

## Narrative Description

### Significance

"All the papers in my office [have] been destroyed," wrote Secretary of War Samuel Dexter to Captain James Taylor on November 19, 1800. [Letters Sent by the Secretary of War, Microfilm Publication No. 6, Records of the Office of the Secretary of War, Record Group 107, National Archives, Washington, DC.] On the night of November 8, a fire had burned all of the papers, records, and books in the War Office. Today the official records of the War Department effectively begin with Samuel Dexter's letter quoted above. *Papers of the War Department 1784-1800*, an innovative editorial project, has set out to change that. It has located copies of some 45,000 of the documents that perished in that fire, and will publish them on line in 2006.

In the era 1784-1800, the War Department was the Federal agency responsible for Indian affairs, veteran affairs, naval affairs (until 1798), as well as militia and army matters. It regularly spent seven of every ten dollars in the federal budget. *Papers of the War Department 1784-1800*, therefore, is much more than a mere military record of the period. It was through the War Department that the people regularly engaged their new government, and through the Secretary of War that the government had contact with them. For that reason, the documents being gathered by the project will be used by a wide spectrum of persons--historians, anthropologists, political and social scientists, biographers, genealogists, and many others. This record is also a window onto a time when there was no law beyond the Constitution, a time when the administration first worked out its understanding and interpretation of that new document. *Papers of the War Department 1784-1800* will fill the most significant gap in the documentary record of the government of the early republic.

The papers dealing with the department's major areas of responsibility--Indian affairs, military and militia affairs, veteran affairs, and naval affairs--have applicability far beyond military history. Between 1784 and 1800 the new nation's diplomatic conflicts with the British in the north, the Spanish in the south, and the French on the high seas became inextricably involved with national policies on Indian affairs, on the advancing frontier, and on the expansion of the army and creation of a navy. Likewise, as families pushed westward, the social history of the nation became entangled with military and Indian affairs on the frontier. The War Department, for example, operated the only federal social welfare program, providing pensions to invalid veterans and the widows and orphans of those who died in the nation's wars. The War Department's connection to the nation's commerce is even more pervasive. During the 1790s, more than seventy percent of the federal budget (debt service excepted) passed through the hands of the Secretary of War. The War Office did business

with commercial firms across the nation; it was the nation's single largest consumer of fabric, clothing, shoes, food, medicine, and, of course, weapons and gunpowder, and in those years its arsenal in Springfield, Massachusetts (though small by later standards) became the largest manufacturing facility in the nation.

The completion of the *Papers of the War Department 1784-1800* will provide new material relative to many topics in the Confederation and Federalist eras. Military records are rich in information relevant far beyond the confines of *military history*. Ration lists reveal much about the diet of the period (and trace its change). Clothing records, including information on the determination of annual allowances and issue, say much about dress of the common man in these years. The movement of supplies (and the recorded cost of such movements) yields a vast amount of unexploited economic data relative to transportation and infrastructure of the era. The detailed record of the Army's contacts along the nation's several boundaries explains much about foreign relations, and the War Department's day-to-day dealings with the Indians illuminates both the formation and execution of policy relative to Native Americans.

Few scholars have ever tried, in any systematic way, to locate those War Department records that do survive--scattered as they are in collections all across the country and in the repositories of a number of other nations. In fact, only three historians--James Ripley Jacobs [*The Beginnings of the U.S. Army, 1783-1812*, (Princeton: Princeton University Press 1947)], Harry M. Ward [*The Department of War, 1781-1795*, (Pittsburgh: University of Pittsburgh Press, 1962)], and Richard H. Kohn [*Eagle and Sword, The Federalists and the Creation of the Military Establishment, 1783-1802*, (New York: The Free Press, 1975)], all writing about the War Office--have examined more than a small fraction of the existing War Department records of this period. Historians, in other fields, who would have profited by consulting the military record as a part of their effort to record the political, diplomatic, social, or economic history of the Confederation and early national periods had neither the time or the resources to do so -- *Papers of the War Department 1784-1800* devoted seven researcher years to the location of these documents. Too often, as a result, their examination of military records (when they have examined them at all) has been limited to the few dozen War Department documents of the period reproduced in the *American State Papers*. [Walter Lowrie and Matthew Clarke, eds, (Washington, DC: 1832-1861)] Even Jacobs, Ward and Kohn saw only a small fraction of the material available--consulting a few dozen collections in a score of repositories. By contrast, *Papers of the War Department 1784-1800* has consulted thousands of collections in hundreds of repositories at home and abroad. [See *Repositories Visited & Manuscript Collections Examined* (Appendix E).]

