The Expanded, On-line
Trans-Atlantic Slave Trade Database

Project Description

The year 2008 marks the bicentennial of the abolition of the slave trade in Britain and the United States. This project proposes to commemorate this important anniversary by creating an interactive educational Web-based resource about the slave trade between Africa and the New World from the sixteenth to the nineteenth century. Using as its foundation information about 27,233 voyages documented in the renowned *Trans-Atlantic Slave Trade Database* CD-ROM (Eltis *et al* 1999), this project will produce a revised and significantly expanded database that will be freely available via the Internet and will contain more than 35,000 voyages—approximately 85 percent of the entire history of the slave trade. The project will present the database and its auxiliary materials, including maps and archival documents, in a two-tier format: one tier designed for professional researchers and another for K-12 and generalist audiences. The Web-based resource will enable researchers to submit new data to an Editorial Board for vetting and inclusion in the database.

The project's major contributions will be in its provision for continued advances in knowledge about voyages with human cargo and in its expanded diffusion of that knowledge to researchers and the broader public via the Internet. Archival research conducted since 1999 by the present applicants has yielded copious amounts of new information, including records on nearly 8,000 entirely new voyages and information that will enrich the data on more than 9,000 of the original 27,233 voyages. Integrating these and other new findings into the data set will profoundly enhance our understandings of the transatlantic slave trade.

The initial publication of the *Database* transformed scholarly research concerning the slave trade. However, its publication format—the CD-ROM—has limited the database's reach in significant ways. First, at $235.00, the resource is expensive, and its use has been confined primarily to professional researchers with access through university libraries. Second, the original publication was geared toward scholars, and did not contain contextual information that could make it useful to a K-12 audience and the general public. Finally, its publication format produced a static resource that has not allowed for ongoing updates as research yields new discoveries.

The present application seeks support for a two-year grant that will accomplish five goals:

1. To create an enhanced open access website for the data, redesigning the delivery system and data selection interface of the 1999 data set;
2. To integrate new data into the existing 1999 data set;
3. To develop and publish contextual materials that are geared toward the needs of a K-12 audience;
4. To establish an editorial distribution system that will enable the integration of new research findings into the database in an ongoing fashion; and
5. To securely archive and preserve the evolving database publication.

This project brings together a premier group of world slave trade researchers with an innovative team of Emory University digital library development experts to create an open access, two-tier research portal. We envisage multiple levels of usage for this database. At the advanced level, scholars will use the database in two forms: a Web-based version and a downloadable Statistical Package for the Social Sciences (SPSS) version. At a more generalist level, a simplified Web-based interface will offer contextual materials to aid lifelong learners and students in understanding the historical significance of this body of data. This project will provide for the first time a solid and continually updated database on the major branch of migration that sustained the early modern re-peopling of the Americas.

Table of Contents

Narrative

Significance

This project proposes to create an open access Web-based resource to fully document the slave trade between Africa and the New World from the sixteenth to the nineteenth century. This new resource will be launched in 2008, commemorating the bicentennial year of the abolition of the slave trade. It will build upon the highly renowned *Trans-Atlantic Slave Trade Database* CD-ROM (Eltis et al 1999), authored by members of this project team. This publication presented data regarding 27,233 transatlantic voyages—approximately 70 percent of the transatlantic slave trade. This project proposes to significantly enrich the existing database and to make it freely available via the Web. Research conducted by Eltis (PI) the Steering Committee, and members of the Advisory Board over the last six years has yielded records for more than 8,000 entirely new voyages, as well as data that will enhance the records of more than 9,000 of the previously published voyages. The current project seeks to incorporate these new materials into the data set, increasing the coverage of the database to approximately 85 percent of the transatlantic slave trade. The project also intends to broaden the reach of the database by creating a two-tier interface, one tier geared toward professional researchers, and the other incorporating contextual materials for K-12 and generalist audiences. Further, the new Web-based *Trans-Atlantic Slave Trade Database* will be constructed so as to enable rolling updates to the data set as new material surfaces.

We begin this section of the proposal by providing a short synopsis of the slave trade and its importance to scholars in a wide range of disciplines. We then offer an overview of the strengths and weaknesses of the initial database publication. Finally, we present a detailed sketch of the new Web-based resource we envision. The ultimate goal of this project will be to provide a dynamically generated Web-based resource that broadens both scholarly and general understandings of the transatlantic slave trade.

Documenting the Transatlantic Slave Trade

Between 1518 and 1866 an estimated ten million people of African descent entered the Americas via the transatlantic slave trade. This forced migration—the largest in history—linked hundreds of communities throughout the North and South Atlantic world. At the slave trade's peak periods in the 1770s and 1780s, slave merchants in Europe, Africa, and the Americas annually shipped 100,000-125,000 enslaved Africans to the Western Hemisphere. In several individual years 30,000 to 40,000 African men, women and children from modern-day Senegal, Gambia, Ghana, Nigeria, Congo, Angola, and Mozambique arrived at each of two or three large slaving ports, such as Kingston, Jamaica, and Le Cap, St. Dominique (Haiti). By contrast, the number of European migrants arriving per year in major centers on the eastern seaboards of the Americas, from Philadelphia to Rio de Janeiro, rarely moved into the five figure range, and, overall, four Africans arrived in the New World for every European before 1820 (Eltis 2002). Today, most African-Americans are descended from people who were forced into the "Guinea trade."

Yet, more than forty years into the revolution in interest, methodology, and funding of the study of slavery in the Americas, much more is known about the role of Europeans in the re-peopling of the Americas than of Africans. This imbalance will never completely disappear, given that Europeans lived in a post-orthographic environment for centuries preceding their first contact with the New World, while an indigenous written record in the non-Islamic sub-Saharan African world did not begin to emerge until the mid-nineteenth century. Still, much can be done to recapture historical information concerning the forced migration of Africans to the Americas. The records of the movement of Africans to the New World are more abundant, more diverse, and richer than are their counterparts describing the parallel and, prior to 1820, much smaller movement of Europeans across the North Atlantic. This is because most transatlantic voyagers to the north retained elements of self-ownership—eventually, indeed, a full measure of

possessive individualism—while those forced to cross the South Atlantic were chattels whose shipment generated a far greater volume and density of records than did the shipping of free migrants.

Impact of the 1999 CD-ROM Publication

As outlined in more detail in the "History" section below, forty years of research on the transatlantic slave trade in international archives has generated a set of records that is unprecedented in the annals of global migrations. A major milestone in these endeavors was the publication of the *Trans-Atlantic Slave Trade Database* CD-ROM, containing records of 27,233 slaving expeditions between 1527 and 1866. These records provide details about vessels, enslaved peoples, slave traders and owners, and trading routes.

The authors of this database believe that the major strength of the resource, as published in 1999, lay not in its interface, but in its scholarly integrity. The data set listed every source for each of the 27,233 slave voyages, and allowed no less than eighteen source variables for each voyage. The records of data set came from multiple archival sources, a difficult and unusual situation for social scientists and historians. Most databases, such as those arising from census and taxation records, are created from a single source. Drawing on more than one source strengthens a database enormously, but it also poses questions of creation and maintenance which do not usually arise for single source databases. The historical record inevitably generates contradictory information about the same event. Double, or even triple, counting of an event or a datum is a much greater possibility when several sources are consulted. Much of the time spent on the earlier slave trade database (and to be spent on the current project) was used to develop a transparent set of protocols to handle such situations, and to implement these protocols to process data prior to its inclusion. We seem to be at the beginning of a shift toward multi-source databases in history and related disciplines, and such data sets likely will become much more common in the future. Their problems are ones that historians are well suited to solve, given their training and preoccupation with evaluating and synthesizing sources.

