-Wise Conference, February 2003, Washington, DC.
9. Bart Ragon, "Castles Made of Sand: Building Sustainable Digitized Collections Using XML," *Computers in Libraries* (June 2003): 10-12, 63-64.

10. "'This Most Dreadful Pest of Humanity': The Philip S. Hench Walter Reed Yellow Fever Collection," in the "Documenting Disease" session, 2003 Society of American Archivists Meeting, 22 August 2003, Los Angeles, CA.

11. "Building It and Then Going the Distance," in the "After You Build It and They Come, Then What?" session, Mid-Atlantic Regional Archives Conference Fall 2003 Meeting, 31 October 2003, Gettysburg, PA.

IV. Project Numbers

1. XML Document totals:

- i. Phase I - 5,498 documents.
- ii. Phase II - 1,950 documents.
- iii. Total - 7,424 documents (less Phase II duplication of Phase I documents).

2. Image totals:

- i. Phase I - 13,502 pages scanned.
- ii. Phase II - 5,392 pages scanned.
- iii. Total - 18,894 pages scanned.
- iv. Grand total, combining GIFs, JPEGs (72, 150, and 400 dpi), and TIFFs - 94,470 images created.

University Of Virginia
Health Sciences Library



Historical Collections

For
comments
&
suggestions
about this
page:
jre@virginia.edu

Last
Modified:
Thursday,
November
04 2004

University of
Virginia
[Health System](#)
Claude
Moore Health
Sciences
Library
1300
Jefferson
Park Avenue
P.O. Box

© 1998-
2001 by the
Rector and
Visitors of
the
University
of Virginia
Disclaimer
Statement

800722
Charlottesville,
VA 22908-
0722
(434) 924-
5591