Given adequate funding, *Papers of the War Department 1784-1800* can complete its work by 2006. By the spring of 2004, the beginning of this grant period, all of the research, copying, scanning, image processing, and initial database records will be done. At that point we will be well into the last major task -- drawing from each document the keywords and phrases that will allow users a near-full-text search capability. Such work is labor-intensive, with no shortcuts. The award of this grant request, a total of \$200,000 over two years, will make it possible to complete the work on schedule.

In July 2003, *Papers of the War Department 1784-1800* was honored in its selection by the NEH as a "We the People" project -- acknowledging it as one of those efforts "that explore significant events and themes in our nation's history and culture and that advance knowledge of the principles that define America." Soon the War Department papers (1784-1800) will be made available everywhere to students and scholars of the nation's early history. This will provide a unique and valuable new window onto the critical and formative early years of this new republic.

## Background of Applicant

East Stroudsburg University of Pennsylvania is situated in East Stroudsburg, Pennsylvania, at the head of the Delaware Valley--literally overlooking the Delaware Water Gap. Located in the foothills of the Pocono Mountains along Interstate 80, it is some seventy-five miles west of New York City and eighty-five miles northeast of Philadelphia.

Founded in 1893 as East Stroudsburg Normal School, East Stroudsburg University is now one of the fourteen schools in Pennsylvania's State System of Higher Education. ESU is a state-owned, tax-exempt, co-educational institution enrolling 6,300 undergraduate and graduate students, with a faculty of 274. It offers 68 undergraduate programs as well as masters degree programs in 9 fields including history.

The *mission* of East Stroudsburg University is to provide high quality programs in both traditional and emerging fields of study which recognize and promote human and intellectual diversity, and to prepare graduates to enter a complex, changing global society with competence and confidence. The University is accredited by the Commission of Higher Education of the Middle States Association of Colleges and Schools, the National Council for Accreditation of Teacher Education, and others.

East Stroudsburg University's Kemp Library houses roughly 340,000 monographs, and 85,000 periodical volumes. The library is a depository for both U.S. Government Documents and Pennsylvania State publications. Kemp Library also holds over 1,000,000 pieces of microfilm, including an impressive collection of microfilms covering the early national period -- many of which were vital to our work: Papers of the Continental Congress, the Washington Papers, the Adams Papers, the Jefferson Papers, the Hamilton Papers, the Knox Papers, the Timothy Pickering Papers, the Winthrop Sargent Papers, the Robert Morris Papers, and the huge Draper Collection.

East Stroudsburg University has been host to *Papers of the War Department 1784-1800* since the inception of the project in 1994. The support and cooperation that the project has received from the University over the years is impressive. Our mutual relationship is a model that could profitably be emulated by other documentary editing projects and smaller schools. Although the University does not pay any of the staff salaries, they have assumed the burden of providing office space, phone, postage, and the myriad of services required by an campus entity. The total value of the direct and indirect costs borne by it on our behalf is impressive. The University's senior leadership has supported us at every turn. The faculty sends us their best students, and, the hand of collegial friendship has been extended from all across the campus.

## History, Scope, and Duration

The project's genesis occurred almost 20 years ago while Dr. Crackel, the project's director and editor, was doing research in hundreds of late 18th and early 19th century manuscript collections in repositories across the country--work that later became his *Mr. Jefferson's Army* [New York: New York University Press, 1987]. At the National Archives he had been informed that no War Department papers for this period survived the 1800 fire. As his work progressed, however, he noted that many of the collections he was using outside the Archives contained copies of War Department documents some version of which had likely been lost in the fire. There were so many that it seemed likely that a useful reconstitution of that

record could be achieved if only these disparate pieces could somehow be drawn together. In 1989, when Crackel was asked to write an article about the early War Department for the National Archives's journal, *Prologue*, he took the opportunity to suggest such an undertaking.

Serious progress toward establishing the project did not come until early 1993, however. That spring Dr. Crackel drafted a detailed proposal and presented it to Dr. James E. Gilbert, then President of East Stroudsburg University. In August 1993, the University agreed to sponsor the project. Crackel was appointed Director and Editor and began immediately to develop a detailed work plan and to establish a fund raising effort. In the fall, Kathryn M. Willis joined in the effort as an Associate Editor.