An anonymous referee of an earlier grant application described the 1999 CD-ROM as the largest database on pre-colonial Africa that was ever likely to exist, and another commentator compared its impact with that of the Hubbell telescope on astronomy several decades ago (Bailyn 2001, 245). Over the last six years, the database has moved to the center of studies of the volume of the slave trade, its regional distribution, the connections between Africa and the Americas, and questions concerning the impact of slavery and the slave trade on Europe, the Americas, and Africa. It has become increasingly rare to see publications in this field that do not cite the CD-ROM, and the volume of queries and suggestions directed to its authors has increased every year since its publication—a pattern quite different from what is the norm in hard-copy academic print publishing. Indeed, much of this feedback has been incorporated into the present project proposal.

Two quite significant overall findings have emerged from the considerable literature that scholars have generated using the database over the last six years. First, researchers have discovered that there were strong, systematic, and enduring links between particular regions of embarkation in Africa and regions of disembarkation in the Americas. This finding has meant that the older conception, in which the slave trade was thought to draw on a mass of heterogeneous peoples and cultures, is gradually being replaced by a recognition that the victims of the slave trade on any given voyage probably had more in common culturally and linguistically than was previously thought to be the case, and that shifts in the provenance of slaves, though large, occurred slowly. A second major general finding concerns the importance of African agency. The structure of the transatlantic slave trade was the result of interactions between Europeans and Africans. The patterns of rebellion, slave gender and age, geographic origin, and indeed the rise and fall of the slave trade itself, were products of values, beliefs, and conceptions of social identity that were as much African as European. More specifically, on just one of these issues, there now seems little doubt that the pattern of African resistance in the form of slave-ship rebellions had a major

impact on the size and direction of the movement of peoples across the Atlantic over three and one half centuries.

Applications of the CD-ROM in related fields have also steadily increased over the last six years, as judged by scholarly citations and letters received by its authors. The database is widely used in research concerned with human migration. Scholars with interests in political, cultural, or economic issues increasingly call on the age, gender, and regional patterns of the slave trade revealed by the 1999 database. The burgeoning field of Atlantic history has made particularly heavy use of the database. The debate on creolization among the black population of the Americas has recently drawn on findings from the database, as does a major political history of Lagos, Nigeria (Mann forthcoming). The expanding micro-historical approach represented by biographies of individual slaves and people on the organizational end of the trade (including Africans) has also made heavy use of the 1999 resource. More broadly, the database is being employed in attempts to make transatlantic links in mitochondrial DNA by pointing to those areas of Africa that had the closest links with a given part of the Americas, and in epidemiological research on such issues as heart disease and HIV. The field of maritime history has made extensive use of the details on ships, owners, captains and routes contained in the database, as most slave ships were not built and used specifically for the slave trade but found uses in a variety of other maritime endeavors. Finally, we should note the wide interest in genealogy among both scholars and the general public, for which the database has become an exciting tool.

Drawbacks of the 1999 CD-ROM Publication

Perhaps because of the intensive use of the database, the need for its improvement became obvious fairly quickly upon its publication. Six major issues have emerged over the last six years, as described below.

First, the CD-ROM format carries several liabilities. The 1999 data set, once published, was effectively fixed in amber. The CD-ROM offers no efficient means of updating the information contained within the publication. Users have been unable to gain access to new research findings, which cannot be presented prior to the publication of a second edition. If we were to pursue a second edition CD-ROM publication this problem would simply repeat itself as the pace of research continues to move swiftly forward. To further complicate matters, many newer computers are incompatible with the CD-ROM, and others (including hand-held machines, i-Macs, and laptops) are manufactured without CD-ROM capacity.

Second, while many people, including both professional and non-professional researchers, possessed information that was missing from the database, no structure existed for ensuring that this new information would find its way into a new version of the data set. Further, there is currently no established procedure or institutional framework for sharing the responsibility of vetting new information with an editorial board of scholars. This means that the future growth of the database, including the addition of new data and new variables to the data set, is overly dependent upon its small group of original creators.

Third, the CD-ROM is effectively unavailable to many potential users due to its prohibitive expense. The price in 1999 was $195.00. While it is still for sale, its price has increased to $235.00. Even at these prices, Cambridge University Press has not been able to recover the cost of creating the original interface. Moreover, new data, new variables, and new analytical modules have become available since 1999. Making these available will require extensive and costly modifications to the interface. A new CD-ROM, or, indeed, an alternative Web-based platform for the database, cannot be created without significant additional expenditures. If pursued by a publisher, the new publication would mean additional costs to users, either in the form of annual fees for a website (*e.g.*, see the annual user costs proposed for the forthcoming Web-based *Cambridge Historical Statistics of the United States*), or as a set price for a new CD-ROM (which would no doubt be released at a price even greater than $235.00).

Fourth, the 1999 database publication addressed only one audience: researchers with scholarly training. Some professional researchers in the social sciences, sciences, and humanities accessed its downloadable SPSS (Statistical Package for the Social Sciences) formatted files, while others, especially humanities scholars, used the interface to search, select, display, and make use of information held in the database. Yet, the database's potential users could include non-professional researchers and K-12 learning communities. In order to serve these other audiences, the database publication needs to contain a second interface that bears contextual information and teaching guides to help users to make sense of the information contained therein. A two-tiered system could be structured to meet the needs, abilities, and experience levels of all those interested in this topic.

Fifth, many users have reported their desire for path graphics or Geographical Information Systems (GIS) tools that could better represent the routes of the transatlantic slave trade's voyages. Although the current publication contains some illustrations of voyage routes, these maps have been pre-generated as static objects, and are not malleable to user needs. The database authors have wanted to address this need by enhancing the mapping capabilities of the current system, and by incorporating new tools that facilitate user interaction and that allow the display of multiple fields of information simultaneously. However, these goals cannot be accomplished without a new publication of the database.

Finally, the coverage of the CD-ROM database was uneven in terms of both width (number and type of variables included) and depth (number of cases). Particularly, information concerning the very large Portuguese slave trade, and to a lesser extent, the U.S. traffic, was not gathered until after the publication of the 1999 database. Users steadily request information about the voyages of the Portuguese slave trade and the U.S. trade, and we have gathered substantial data over the last six years that nearly closes these significant gaps in coverage. In addition to requests for such voyages, many users have inquired about variables that were not incorporated into the 1999 publication, including slave prices and the ethnicities of slaves. Although the authors have been steadily building this research in the last six years (as discussed below), they cannot add it into the database until it is republished.

Interim Solutions

The authors of the CD-ROM have provided partial solutions to these problems wherever possible. We have provided assistance to users, encouraged users to submit new information, and made new data available to individual users. However, as mentioned above and detailed in "History" below, new data must be processed using established protocols prior to its inclusion in the master data set. As a result, when we have received requests for new data, we have only been able to accommodate those users who could handle it in its raw form. As requests of this nature have increased each year, the authors now face difficulties in responding to user inquiries and needs.