The initial aim was to produce only a calendar of those War Department papers that could be found, indicating their location, and supplying just enough description to facilitate access through indexing. This alone would be a boon to historians of the early military establishment. After an initial survey, however, it was clear that the project would find thousands of documents (early estimates were between 12,000 and 15,000) and that a published calendar would be over a thousand pages in length. Such a calendar seemed an unlikely undertaking for any publisher. The solution, it seemed, was to put the *calendar* on CD-ROM. It was only a short step further to the notion of putting the document images on the CD as well, and to include in the database keywords and phrases from the text of the document -- the persons, places, things, ideas, and issues mentioned. This would allow near-full-text searching, and would provide a uniquely valuable tool for scholars. The great advantage of this approach over traditional microfilm would be its searchability. Its advantage over traditional published editions would be comprehensiveness, cost, and speed.

By the late fall of 1993, the project's output had taken final form--digital images linked to searchable text describing them and driven by a sophisticated search engine with a powerful display capability and graphical user interface (GUI). This promised a highly cost effective means of giving unprecedented access to the papers of the War Department.

The first major grant proposal was made to the National Historical Publications and Records Commission (NHPRC) in January 1994, detailing a project that would be launched that fall, and whose principal output would be a CD-ROM with digital images and searchable text tags. On June 16, *Papers of the War Department 1784-1800* was offered an NHPRC grant of up to \$34,000, on the provision that it be matched dollar-for-dollar by September 1st. The project staff (with assistance from the University grants office) spent the rest of June and the months of July and August 1994 in an effort to raise the matching funds. By September 1, the project had more than matched the NHPRC grant, raising \$40,000. With this grant and matching funds in hand, the project was formally launched on October 1, 1994.

The next *two years--project years one and two* [1994-1996]--were spent searching for copies of the lost documents, designing the project data base, and raising money. The project staff visited hundreds of repositories and contacted hundreds more by mail, obtaining copies of thousands of war department documents. At the same time, we selected DataPerfect as our data base program and created and tested the data entry formats. Fund raising, in this period, was remarkably successful. By July 1995, when the project was just nine months old, we had raised over ninety-five percent of the money we originally anticipated needing--some \$671,000 towards the anticipated budget of \$705,000. (Funds raised in this period included a \$385,000 grant from the Legacy Program of the Department of Defense, two grants covering four years from the NHPRC totaling over \$146,000, numerous grants from individuals and private foundations amounting to in excess of \$52,000, and a commitment from East Stroudsburg University for roughly \$88,000 in indirect costs.)

Our celebration of this fund raising success was soon tempered, however, by the bittersweet discovery that the original scope of the project was badly understated. Although the National Archives steadily maintained that they held few War Department documents for the period before 1800 (and neither Jacobs, nor Wade, nor Kohn had found much), the project soon learned that, in fact, there was a huge cache of relevant material there. The first inkling that the project had of this came when a reviewer's comments about the project were furnished to us. This reviewer, obviously an archivist at the National Archives, noted that, "I am currently compiling a comprehensive finding aid of about 100 pages entitled 'Records of the War Department and U.S. Army, 1784-1800, in the National Archives.'" The fact that a guide to the Archive's holdings of pre-1800 War Office documents would take 100 pages to describe was an eye opener. The archivist proved to be Howard H. Wehmann, who had spent many years working in this area, and who, a year later, retired from Archives and came to work for the project.

The largest components of the cache that Wehmann had identified were the papers of the Accountant of the War Office (papers which had been saved at the time of the fire) and the papers of the Purveyor of Military Stores (who, with his papers, had not moved from Philadelphia to Washington). This discovery promised to double, at least, the number of documents the project would deliver. We were delighted. There was a down-side, however. Our very success became a problem. Processing such a large number of documents in Washington meant increasing the staff there and at ESU, extending the term of the project, and, most problematic, enlarging the budget. Normally, we would have returned to our major funder for additional support. But, by the time the impact of this new material became clear, Department of Defense had eliminated the Legacy Program--our main underwriter--and, with its demise, went any further hope of DOD funding.

The *third and fourth years* of work [1996-1998] saw the completion of the document searches in the field, although search work at the National Archives continues into year five. In total, Crackel and Willis visited over 200 repositories in the United States and overseas, and consulted some 3,000 collections. [See Repositories Visited & Manuscript Collections Examined (Appendix E)]. Hundreds of other sites were contacted by mail or phone, and a number of these contacts also yielded documents. During project year "four" we also realigned our project calendar to coincide with a new NHPRC funding schedule. Project year four extended from Oct 1997 through Feb 1999.

During this same period, the director and student staff at East Stroudsburg University began the two-part process of (1) transforming the paper and microfilm copies we had collected to the digital images and (2) creating the searchable tags--text descriptions of the documents--that will ultimately be placed on the CD-ROM and allow the user easily to find documents of interest. This work was hampered during the latter part of year three and much of year four by funding constraints, which severely limited staff size and hours and delayed equipment purchases and modernization--including upgrades required by Y2K considerations.

In project *year five* [1999-2000] we solved part of the funding problem by convincing the NHPRC to fund the project at a level on a par with other similar projects that they supported. We were then able to restore hours, to increase the staff, and to put in place the Y2K compliant equipment and software necessary to move more rapidly toward completion. Project *year five* also saw the completion of document searches and scanning at the National Archives.

*Project years six and seven* [2000-2002] began with the very welcome news of two significant two-year grants, one from the NHPRC and another from the NEH. The NHPRC grant was for \$105,000 for project year six (of which \$10,000 was earmarked for consultation on Computer Output Microfilm [COM] and search engine development), and \$85,000 for project

year seven. The NEH grant -- most significant because it was the first NEH funding the project had received -- was also for two years -- an outright grant of \$50,000 and a matching grant of up to \$50,000. (The project was able to raise only \$15,000 toward the matching grant.) In these years the project completed much of the copying, scanning, and image processing, and continued the initial data entry -- the effort to create a data record for each document (what we call Data Entry I).

During *project years eight and nine* [2002-2004] the focus has been on Quality Control, Data Entry I and the beginning of the content data entry or Data Entry II. Data Entry II is the process by which all names, all places, and all events, issues, ideas, and items mentioned in a document are listed for each document. These keywords and phrases provide the basis for document searches based on the textual content of the documents. The work was begun, but had barely gotten under way when another funding crisis developed.

In August 2001, when we applied for our 2002-2004 NEH grant, we suggested that we should be able to complete the project in two years, or by mid-2004. This projection was predicated on two conditions: first, that we would be fully funded, and second, that the document count would be between 30,000 and 35,000 documents. The first condition would have allowed us to add the staff needed to finish the work in a timely manner. Full funding did not materialize. In fact, the project received barely half the funding it requested from the NEH and the NHPRC -- our major funding agencies. Funding was so critical in year eight that Dr. Crackel (Director/Editor) accepted a one-year Visiting Professorship in History at West Point, maintaining only an oversight relationship with the work of the project. A year later he returned full-time to the project.

As we began our planning for *project years ten and eleven* [2004-2006] we again looked toward completing our work in just two more years. In August 2003, based on the fixed expectation of delivering approximately 30,000-35,000 documents, we began our work on this proposal. There seemed no reason to believe that we could not, with minor staff augmentation, complete the project in the two additional years addressed by this grant proposal.

In September, however, we discovered a new reality in the form of a large cache of new War Department documents. These documents were some that had been found and logged by our Washington team at the National Archives toward the end of their effort but never reported to the project office at ESU. It was during a routine foray into the working papers of the Washington team (which had been shipped to ESU for storage) that the problem and its scope came to light. In excess of 10,000 documents had never been identified to ESU for processing. We immediately ordered the microfilm that holds these images, reinstated a microfilm scanning contract, and began the process of putting together a new team of undergraduates to do the required image processing and preliminary data entry. We expect to have this work completed by the end of the summer of 2004 (early in the first year of this grant period).

This means that the total number of documents will likely reach and may exceed 45,000 documents. To complete our work by mid-2006 will mean adding at least two new assistant editors to the staff early in the summer of 2004. With this augmentation devoted fully to Data Entry II, we believe that we can still complete our work in two years. The budget and work plan that accompany this proposal are structured with that end in mind.

The work plan for 2004-2006 includes the early completion of residual image processing and data entry for the newly discovered documents, followed by a quality

control program that will insure that we have found and copied (and scanned and processed) all of the War Department documents that are accessible. The major focus, of course, will be the completion of Data Entry II. Once we are sure we have captured all of the accessible documents and have completed all of the data entry (I & II) we will prepare our text and image databases for hand-off to the E-Text Center. Throughout the two years we will be working closely with Virginia's E-Text Center to prepare to prepare for the moment when the text and images are hung on their web site. We believe that with the funding requested from the NEH (and NHPRC) we can complete this work by mid-2006. Without that funding, with staffing at current levels, we anticipate at least two additional years of effort.

## Methodology and Standards

From a methodological perspective the project is divided into three processes: *collection*, *digitization*, and *preparation for web-based access*.