To address concerns regarding the scope of the database, two of the authors of the 1999 data set (Eltis and Richardson) applied for and, in 2001, received a three-year grant from the AHRB, or Arts and Humanities Research Board (as it was then termed), located in the United Kingdom. This grant was for research only, and was specifically aimed at expanding the 1999 data set. As a result of this support, we have made great inroads toward creating the informational basis for 1) adding new variables on slave prices and ethnicity to the data set, and 2) incorporating archival materials concerning the Portuguese slave trade into the database. Specifically, Eltis and Richardson located additional information on 16,597 voyages, 7,501 of which were new to the 1999 database. Although the coverage of the Portuguese slave trade still does not equal that for the British, French, and Dutch traffic, in large part due to the highly dispersed nature of primary sources on the Portuguese traffic, we have made dramatic headway. In addition, we have increased our U.S. slave voyage component by about fifteen percent. The research findings of the last six years will be discussed in detail under the heading "Data Gathering, Analysis, Processing, and Integration."

Creating the Enhanced Transatlantic Slave Trade Database Resource

This project seeks to create a new incarnation of the *Trans-Atlantic Slave Trade Database*, one that both addresses all of the identified shortcomings of the previous version and builds in a capacity for continual updating in the future. The new publication will also provide users with a freely accessible website that presents the database materials in a new two-tier system to ensure the widest possible audience.

The advanced tier will be designed for scholars and professional researchers. Like the current database, it will include a downloadable version of the raw data set that researchers can use with such professional statistical analysis software packages as SPSS. It will also include an enhanced database display interface that will allow users to browse the full data set, conduct searches for particular material, dynamically generate displays of material in response to user queries, and produce interactive maps that illustrate this material with layers of information that are chosen by the user.

The generalist tier will address K-12 teachers and learners, as well as the general public. We will produce teaching guides and contextual materials that enable users to gain knowledge both of the subject area and of the ways that databases work. These materials will include lesson plans, structured assignments, and learning modules that build upon the database and its mapping capabilities. All materials will be produced in conjunction with specialists, and comply with national education standards (see Appendix 3).

Both tiers of the interface will benefit from the new data that will be incorporated into the database, including new records concerning the Portuguese slave trade and such new variables as ethnicity and slave price. Researchers will further benefit from improved path graphics and GIS tools that will encourage users to view graphical components of the database in map forms. As new material becomes available, it will be integrated with the existing data set for instant access by researchers of all levels. Such changes will also be made transparent, so that researchers may see exactly what changed, when it changed, and why it changed. Further, users will be encouraged to become contributors to the data set by submitting their own research findings to an editorial distribution system. These submissions will then be vetted by an Editorial Board, and incorporated into the data set as appropriate.

The tasks involved in this proposed project fall into five broad categories: system design; data gathering, analysis, processing, and integration; production of contextual materials; editorial distribution system design; and storage and maintenance. Below, we provide an overview of each of these tasks and the staff who will be involved in completing it. These tasks are also detailed in the “Work Plan” section.

System Design

This proposal draws upon Emory University as a leading site of digital library research. As executive director of Emory’s MetaScholar Initiative, Martin Halbert has served as principal investigator for many projects in the areas of scholarly research portals and digital preservation, of which the following three examples are representative (see http://MetaScholar.org for details concerning other projects). The *Southern Spaces* Internet journal (http://SouthernSpaces.org) is an innovation in humanities publishing that takes advantage of the capabilities of the online medium for scholarly research by, for example, pioneering ways of updating Web-based scholarly materials after their initial publication while making such changes transparent to researchers by documenting revision histories (*e.g.*, see http://www.southernspaces.org/contents/2004/wilson/revhist.1.htm). The projects AmericanSouth (http://AmericanSouth.org) and MetaCombine (http://MetaCombine.org) have experimented with methods of meaningfully combining improved techniques for organization and access to scholarly information via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) as well as the World Wide Web, exploring combinations of information and services at various levels of abstraction, including: combined search of OAI and Web resources, combined semantic clusters of information, and combined digital library components. The MetaArchive project (http://MetaArchive.org) is one of eight

National Digital Information Infrastructure and Preservation Program (NDIIPP) centers chartered by the Library of Congress to preserve the digitized cultural materials of the nation. The range of these projects and their intrinsic collaborative nature have produced a very strong capacity at Emory for the design and creation of digital scholarly research tools such as the Web-based *Trans-Atlantic Slave Trade Database*.

Creating the new system design and infrastructure for the *Trans-Atlantic Slave Trade Database* will require the completion of three main tasks: 1) transforming the CD-ROM database framework into a dynamic Internet module; 2) designing an appropriate interface for a two-tier Internet system, one tier geared toward professional researchers and another toward the K-12 community and the general public; and 3) producing path graphics and Geographical Information Systems (GIS) tools to illustrate the geographical movement of peoples during migration.

The first task at hand will be to transform the existing database framework into a dynamic Internet module. The project PIs, Halbert and Eltis, will first consult with staff members at Emory University, including the programmers, the Web Design consultant, and the GIS consultant. This group will begin designing an Internet-based module that uses a relational open-source database, MySQL, to serve as the underlying framework for data set delivery. Convening this group at Emory University as the project begins in June 2006 will ensure that the initial transformation of the database framework will produce a component that will be compatible with both the website design and the application of path graphics and GIS tools within the overall systems design.

Under Halbert's guidance, programmers will then create one database platform that will dynamically feed into 1) the downloadable version of the database that is compatible with such statistical software packages as SPSS; 2) the web-based display of the database that can be browsed and extensively queried in each tier of the system (advanced and general); 3) the GIS toolkit that can be used to map various elements of the database; and 4) a tracking system that can display the database as it existed at different dates, and include information regarding what changed, when it changed, and why it changed from its 1999 publication forward. This final element is of utmost importance, as researchers will need to be able to visit different versions of the database when tracking citations of this resource in scholarly works. A first prototype of this system will be issued during month six (November 2006) of the project, and a second system will be released in July 2007 for usability testing with the Steering Committee and the Advisory Board and for use by the education specialists in designing the K-12 contextual materials.

As the programmers work to create the underlying database structure, the Web Design consultant will begin working with the project PIs (Halbert and Eltis), the Steering Committee Members (Richardson, Florentino, and Behrendt, all premier experts on the transatlantic slave trade and long-time contributors to the data set) the Advisory Board members (Miller, Lovejoy, Klein, Werum, and Hahn—see "Staff" for details on each), the Post-Doctoral Research Associate (see "Staff"), and the Research Assistant (see "Staff") through in-person meetings, conference calls and email correspondence in months 1-4 (June-September) to design the two-tier website. In particular, Miller, Lovejoy, and Klein will assist in constructing the navigational system and graphic design of the advanced tier, and Miller, Werum, and Hahn will help to determine what features and navigational system will best suit the K-12 tier of the website (as described in more detail below). The Web Design consultant will then create the navigational system, framework, and graphical design elements for the website across both tiers. After the GIS consultant has designed the path graphics and the GIS toolkit, the Web Design consultant will incorporate these modules into both tiers of the website. Like the database platform, a prototype of the website will be issued during month six (November 2006) of the project. A second prototype of the website will be made ready for usability testing and for use by the education specialists who are designing teaching materials for the K-12 tier in month 14 (July 2007) of the project, and the beta version of the website will be publicly launched at the end of the project period.

The GIS consultant will collaborate with the project PIs (Halbert and Eltis), the Steering Committee Members (Richardson, Florentino, and Behrendt), the Post-Doctoral Research Associate and the Research Assistant in months 1-4 of the project through in-person meetings, conference calls, and email correspondence to determine what database elements best lend themselves to graphical display as mapped elements. The GIS consultant will begin gathering existing shape files that correspond to transatlantic slave trade locations at different points in time. Where necessary, the GIS consultant will also work in conjunction with the Post-Doctoral Research Associate and the Research Assistant to identify maps that can be scanned for conversion into shape files for use with the data set for regions and time periods that do not yet have such shape files. The GIS consultant will construct an initial prototype system during the first six months (June-November 2006) of the project that allows a user to dynamically call upon the data set and the shape files in order to create layered maps of data on demand. The GIS consultant will then create a second prototype of this module for release in July 2007.