**1. Collection.** Planning for the project's extensive collection effort began with a definition of a "War Department document." Obviously, all communications to or from the War Department were included, but we soon concluded that other categories of documents were also likely to have been among the War Department's papers at the time of the fire. Among these were documents that originated in other executive offices (including the President's) and the Congress that touched on subjects under the purview of the War Office, even when they were not addressed to the Secretary of War. In addition, many letters between third parties seemingly found their way into the department's files--usually as enclosures to other official correspondence. In the end we included such documents any time there was strong reason to believe that a copy of the document was among the papers destroyed in the fire.

War Department officials likely to be on one "side" or the other of the correspondence, included: the Secretary of War and any assistants; the clerks of the War Office; the Accountant and his clerks; and the Superintendent of Military Stores and his clerks. All correspondence related to activities of the War Department that was written to or by these persons was included. In addition letters written by certain Army (and Navy) officers are counted War Department correspondence when those officials wrote from the seat of government (or from wherever the Secretary of War was) *and* wrote at the behest of the Secretary of War. This often applied to the army's adjutant or inspector (or adjutant general and inspector general). It might also apply to any of the several commanding generals, quartermasters or quartermaster generals, commandants of the Corps of Artillerists and Engineers, paymasters, and judge-advocates.

The first effort to identify the location of collections of interest to the project was to determine where others had found them. We examined the secondary literature and noted where other scholars had worked. We expanded this list by consulting NUCMC, RLIN, and OCLC, looking for papers of the various Secretaries of War and their subordinates, of all army and navy officers of the period, of the superintendents of Indian affairs and Indian commissioners, and of the Governors of the states and territories and their county lieutenants. The records of Congress and the papers of key members of Congress were also identified. Finally, we consulted the *Directory of Archives and Manuscript Repositories in the United States* and the *Directory of Historical Agencies in North America* to identify additional repositories that might hold collections with War Department documents. We wrote or called over 700 other repositories to determine whether or not a visit was warranted. Somewhat later, we examined the files of other

related documentary editing projects to ferret out the War Department documents they had found.

In the end, we visited over 200 repositories and consulted over 3,000 collections in the United States, Canada, England, France, and Scotland. The effort at the National Archives identified over 160 series which were searched—roughly 4,000 boxes and 170 volumes. As of the date of this application, we anticipate publishing roughly 45,000 individual documents.

**2. Digitization.** The second methodological area is digitization--the preparation of both the digital images and the searchable text tags that allow those images (documents) to be identified by a search engine provided on the CD or DVD. This area includes three discrete activities--scanning, image processing, and data entry.

a. **Scanning.** Copies of documents came to us for scanning in four forms--originals (which we scan on site), xerographic copies, microfilm copies, and digital image files. At the few repositories where no capability existed to provide any form of copy from bound documents, the project Director filmed the documents using a 35mm camera and high contrast black and white film. This film is processed in the same manner as microfilm. In some cases we were allowed to scan original documents on site using our Fijitsu M3096GX flat-bed scanner. In rare cases, repositories scanned the documents for us, giving us digital image files rather than hard copy.

The electrostatic copies collected in the field were scanned at our East Stroudsburg University facility using the M3096GX scanner. Microfilm (and 35mm film shot by the project staff) is scanned either at ESU using our Canon Microfilm Scanner 100 or by a contractor who does bulk scanning for us on a high-speed Sunrise scanner. In late 1998, the project's flat-bed scanner and two dedicated computers were moved to the National Archives. There, loose originals are scanned on the flat-bed scanner while bound documents are microfilmed and then scanned. In 2000 this equipment was returned to our ESU facility and is currently operated there as needed.

In all, about twenty-five percent of all documents came to us as electrostatic (hard) copies. About forty percent of the documents are on microfilm (much done specifically for the project), and the balance (about thirty-five percent) are scanned from the originals.

The 3096GX can provide gray scale images of documents up to 11.7"x 17". Scan time is 2.7 seconds (@ 300 dpi), allowing through-put as rapidly as copies can be retrieved from their boxes, laid on the scanner, and then refiled. The Canon Microfilm Scanner 100 produces bitonal (black and white) images at 400 dpi in about 4 seconds. In both cases the resulting files are saved as TIFF (Tagged Information Format Files) images. To provide secure, yet accessible storage, the files are burned onto a CD-R (writable CD) at this point and filed to await further processing.

At the National Archives, where extensive scanning was conducted from the fall of 1998 to the fall of 2000, files were written onto a removable hard drive (SparQ) as they are scanned. As soon as the files begin to approach the 650Mb limit of the CD-R disk, the drive is removed and transported by "sneaker net" to an adjacent computer dedicated to CD-R burning. There the contents of the SparQ drive are burned onto a CD-R. While this CD and a duplicate are being burned (roughly a 35 minute process), scanning continued using a different SparQ cartridge. The CD's were then mailed to the ESU facility for further image processing. At ESU the scanner is connected to the local area network and scanned images can be moved to any machine on the network as required.

b. Image Processing. Faced with the need to offer as many as 150,000 separate images to the user it was essential to reduce file size to a minimum. Uncompressed gray scale images averaged 10Mb in size and without being reduced would require more than 2,000 CD-ROMs. In practice, by converting them to black and white images and by compressing the files, we are able to reduce their size to about 60Kb .

But, image processing is more than a mere reducing of file size. Utilizing Adobe Photoshop 4.0.1 we crop the image and, in the process of conversion from grayscale to black and white, do some clean-up as well. Often bleed-through from the back side of the page and other stains that partially obscure the text can be removed digitally. This process is largely automated. Recent upgrades in computer equipment and the new capabilities of Photoshop 4.0 have allowed us to automate the work, dramatically reducing processing time from seven minutes to about twenty seconds per image. This is a huge savings; every minute spent processing a single image becomes a man-year of effort when repeated for each of our 150,000 images.

The bitonal images produced by the microfilm scanner are processed separately. In addition to Photoshop, some of these images are run through a program called ScanFix, which we use to remove small clusters of black pixels (noise).

When images are then finished in Photoshop (and ScanFix), all of the pages (images) of a single document are concatenated into one file--a multi-page TIFF file. Most of these files contain only a few pages, but a few are much larger. Letter books, for example, which contain largely War Department materials are presented in their entirety to the user as a single file, some of which are hundreds of pages long. For these large multi-page files, the search engine will open to the page on which the desired letter begins. Once in the letter book, however, the user will be able to leaf backwards and forward through the whole volume.

When image processing is completed each file is examined in a **Quality Control** process to insure that the images have been properly processed and that the images are in the correct order. When this process is completed, the files are run through one last program, Stamper, which applies a small strip of text to the edge of the image identifying the project and the image file name. Page numbers are also printed on all files over ten pages in length. The finished files are then burned onto a CD-R for storage until needed in the final mastering process.

c. Data Entry. The project database was created with DataPerfect 2.3, but Y2K issues convinced us that we must migrate to a different database. In June and July 1999, the project data was successfully migrated to Microsoft Access 97. Information is entered into the database about each document in a two-phase operation. This data is necessary to provide the user with information that fully describes each document. As discussed below, it is this information that is examined by the search engine, for our "documents" are mere digitized images--a digital picture, not text that could be read by a computer.

During Data Entry Phase I, the student data entry team creates a record for each document. Those records include, for example, the document date, the author and recipient, where the project found the document, and a certain amount of project management information. Each record is checked in a **Quality Control** process and corrected (if necessary) before being passed to the next phase.

Data Entry Phase II adds information about the *content* of each document. During this phase, a very brief abstract of the document (which we refer to as the "gist") is prepared and keywords for indexing the document are selected and entered into the database. Keywords will include *all* persons and groups mentioned in the document, *all* places mentioned, and *all* events, ideas, issues, or items mentioned. It is project policy to enter keywords exactly as they appear in

the document, without further elaboration. The “gist,” however, offers us a bit more latitude. Here we employ a glossary of approved terms to describe events mentioned or even alluded to in the text. The full list will be available to users. For example, the “gist” of a letter describing the uprising in Western Pennsylvania in 1794 would employ the term “Whiskey Rebellion” even though contemporary descriptions would not use that term.

Date Entry Phase II (“content” data entry) will be the focus of the project’s effort in the last two years of effort -- the two years for which funding is herein being requested. At the end of Phase II Data Entry each record is examined once more, for **Quality Control** purposes, by a different member of the professional staff. Funding for a new Assistant Editor is included in the current budget. This is another key area where timely completion depends on full funding.

Documents which are *mentioned* in other documents (cited), but for which copies are not located, are treated as if they are “real” documents. They are given a document number and included in the database. The content of these documents will be described to the degree possible. These “cite” records will be made available to the user *via* the same searches that will identify other documents. The inclusion of these *Cite Only* documents aids the scholar for it is often useful to know simply that two persons were corresponding at a particular time, and sometimes the cite tells us something about the substance of the missing letter. In addition, the “cites” give us a fuller picture of what documents were at the War Office at the time of the fire. They will also give us one means of accessing what portion of the whole record we have recovered.