Data Gathering, Analysis, Processing, and Integration

Since 1999, as previously mentioned, Eltis (PI), Richardson, Florentino, and Behrendt (Steering Committee members) have continued to gather data from international archives to add to the substance and coverage of the data set. Along with Klein, Lovejoy, Miller, and Nwokeji (Advisory Board members), these scholars number among the most prominent and respected researchers of the transatlantic slave trade, and have published extensive work interpreting the slave trade in leading journals. The data-gathering initiatives of these researchers have been immensely successful, yielding a data set that now includes comprehensive information regarding the forced intercontinental movement of nearly ten million people spanning three and one half centuries. This data set is unrivaled in coverage and density, and is, indeed, quite unique among the records of the great long-distance migrations in global history.

The raw data gathered by the PI and the Steering Committee Members over the last six years at last provide a solid corpus of information concerning the Portuguese slave trade, which, as earlier described, had been identified as the largest gap in the contents of the data set. Figure 1 (see Appendix 1) shows the dramatic advances since 1999 in data about the Portuguese slave trade. As the figure demonstrates, the temporal profile of coverage has greatly increased, and the new data set concerning the Portuguese slave trade is fully three times larger than it was in 1999. The major remaining rich troves of information are located in Havana, Cuba, and in Seville and Madrid, Spain. We have assessed the remainder of the relevant collections in these locations and have determined that we can complete research on site and check problematic entries from the initial data searches with two months of work, as described below.

This recently gathered, raw, unprocessed data alone already reveals two startling findings that are new to the database and to slave trade scholarship. Prior to the gathering of these data, it had long been assumed that the Portuguese were the largest national carrier of slaves among Western nations because the Portuguese carried slaves to every part of the New World over the whole period of the slave trade. The new database indicates that while British participation spanned a shorter period than did the Portuguese, the British carried more slaves than any other Western nation during those years that they were involved in the slave trade. Second, the raw data shows that the U.S. share of the transatlantic slave trade expanded rapidly after the American Revolution. Neither of these findings can be deduced from the 1999 CD-ROM.

But the new information gathered by researchers since 1999 extends far beyond the Portuguese slave trade. Table 1 (see Appendix 2) focuses on some of the most frequently used variables to demonstrate the effect of the new data on our overall data set. Even after allowing for the inevitable deletion of duplicates and elimination of produce (rather than slave) vessels, we have gathered significant amounts of data on more than 7,500 new voyages, a total that we expect to reach 8,000 by the time that this proposed project will begin, as well as data that enriches the voyages previously included in the 1999 database.

For example, rows 5 and 6 in the table show increased numbers of Africans imputed to have embarked on and disembarked from slave vessels. We now have evidence (slated for publication in the new database) regarding nearly 10 million slaves who departed Africa and almost 8.5 million who landed in the New World. These figures are nearing the current estimated size of the slave trade. The new database will provide information on between eight and nine voyages out of every ten that ever went out on a slaving expedition, and on only a slightly lower proportion of all who were forced aboard those vessels. We know of no databases on the long-distance movement of peoples that can compare with this degree of coverage.

Rows 11 and 14 of the table demonstrate that we now know 40 percent more about exact embarkation locations in Africa, and nearly a third more about American disembarkation locations than we did in 1999. We have identified several hundred locations on each side of the Atlantic and have grouped them into regions that allow us to trace the detailed geography of the slave trade in a way not previously possible. It is here, it should be noted, that the path graphics and GIS tools that we propose to employ will be of particular use. With them, we can now trace and make readily accessible to a wide audience the movements of people from Africa to the Americas in detail.

Most of the remaining rows in Table 1 speak to the human experience of the voyage. By combining information about voyages for which we know the precise number of Africans embarked and disembarked (rows 12 and 16) with information about a vessel's crew size and tonnage (rows 7 and 8), we can assess the proportion of crew members to slaves and the degree of crowding on board the vessel. Row 17 shows an increase in the number of journeys for which we know exact numbers of deaths on board, which augments our ability to accurately calculate mortality on voyages. The more than 6,000 records on slave deaths also provide the largest amount of data on mortality during long-distance migration of any sort that is ever likely to be available. Additional records on the age and gender of slaves (rows 13 and 18) provide some sense of the demographic mix of peoples on board slave vessels and a basis for determining disproportionate mortality rates among men versus women or among children and the elderly versus adolescent or middle-aged slaves. Rows 3 and 4 provide more than 4,000 new names of ship captains and owners that allow us to identify the people of European descent who were involved in organizing and running slave voyages. Given that there are sixteen owners' names variables and three captains' names variables, the new data set actually identifies close to 50,000 separate owners and captains—an increase of about one third over the 1999 data set. One way of looking at these new data is to see them forming a capstone to forty years work beginning with Curtin and Klein's efforts in the 1960s, but in turn, providing a basis for augmentation into the foreseeable future.

Not reported in this table are the new variables that we propose to add to the data set. Information on the prices for which slaves were sold in the Americas is now available for 1,200 voyages and the approximately 250,000 human chattels that these voyages disembarked in the Americas. These data cannot be entered in their raw state because of the variety of currencies in which the sales were recorded and because of the range of age and sex combinations which formed the subject of these transactions. Considerable preparatory work is required to standardize prices and make comparisons possible.

At this stage, we propose to create four additional variables, at least one of which will be an appropriately standardized valuation. A full explanation of how this information will be derived, as well as some preliminary trends and patterns that have emerged from these new data, have already appeared in print (Eltis and Richardson 2004). We will construct a second set of new variables through our links with a related database of 67,000 African names taken from the records of the International Courts of Mixed Commission, which adjudicated slave ships in the nineteenth century (Nwokeji and Eltis 2002). Full descriptions of the Africans found on board such vessels make it possible (with some additional research during the data processing, gathering, and integration stage of this project) to construct ethnic profiles of the catchment area of the regions from which these slave ships drew.

It is imperative that the new, raw data be incorporated into the new Internet-based publication. However, the process of assembling and integrating data from many different sources over a period of several years in a database that will have close to 250 variables involves potentially changing nearly nine million cells of data (35,000 voyages x 250 variables). As a result, the data must be processed, such overlaps and incongruous findings must be identified, and the data associated with these problems must be verified at their sources. Many of these errors can and will be corrected without travel. However, in some cases it will be necessary to resolve problems in the data with an on-site visit. For most of the major collection sites, *i.e.* in Rio de Janeiro, Lisbon, and London, Steering Committee members will carry out this checking, but for other centers, specifically Havana, Cuba, and Seville and Madrid, Spain, we will need to send the Post-Doctoral Research Associate to the collection site in order to complete the data verifications. This task can be conducted with two weeks of work in each of the two countries.

Further, our work has uncovered the presence of three separate caches of documents that could be added quickly to our data set to make it more complete. These are not extensive, but are of great importance. The first is in the Archivo General de Indias, Seville, Spain. There are about 30 legajos (or document boxes) in the following Seccións—Santo Domingo, Indiferente General, and Contaduría. Second, in the Estado Sección at the Archivo Historico Nacional in Madrid, there are about 60 legajos covering the slave trade, mainly to Cuba in the nineteenth century. We have covered half of these in earlier work and completed our coverage down to 1843. Additional support for two weeks would allow us to complete our coverage to 1867—the year in which we think the last transatlantic slave voyage occurred.

Third, we know of 25 legajos in two collections in the Archivo Nacional de la República (ANC) de Cuba in Havana, some of which we have been able to consult. Two weeks of further work would allow us to strengthen and broaden our coverage of the final and illegal phase of the slave trade. We would like to emphasize that we know the location and content of these sources, and have worked on similar boxes in the same locations. As such, we are confident that this work can be completed in the proposed timeframe.

In all three of these cases, we wish to combine data checking with these final data collection trips. In order to complete this work, the Post-Doctoral Research Associate and the Research Assistant will spend the first three months (June-August) of the project intensively processing the raw data. During September 2006, the Post-Doctoral Research Associate will spend two weeks collecting new data and two weeks verifying data gathered in Seville and Madrid, Spain. In October 2006, the Post-Doctoral Research Associate will spend two weeks verifying and two weeks collecting data in Havana, Cuba.

Following the on-site data verifications, the PI together with the Steering Committee Members, the Post-Doctoral Research Associate, and the Research Assistant, will meet at Emory University during November 2006, for one week. During this meeting, they will devise tests to ensure the internal consistency of the data. They will also construct transparent protocols that will be used to develop imputed variables (see Appendix 5). Developing these imputed variables requires considerable preliminary analysis and programming, and will be completed with the SPSS statistical software package. Work conducted during this meeting will ensure the continued scholarly integrity of the database.

The Post-Doctoral Research Associate and the Research Assistant will then prepare the data for the new interface over the next six months of the project (December 2006–May 2007), using the protocols developed in conjunction with the PI and Steering Committee Members. They will also begin to incorporate the data into the master data set, thus readying it for publication.

Production of Contextual Materials

Our design for the generalist/K-12 tier of the website carefully balances two needs: 1) to create a user interface that appeals to both K-12 learners and the general public, and 2) to design clear teaching guides

and lesson plans that teachers can adapt to their state-designated standards and use with the materials available on this interface. This will ensure that teachers and K-12 students will have clear-cut guides to consult as they use this material in classrooms, while also accounting for the desires of adult learners to access and make use of the materials of this tier of the site well beyond the classroom environment.

The K-12 tier will offer a section on teaching materials for classroom use that specifically address the key skill sets that students are to develop at three distinct levels: K-4, 5-12, and 9-12, as designated in the current national standards for K-12 history education. These teaching materials will include lesson plans, activities, and assignments that draw on the resources—the database, the contextual essays, the path graphics, and the GIS mapping tools—provided in the generalist/K-12 tier of the website. We have outlined the standards that we intend to address with this tier of the site in Appendix 3.

The planning and execution of both the site design and the teaching resources will be undertaken in three stages. First, Regina Werum and Carole Hahn (Advisory Board members and education specialists) will consult with Eltis (PI), Halbert (PI), the Web Design consultant, the Post-Doctoral Research Associate, and the Research Assistant at Emory University during the first two project months (June-July 2006) to help design the prototype system of the generalist/K-12 tier of the website. The second planning stage will commence in month seven (December 2006), after the prototype system has been developed. During this stage, Werum and Hahn (Advisory Board members) will consult with Eltis (PI), the Post-Doctoral Research Associate, and the Research Assistant in order to determine what contextual materials will best address the needs stipulated in the national standards for history. They will issue an international call for materials by March 2007, which will include detailed information concerning what elements should be incorporated into the teaching guides, and will spend months eleven through fourteen (April-July 2007) selecting candidates to produce these materials. The third stage will begin in month fourteen (July 2007), when the second prototype of the system will be issued and made available to the final candidates for use as they construct lesson plans, teaching materials (including essays and timelines), and guides to the use of the database, path graphics, and GIS tools. These candidates must submit their materials for review, editing, and approval in month eighteen (November 2007) of the project period.

Editorial Distribution System Design

When complete, the new database website will offer an interactive experience on the slave trade in two senses. At one level, “interactive” simply means the ability of users to search the site, pose questions, and to select and analyze the data that interests them. In addition, however, “interactive” means that users will have the opportunity to contribute to, and thus shape, the database that they are using.

In order to ensure the flexibility and sustainability of this new publication, we will design an editorial distribution system as part of this project. The system will incorporate two distinct components: 1) naming an international team of prominent scholars to serve as the first Editorial Board, setting the duties of this group, and establishing transparent protocols that will be used for the vetting, processing, and integration of future submissions to the data set; and 2) the creation of a software toolkit that will facilitate the editorial process, including submission processing, vetting, and publication, and the tracking of all changes made to the data set by date and by substance.

To begin this work, Halbert and Eltis (PIs) will initially collaborate with Florentino, Richardson, and Behrendt (Steering Committee members) during the first all-project meeting, which will be held at Emory University during November 2006. Eltis, Florentino, Richardson, and Behrendt will comprise the first Editorial Board. During this meeting, the Editors will work in conjunction with the Post-Doctoral Research Associate, the Research Assistant, and the programmers to set the protocols for cleaning, checking, and incorporating new data into the master data set. The procedures and protocols that they create during this meeting will have general application to the website’s editorial distribution system.

These will later be developed by the PIs, the Steering Committee Members, the Post-Doctoral Research Associate, and the Research Assistant into guidelines for other scholars and researchers who wish to contribute to the site, and will be used to facilitate the process of vetting, processing, and integrating new raw data and variables into the data set in the future. The Editorial Board will also use this first all-project meeting to create the standard practices of the Editorial Board, including determining the length of member terms, the number of Board members needed, and the duties that will be expected of the Board. The Editorial Board will also begin to work on a call for submissions and a set of guidelines for such submissions, to be released at the time of the new database's publication.

Following this meeting, the programmers will begin working on a software tool that will enable the Editorial Board to make changes directly to the data set that underlies the Web-based publication of the *Trans-Atlantic Slave Trade Database*. When processed through this software tool, changes will be automatically incorporated into each module of the website—the downloadable SPSS file, the viewable database component, and the path graphics and GIS applications. The seamlessness of this updating will be necessary to ensure the publication's sustainability over time. The software tool will be designed such that it will store information regarding all changes that are made to the database, enabling researchers to view the database as it appeared on specific dates, and to see *exactly what changed and when*—information that rarely accompanies Web-based publications. Such information is imperative as it allows researchers to verify scholarly citations from particular dates with the system.

The programmers will also work with the PIs, the Steering Committee, the Post-Doctoral Research Associate, and the Research Assistant to begin designing the editorial distribution system between November 2006 and the second Steering Committee meeting at Emory University (July 2007). This system will lay out each step of the submission and publication process. It will include a management system that tracks 1) the uploading of a new submission; 2) the receipt of this submission by the Editorial Board; 3) its peer review; 4) its acceptance or rejection; and 5) its publication, where applicable. The system will automate and record all correspondence that takes place between the submitter, the Editorial Board, and the referees. (For an existing model that is similar to our proposed management system, please see the Public Knowledge Project's Open Journal Systems, http://www.pkp.ubc.ca/ojs/). During the July 2007 meeting, the Editorial Board will test this system and suggest modifications as needed. The programmers will complete these refinements to the system between July 2007 and April 2008.

Storage and Maintenance of Data

The Web-based publication of the database will be stored and maintained in the Woodruff Library of Emory University, the lead institution for a National Digital Preservation center established by the Library of Congress (http://MetaArchive.org). The database will be archived through a secure distributed digital preservation system using the LOCKSS software created by Stanford University.