The bibliographical and content data in these document records will later be used to produce searchable text records which will be “tagged” to each of the image files. The information on these “tags” will allow scholars highly sophisticated access to the documents of this collection. If, for example, one wanted to locate all of the War Department documents written by a particular individual, in a particular year (or years), that mentioned a specific, person, place, event, issue or item (or some combination thereof), the project’s search engine will quickly locate and list them. The researcher, writer, student, or teacher can then select any or all from the “hit list” to be viewed, printed, or saved for future reference. It is the searchable “tag” that makes this possible.

**3. Preparation for Web-Based Access.** This methodological phase of the project includes the final quality control checks of text and image data, the preparation of front material, the ongoing coordination with the E-Text Center as the web-based product is designed. It also includes the preparation of sample data for testing, the participation in E-Text Center testing, the final hand-off of the complete data set to E-Text Center, and final coordination with the center as the materials are mounted on the web.

## **Plan of Work**

This is an ongoing project which is requesting a grant for what will be its last two years of operation. A number of activities described above will have been completed when the grant period opens in the spring of 2004. Among them are Data Entry I and image processing. Data Entry II (“content” data entry), which was begun in project year eight, will continue as our major effort in the coming two years. Also important will be the preparation of front material for the on line product and the final quality control checks before handing off the data and image files to UVa’s E-Text Center.

*First six months.* Data Entry II will continue as the major focus of the professional staff. Residual image processing and Data Entry I will be accomplished largely by the student staff under the close supervision of Assistant Editor Sarah Lambert. As soon as that work is completed, we will undertake a detailed quality control enterprise to insure that we have captured and processed all of the identifiable War Department Documents. This will involve: (1) renewed contact with repositories to check on new accessions, (2) a check of our own records to insure we captured all the documents that we identified, and (3) finally a check of secondary sources (with a particular emphasis on recent works) to insure that we have not missed any documents or collections. Also, in this period, will be the initiation of preparation of the front material for the on-line edition, and the continued coordination of the web-based product with the E-Text Center. Toward the end of this period the project will prepare sample text and image database to be used by the E-Text Center in testing their on-line designs.

*Second six months.* Data Entry II will continue to be the major focus of activity. Also continuing will be quality control efforts and the coordination with the E-Text Center and the preparation of front material. Finally, the project will participate, with the E-Text Center, in testing their on-line designs.

*Third six months.* Data Entry II will continue to be the central focus. At the conclusion of the quality control effort to insure that we have captured all of the identifiable War Department documents, Assistant Editor Lambert will begin the process of obtaining formal permissions from all of the repositories from which documents were drawn. Discussions were held with each repository as the searches were conducted, but it was thought better to delay seeking formal permissions until it was clear what we would want to use. This process will include a listing of documents to be included in the final production by source repository and, using that, a final communication with the repositories seeking final permission to publish them on line. It will also include a final detailed check to insure that we have in our databases the text and images for all the documents we intended to include, and the proper links in place to insure that the proper images come up when called for. (A part of the Data Entry II process is a detailed QC check of the text and image files. This final check will simply insure that what we have put in place, stays in place.) Also, in this period, will be the continued coordination with the E-Text Center and the development of front material for the on-line product.

*Last six months.* Data Entry II, the final quality control checks, and the preparation of front materials will all be completed about mid-way through this period. When all data entry has been completed a final preservation microfilm (Computer Output to Microfilm) will be made by a contractor. (In 2002 an early version of this microfilm was produced at the request of the NHPRC as an emergency back-up for the project.) At the same time a final text and image database and copy of the front material will be delivered to the E-Text Center as the basis for the project's on-line publication. Finally, the professional staff will coordinate the final mounting and testing of the web-based version of *Papers of the War Department 1784-1800*. [See the Graphic Work Plan on p. 13.]

## Project Staff

Director and Editor. Dr. Theodore J. Crackel (author of *Mr. Jefferson's Army*, *The Political and Social Reform of the Military Establishment, 1801-1809* and *West Point: A Bicentennial History*) is Director and Editor of the project. Crackel's extensive research into manuscript collections while working on *Mr. Jefferson's Army* gave him an acute sense of the collections that the project would search. The National Archives recognized his expertise in this area in 1989 by asking him to tell the story of the early War Department in *Prologue* as a part of their bicentennial celebration of the ratification of the Constitution and creation of the new Federal government. A long-time member of the Association for Documentary Editing, he attended the NHPRC Documentary Editing Institute ("Camp Edit") at Madison, Wisconsin in 1980. On the technical side, he boasts three decades of computer and systems analysis work.