Halbert (PI) and the programmers will work from August 2007 to April 2008 to create and establish a secure preservation network system that replicates the data repository, deploying it at Hull University in the U.K., Victoria University of Wellington in New Zealand, and at the Universidade Federal do Rio de Janeiro in Brazil, the base universities of our Steering Committee and long-time affiliates of the *Trans-Atlantic Slave Trade Database*. The system will regularly check the data integrity of the files at each location, and will also assist in maintaining the data during future format migrations. The creation of this system will ensure the database's long-term preservation. The system will be deployed in April of 2008.

History, Scope, and Duration

Transatlantic Slave Trade Research

As noted, the transatlantic passage phase of the re-peopling of the Americas is better documented for Africans than for Europeans, as most transatlantic slave voyages left behind some written record of their activities. This proposal grows out of more than four decades of data collection in international archives concerning these transatlantic slave voyages. From the mid-1960s, researchers have used such records to identify thousands of voyages and to encode them into richly detailed individual data sets.

Each of these early individual sets was based upon the records of one specific European nation or one particular port from which slaving voyages originated. The available information was fragmented; therefore, the individual data sets reflected the nature of the records that had survived rather than the structure of the voyage they described. Further, scholars of the slave trade had spent the first quarter century of the computer era working largely in isolation of each other, each usually using only one source, and recording its contents in idiosyncratic, rather than standardized, formats. The Mettas and Richardson collections were the only significant early exceptions to this pattern (Mettas 1978; Richardson 1986-1996). Voyage records that combined data from several national or port archives were unusual.

By the late 1980s, Eltis (PI), Klein (Advisory Board), and Richardson (Steering Committee) had encoded approximately 11,000 individual voyages in sixteen separate sets. However, when they began comparing and analyzing the data, they discovered that much of these data overlapped in content, that not all of the voyages were transatlantic, and that some were not actually slave voyages. In addition, several listings of voyages that had been extracted from more than one source had also appeared in hard copy form (Mettas 1978; Mettas and Daget 1984; Richardson 1986-1987; Daget 1988). It became apparent that the multi-sourced data contained duplicate and inconsistent records, and would require extensive remediation before it could be combined into a central data set. Even so, the idea of creating a single multi-source data set of transatlantic slave voyages seemed an obvious step forward from these individualized efforts. In July 1993 such a project, led by Eltis and centered at Harvard University, received funding from the National Endowment for the Humanities with supplementary support from the Mellon Foundation. By the time the project began, Postma's Dutch data had become available, as had Behrendt's compilation of the extensive British trade after 1779, and also the large, complex Richardson, Beedham, and Schofield pre-1787 Liverpool Plantation Register set, all of which were held in machine-readable formats. Quantities of smaller sets of published material available only in hard copy form were encoded for inclusion, and scholars volunteered their own unpublished data for encoding as awareness of the project increased.

Between 1993 and 1998, a Harvard-based project to assemble and enlarge the available data on transatlantic slave-trading voyages moved forward with four specific objectives: 1) standardize in one voyage-based data set the information collected by scholars over the previous 25 years on the shipping of slaves; 2) encode additional slave voyages; 3) ensure wide dissemination of the results of the work to specialists and non-specialists; and 4) use the database to initiate a new era of research on the cultural and social history of the movement of peoples from Africa to the Americas. These objectives were realized. After integrating about 11,000 voyages from eight existing data sets, the editors searched primary sources around the Atlantic basins and assembled a data set of 27,233 voyages covering perhaps two-thirds of the transatlantic slaving expeditions ever made. This comprised the data set published in the 1999 CD-ROM. It is reasonable to claim that the historiography of the slave trade divides into two phases, one before and one after the creation of the data set (see the *William and Mary Quarterly*, January 2001).

The major gaps in the depth of coverage in the 1999 database were in the Portuguese, U.S., and to a much lesser extent, the Hispanic slave trades. More slave ships sailed under the Portuguese, and later, the Brazilian flags than that of any other nation, yet the data set contained only 6,183 records of such vessels

compared to nearly 11,000 British voyages. The only substantial part of the slave trade based in the U.S. that the data set covered was the Rhode Island trade, comprising perhaps half of the U.S. total. Moreover the depth of coverage of these branches of the traffic, as reflected in the number of variables completed and the range of sources listed for each voyage, was weaker than what researchers had been able to find for the British, French, Dutch, and Danish slave trades. As a consequence, in 2000, Richardson and Eltis launched a successful application for funding to the AHRB of the United Kingdom, with the University of Hull as the designated host institution. This support enabled the elimination of most of the aforementioned gaps in coverage. In the past four years, archival research in Rio de Janeiro, Salvador, Luanda, Benguela, Lisbon, Mozambique, Seville, Amsterdam, and a variety of U.S. cities, has added an additional 7,501 transatlantic slaving voyages to the data set, most of them Portuguese.

The density of information and sources for the Portuguese slave trade is beginning to rival that of the better-known trades of the north-western European nations. In addition to the new voyages, the AHRB grant has made possible major editorial work on the 1999 collection of voyages, including the elimination of 598 duplicate records. Additional information will have been added to nearly 9,000 existing records of slave voyages, or one third of the 1999 data set, by Spring 2006. We have accumulated information from several hundred copies of slave voyage logbooks that indicate the routes and places of trade in Africa. As noted in "Significance," the implications of this activity for the African end of the business are substantial given that we now have some information on the African region of trade for 26,526 slaving expeditions (See Appendix 2, row 11).

Further gaps were discovered in the width of coverage in the 1999 database publication. In particular, Eltis and the Steering Committee received repeated enquiries on two major topics—the ethnicities of the captives and the prices at which slaves were bought and sold. Eltis, Richardson, Florentino, and Behrendt have been addressing these issues. They will build imputed variables with the Post-Doctoral Research Associate and the Research Assistant during this project in order to integrate information on prices and ethnicity into the database. The nation/ethnicity/language issue has proven more challenging, but we have begun to work with these data in two ways. Descriptions have survived for 67,000 people on board 295 slave vessels, including their African names, scarifications, ages and sexes, and for many, their nations or countries as well. A separate project is in the process of encoding and interpreting these data. In collaboration with Advisory Board member G. Ugo Nwokeji, these data will be linked with the database during this project, enabling users of the new database to bring up the African names and estimated ages of these 67,000 people, and to see also modern estimates of the ethnicities of these individuals (Nwokeji and Eltis 2002). For other vessels we will be able to infer origins of slaves using data gathered at the regional rather than the voyage level. By 2008, we will reconstruct and incorporate information into the database about broad language and nation groupings in ways that would have been considered impossible a decade ago.

The ultimate significance of this proposal lies in bringing this new, enormously enhanced data together with technological and conceptual strategies to broaden the database's user base and to establish an organic, dynamic publication that will live and change well beyond the date of its initial publication (indeed, the very terms "publication" and "re-publication" appear irrelevant in this context). As already indicated, we aim to engage audiences with a wide range of sophistication and knowledge bases at both user and input ends of the operation. The new database will permit scholars to carry out new work on a wide range of topics, making the issues of African agency and the cultural exchange between Africa and the Americas even more central to Atlantic history than they are at the moment. On the slave trade itself, we expect the debate on the dimensions and the impact of the slave trade (in all continents bordering the Atlantic basins) to be sharpened as well as broadened. But we also expect the database to become a basic reference tool in fields of study far beyond the Atlantic. From studies of ethnicity to business history, many topics in early modern European, American, and African history that have little or nothing to do specifically with the slave trade will be able to draw on this resource.