Associate Editor. David A. Finken has a BA in English Literature and Language from East Stroudsburg University and is presently working on a MS in Computer Science at ESU. Finken joined the project in 1996, as a sophomore. During his senior year he was made Assistant to the Editor. Upon his graduation, Finken joined the project full time as an Assistant Editor. In the absence of the Director/Editor he is responsible for the day-to-day supervision of the project facility at ESU. He is also a member of the Association for Documentary Editing and a graduate of the NHPRC's "Camp Edit".

Assistant Editor. Sarah Lambert holds a BA in History from East Stroudsburg University and is currently enrolled in the masters degree program (history) there. Lambert came to the project as a sophomore in the spring of 1999. Upon her graduation in May 2001 she joined us full-time. Lambert has worked in data entry, image processing, and quality control. She is also responsible for supervising student help. She is a member of the Association of Documentary Editing and a 2003 graduate of "Camp Edit."

Student Assistants/Interns. During the life of the project the History and Political Science faculties send us many of their best undergraduates. We have been pleased that a very significant number of those who have graduated with honors in History in the last several years have been members of the *Papers of the War Department 1784-1800* staff. We have employed as many as 12-15 part-time undergraduates at some periods and many of these stayed with us for two or three years. To date, three of these students continued in documentary editing after graduation -- two with this project, and a third (David Ranzan) with the Thomas Edison Papers project at Rutgers. This year we intend to hire three to four more students and we will accommodate those interns that the History Department wishes to place with us. Our undergraduates have proved to be extraordinarily productive members of the team and have been enthusiastic about the opportunity to do "hands-on" history.

## Dissemination.

In the early days of the project we considered carefully two modes of disseminating our product to the user--the Internet and CD/DVD-ROM disks. Initially, we favored a published CD/DVD-ROM product that we would distributed free of charge to key libraries. Although the

Internet solution was enticing, there were several apparent drawbacks. The two most important were site maintenance and access. First, the project, itself, had no means of maintaining an Internet site which could serve up the documents and at that time there were no sites such as today's E-Text center which could host the project's text and image databases. The second issue was access. Neither speed nor band width were adequate, we concluded in 1993, would handle the demand. And, it was by no means certain that they would be available any time. Scholars, we believe, will want to access, not one or two items, but dozens, hundreds, even thousands of documents and will probably need to print working copies of many of them. In the early years of the project we did not believe that the Internet would support this kind of access.

The CD/DVD-ROM seemed the better solution. One drawback was the high price often demanded for such publications. We proposed to solve that difficulty by producing and distributing the product ourselves -- giving at least 1000 sets of the publication free of charge to the major historical research libraries in the United States and to selected repositories overseas. Sets would also be made available to other interested institutions or individuals at a price of \$100-\$200.

A CD-ROM disk is able to store some 10,000-11,000 images of approximately 60Kb each--the average file size we are able to achieve with bitonal images at 300dpi in TIFF Gp 4 Compression format. The then anticipated 35,000 documents (with multiple images) would occupy as many as a dozen CD-ROM disks. Although workable, a set of this size is unwieldy and inconvenient for the user. DVD-ROM, however, would hold 7 Gb or more of data. We believed it was possible that the entire publication might ultimately fit onto one or two DVD disks. The cost of the DVD-ROM publication would have been about \$10 per set assuming the pressing of 1000 or more sets.

Fortunately, the nature of the project has allowed us to keep options open that would insure the publication's continued availability to institutions and individuals. In recent years high-speed access to the Internet has lessened our concerns about access. At the same time several web sites and major universities have begun to host projects like our own. Therefore, in the summer of 2003, we began negotiations with the University of Virginia's Electronic Text Center with the object of having the center host the project on its web site. Mounted on that site we feel that the first of the disadvantages that initially dissuaded us from Internet publication no longer pertain. At the E-Text Center, the documents will be kept by an institution committed to "build and maintain an Internet-accessible collection of SGML and XML texts and images," an institution with the resources to maintain the integrity of the *Papers of the War Department 1784-1800*. [See Letter of Interest, E-Text Center, University of Virginia, Appendix F.] We have, in the end, concluded that we should go with an on-line publication and we believe that the University of Virginia will provide the user accessibility we desire for the students and scholars of Early American History.