Methodology and Standards

To review, the major goal for this project is to promote scholarly and public awareness of the role played by the African slave trade in shaping the development of the Atlantic world. In order to achieve this goal, we propose to undertake five major tasks in this project: 1) create a significantly enhanced, two-tier, Web-based *Trans-Atlantic Slave Trade Database* resource that is freely available to the public; 2) to prepare and process the data that has become available since the original 1999 CD-ROM publication; 3) to publish contextual materials that can be used in conjunction with the database to teach history to K-12 students; 4) to create an editorial distribution system that will facilitate continual updates to this resource in the future; and 5) to establish a secure distributed archive for the storage of this digital publication.

As we transform the database from the hard-copy CD-ROM format to an open access website, ease of distribution and availability need to be combined with simplified, yet more powerful interactive software. The new software will make data selection much less cumbersome, incorporate a greatly enhanced mapping facility, and offer greater flexibility in the analysis of output from any given selection. A part of the new software will also be aimed at the K-12 educational segment. All of the development work we will undertake in completing these modular components of the site will occur in conjunction with a highly qualified group of Steering Committee Members and an Advisory Board with a range of strengths, skills, and experience in designing and implementing Web-based education projects.

All data incorporated into the database since the process began in 1993 are taken from the public domain. There are accordingly no issues of intellectual property in the creation of a new database. Cambridge University Press held the rights to the 1999 CD-ROM, but has relinquished those rights to David Eltis (PI) and Emory University. The starting point of the present proposal is thus the CD-ROM interface.

Procedures and Standards for the Current Application

Designing the Interface

The core of this application is the creation of a flexible Web-based infrastructure to deliver and analyze data via a redesigned interface. The existing CD-ROM interface will be the starting point for this infrastructure, but we intend to expand significantly the capabilities of the CD-ROM through new query capabilities and ongoing enhancements. Establishing effective mechanisms for ongoing maintenance of the database is a necessary element in this infrastructure. We intend to develop a comprehensive means of aggregating and vetting additional records and a means of secure distributed digital preservation.

A set of functional requirements has been developed for the online database, informed by several years of study by Eltis and colleagues, as well as systems analysis undertaken at Emory University and Cambridge University Press concerning the needs for such a system. Our proposed software design incorporates a set of modular components that will be programmed using only open technical standards and open source software systems. Each module will draw upon an underlying MySQL database in which all of the data records associated with the transatlantic slave trade will be stored. This will enable the overall system to continually adjust to any updates that have been made to the underlying MySQL database. Further, when updates are made to the system by the Editorial Board, the MySQL database will track those changes by date, and will enable the Editorial Board to make notations concerning what changed and why. All former versions of the database will be maintained within a live archive, enabling researchers to access the database as it appeared on any given date in the past.

These software components are described below. A system schematic is included as Appendix 6.

System Modules

A Query Module will allow search functions equal or superior to the capabilities in the original CD-ROM. This system will be constructed with PHP and Perl scripts, utilizing an underlying MySQL relational database. Searches will be enabled by time period, region, and other filters associated with metadata elements of the database. The metadata elements for the database will include all of the elements of the original system plus the new imputed elements created in the course of the project.

A Mapping Module with map generation and Flash animation display capabilities will allow for multi-tiered hotspots, zooming, and data interface with the main query engine. The popular MapServer package from the University of Minnesota will be used for this module. This system utilizes open standards for GIS applications and is also open source software, which will enable straightforward integration with the query module by the project programming team. The MapServer application has strong support for both Flash animations and multi-lingual display. This module will also include a "path graphics" capability, enabling the display of volume of voyages from one region to another.

A combined Help and FAQ Module will be provided. This module will include a dual level help component (professional and generalist) that will be written and implemented contextually. The help system will also include documentation concerning the data codebook, data provenance/preparation, and other information about the database relevant to researchers. Additionally, it will include a FAQ of key questions asked by new users about the nature of the database and the interfaces.

A Supplementary Materials Module will provide access to images and textual content pertaining to specific voyages, including navigational information (where available) that can be mapped dynamically. This module will enable the display of information from selected logbooks, showing the relevant entry information associated with particular locations during a ship's voyage. This capability will draw on the functionality of the Mapping Module.

Two Graphical User Interfaces (one for K-12 level student use and the default interface designed for university-level research with options for more advanced analysis). Each interface will include appropriately modified versions of the above system components. These interfaces will be implemented using cascading style sheets (CSS) for maintainability.

An Editorial Administration Module will automate administrative functions, including uploading of new records, data review and integrity checks, and database maintenance functions. This software component will allow users to submit new information to the Editorial Board. A procedure will be developed for uploading the new data after the Editorial Board has approved the proposed changes or additions. This procedure will include evaluation of the sources from which the new data have emerged, approval of the impact of the new data on the imputed variables, and the creation of an open access and permanent record of the changes made to the original database as a result of user submissions. A central feature of this procedure will be the maintenance of a copy of the original Web-based database and every subsequent revision, each of which the interface can access at any time. Thus, users will always have the option of selecting the latest database of voyages for their analysis or to return to any earlier version (including the unedited version as originally launched), and will have a sense of the growth of the database over time.

Preparation and Processing of Data

Over the last six years, Eltis (PI) and the Steering Committee Members of this project have continued to gather data following the procedures and protocols that they constructed in the late 1990s. These new data have presented opportunities for significantly revising, broadening, and enhancing the data set published in 1999. During the production of the Web-based *Trans-Atlantic Slave Trade Database*, the 1999 data set will be changed in three ways. First, many of the blank cells that appear in the existing records will be filled with new data collected since 1999. Second, additional voyages will be added, as described above. This will lengthen the database from 27,233 voyages in the 1999 version to 35,000 by the end of the grant period. Third, additional variables will be developed, adding substantial new width to the database. We estimate that we will add at least seven new variables to accommodate the new information we have

gathered, particularly on the prices of slaves sold from slaving expeditions and on the ethnicity of slaves carried. The rules and standards for adding data to the data set are carefully spelled out on pages 6 to 27 of the "Introduction" to the 1999 CD-ROM, a copy of which forms Appendix 4 of this application.

The fundamental data structures utilized for the presentation of slave trade information were established during the creation of the 1999 database. This project will build upon and extend these data structures to incorporate new data and variables into the master data set. There are two essential data preparation tasks. The first is to devise tests to ensure the internal consistency of all the *data variables*, which incorporate information drawn directly from archival and published sources and comprised 162 of the total 226 variables in the published data set. Additions of and changes to almost 16,000 voyage records (involving approximately 300,000 separate cells) will be made to the original database. As with any multi-source data set, the new data will inevitably disrupt the original data set. During the first three project months, Eltis (PI), the Post-Doctoral Research Associate, and the Research Assistant will develop a transparent set of protocols that will be employed to process and test the integrity of the data prior to its inclusion.

The second task relates to constructing the *imputed variables*, which are inferred by the project researchers from their knowledge of the voyages or calculated directly from data encountered in archival or published sources. The 1999 data set contained 64 imputed variables that were created to make the data set more accessible to users. These imputed variables were used primarily to construct tables, graphs, and path graphics within the CD-ROM publication, and were heavily used by the researchers who consulted the data set. Developing these imputed variables, which comprise about one third of all the 1999 variables, requires considerable preliminary analysis and programming. While some of the 1999 tests will be re-used, the extra material uncovered and entered since then makes possible more refined inferences. The SPSS "program to create imputed variables" that the project team developed and used for the 1999 publication will be modified and enhanced for use in data preparation in the present project. An annotated excerpt from this SPSS program is included as Appendix 5 of the present application.

In addition to these central tasks, cleaning and checking procedures will be required, as for any large data set put together from many different sources. These will be conducted by the Post-Doctoral Research Associate and the Research Assistant in an ongoing process throughout the two-year project period.

Developing Contextual Materials

The generalist/K-12 interface of the website will contain supplemental materials geared toward the needs of teachers and students. Our proposed design will produce teaching guides, lesson plans, contextual essays, and specific assignments that students can complete using the database and the mapping capabilities of the site. To ensure the applicability of the contextual materials to the K-12 audience for which they are intended, we will follow the National Standards for History as presented by the National Center for History in the Schools (http://nchs.ucla.edu/standards/toc.html) and described in Appendix 3.

In order to ensure our compliance with national standards, we will work with Advisory Board members Werum and Hahn as we plan and implement this portion of the project. The tasks associated with this occur in three stages. *Stage one* will begin with the project in June 2006. At that time, Werum and Hahn will consult with Eltis (PI), the Post-Doctoral Research Associate, the Research Assistant, and the Web Design consultant in order to plan the design for the prototype K-12 interface. During *stage two*, beginning in December 2006 when the first prototype of the system will be available, Eltis, Werum, Hahn, the Post-Doctoral Research Associate, and the Research Assistant will: 1) produce a blueprint for the creation of the contextual materials; 2) issue a call for materials to the education community; 3) review submissions; and 4) select candidates to produce the materials. As the second prototype is issued in July 2007, *stage three* will begin. The prototype will be shared with the selected candidates as they begin to create lesson plans, teaching guides, essays, and assignments for the site. The candidates will

submit their materials for review, editing, and approval by Eltis, Werum, and Hahn by November 2007, and the contextual materials will be posted to the site in December 2007.

Developing the Editorial Distribution System

A workflow for the submission, editorial review, and acceptance of new records from international contributors will be developed in the course of the project. There are many potential contributors of slave trade records throughout the world, and their assistance in continuing to grow the database will be invaluable in coming years. Contributed records must be peer reviewed for accuracy and de-duplication by responsible experts before addition to the database. Further, imputed variables must be interpolated for purposes of data quality. We will develop an editorial process whereby Eltis and the Steering Committee will be able to review and approve new submissions to the database through an administrative interface. Crafting this editorial process and beginning to use it is an essential part of the project, and requires a face-to-face collaborative meeting of Eltis (PI) and the Steering Committee in November 2006. These Steering Committee members will comprise the first Editorial Board of the project.

Development of this system will begin with modeling of the current practices for adding records to the database, will include systems development of submissions workflow in the administrative module, and will conclude with the Editorial Board's use of the system to mediate new additions to the database. The system will comply with current double-blind peer-reviewing strategies. The submitter will remain anonymous to the reviewers, and vice versa. It will also be constructed such that approved submissions will be added directly to the database that underlies the entire dynamic system, allowing the changes to take effect within each component of the dual-tier interface. The system will track the date of each change made to the data set, and will archive all previous versions of the data set by date.

Developing the Secure Preservation Network

The database will be archived at the Woodruff Library of Emory University, the base of one of the two project PIs. Emory is the lead institution for the MetaArchive project, which has developed a low-cost technological solution for secure distributed digital preservation using LOCKSS software. This technology will address long-term preservation concerns for the new data repository. A secure preservation network system for replicating the data repository will be deployed under the guidance of Steering Committee members at Hull University, the Federal University of Rio de Janeiro, and Victoria University of Wellington in the second year of the project. This strategy will ensure that the database will be archived at multiple geographically distributed locations. The secure preservation network system will feature automated techniques for data integrity checks and format migration. As an NDIIPP lead institution, Emory is well-prepared for the requirements of long term preservation of digital information.

Prototype Construction

Two prototypes of the overall system will be constructed during this project. The first will be released in November 2006, and the second will be released in July 2007. Both prototypes will be produced under the guidance of Halbert (PI) by the programmers, the Web Design consultant, and the GIS consultant. The construction of these two prototype systems will serve two related purposes. First, they will enable the team to test the system with the project staff, as well as the Steering Committee and the Advisory Board prior to its launch in 2008. Second, they will enable the creation of contextual materials by the candidates who are selected by Werum and Hahn, the K-12 specialists of the Advisory Board. This second prototype system in particular is imperative for the construction of these supplementary materials, as the teaching guides, lesson plans, essays, and assignments are expected to be designed for use with the database, as well as with the GIS and path graphics tools.

The first prototype will include early versions of: 1) a two-tier Web-based interface; 2) the underlying MySQL database system; 3) the downloadable SPSS database file; 4) the GIS and path graphics module;

and 5) the editorial distribution system. This prototype will be tested for usability in small focus-group sessions in July 2007 with an advisory group, as described below. The second prototype will include the same basic features as the first prototype, but all of the modular components included in the prototype will have been enhanced and further developed during the interim period. Again, the prototype will be tested for usability by a group of advisors, including the project staff, the Steering Committee, and the Advisory Board, as described below. We also expect to gain valuable feedback from the specialists who prepare the teaching materials for the K-12 tier of the site.

Project Evaluation

The project has been designed to include several evaluative components that will commence after the production of each of two prototype systems. As mentioned above, the first prototype will be issued in November 2006, and the second, in July 2007. We plan to make the first prototype available to the Steering Committee, the Advisory Board, and our extended project staff members in November 2006 for testing and feedback. Halbert (PI) will also give a presentation and demonstration of this prototype during the Spring Forum of the Digital Library Federation in 2007 in order to gain feedback from the extended digital library research community. All feedback will be documented in project reports, and where appropriate, will be used to modify the prototype for its second release.

In July 2007 (the second year of the project), we will engage the project staff, the Steering Committee, and the Advisory Board as advisors to test the second prototype system. These usability tests will be hosted by Halbert (PI), Werum (Advisory Board member), the Post-Doctoral Research Associate, the Research Assistant, and the programmers. They will be designed as two small focus-group sessions, one to take place at Emory University with the Steering Committee and project staff members during the second collaborative group meeting, and the other to take place via conference call with all Advisory Board members. During each session, advisors will be asked to provide guidance as they complete a sequence of structured activities. They will first be asked to complete a list of ten tasks, including simple and complex searches, database queries, and mapping exercises. This task list will be constructed such that it should take approximately 10-15 minutes to complete. We will track the movements of the advisors through automatically generated user logs. After the advisors have completed these tasks, they will be asked to provide feedback concerning the interface and the operability of each module through a 15-20 minute group discussion with the project staff members who are proctoring the session. The feedback and discussion of each session will be recorded. Feedback provided in these sessions will be written up as a project report and will be used to appropriately modify the system wherever needed.

In July 2007, we will also make this second prototype system available to leading scholars in transatlantic slave trade history, to the candidates who have been selected to create the contextual K-12 materials, and to museum directors and curators who express interest in creating installations around this resource (see “Dissemination”). We will track all site usage through automatically generated user logs, and analyze these to improve the navigation system where necessary. We will invite voluntary comment about this prototype system from all users through a feedback system. Similarly, the initial launch of the Web-based resource in 2008 will track user movements with user logs that will be analyzed by Halbert, the Post-Doctoral Research Associate, and the programmers to assess the usability of this site. This launch will also possess a feedback system that will allow users to provide voluntary comments about the resource.