

This proposal was submitted to the National Science Foundation in 2000. It outlines an approach to the design of a world history databank, in which systematic work of collection, transformation, and display of comparable data for the world as a whole (over as much as five centuries) would lay the basis for still larger scale research. The results would provide data permitting identification and analysis of long-term, large-scale patterns in commerce, demography, and family life.

The proposal was not funded, but I believe the need for such research has only grown with time, and hope to find collaborators who will join in developing this project.

Patrick Manning



World History Center

Department of History
Northeastern University
Boston, MA 02115
(617) 373-4060
www.whc.neu.edu

World History Databank

World historians and social-science analysis.

Sociologists, historians, anthropologists, political scientists and economists are becoming more global in their perspectives. Globalization studies, transnationalism, and comparative studies are all changing the way we see the world and understand social processes. Yet the nation-based frameworks of previous studies, combined with the difficulties of converting and comparing data, mean that empirical verification is falling behind theoretical developments.

This project will assemble world historians and other social scientists into a collaborative network, to create a World History Databank addressing the past four centuries in systematic and global format. Three categories of data have been selected for emphasis because of their relevance to major interpretive issues and because they are relatively available and tractable in analysis: trade volume and value, migration flows, and family size and structure. A fourth set of data provides the conversions enabling the comparison and aggregation of the first three: it includes the widely varying weights, measures, regions, and other dimensions of the principal data. Project staff and external collaborators will work systematically to collect, transform, display, aggregate and analyze these data in world history.

World historians. World historians are social scientists specialized in addressing problems in broad regional and temporal scope. We consider the past in geographical units broader than nations or civilizations. Extending this practice of looking beyond boundaries, world historians also look across boundaries in time periods, in interpretive themes, and in social-scientific scope. Through their experience in considering a broad range of phenomena and viewpoints in time perspective, world historians are well equipped to join and sustain discussion with other social

scientists. Among the issues under debate in the field of world history are the long-term implications of the development of agriculture (Diamond 1997), the connections of religion and commerce across Eurasia (Liu 1996), the spread of disease and plants with the opening of global maritime contact (Crosby 1986), the rise of nationalism (Hobsbawm 1992), national and regional competition for global economic leadership (Frank 1998), the acceleration of migration (Cohen 1995), and the transformations in modern families.

Global analysis. The social science disciplines, in turn, have embarked on analysis in global and historical perspective. Each of the major social-science disciplines has moved, in its own way, from comparison of isolated case studies toward study of interactions in global analytical perspective. Economics has moved from national accounting (Kuznets 1971) to estimation of global economic interactions through computable general equilibrium models (O'Rourke and Williamson 2000). Demography has moved from stable-population analysis (Coale and Demeny 1966) to a focus on migrations. Family history studies that began with a focus in local context, abstracting from migration (Laslett 1968) now seem poised to explore global comparisons and patterns in family structures (Canny 1998).

Improved computational tools and improved access to historical data combine to facilitate this move to the study of interactions. Computers can handle data sets that are large, complex, or both. Spreadsheets and database programs, though a relatively basic innovation, bring sophisticated data analysis to the ordinary user: life tables, for instance, are calculated almost effortlessly with spreadsheets.

Time perspective. Social scientists, in turning to address social interactions, have begun with a focus on the present. In interpretive work within each of the major social science fields, a form of the "globalization" thesis has come to the fore: observers hypothesize that the contemporary globalization of society is new and governed by contemporary technology. Such a thesis cannot be sustained, however, unless tested by a global analysis of the past. For these purposes, we need research that is long in time-frame, global in its reach, and interdisciplinary in its thematic emphasis.

Hurried adoption of the globalization thesis provides an example of generalizing from available data rather than seeking out relevant data. While the formulation of social science debates emerges principally from the conflicts of our own time, assessing current hypotheses and their implications requires posing the same questions across space and time. If the spread of McDonalds's restaurants today suggests a diffusion of American corporate capitalism, the spread of Thai restaurants today and of Chinese restaurants in earlier times suggests that there is more to globalization than Americanization. The task of world historians is to locate and develop the relevant data for testing major social science hypotheses. This task involves reaching across large

spans of time and space, to locate, produce, and transform the relevant data for addressing major theses.

Global data sets. Global historical analysis is still conducted, principally, through individual assemblages of local and national data. A few individual scholars may be credited with signal achievements in assembling large data sets (Mitchell 1998), but in general the creation of systematic global sets of data exceeds the capacity of individual researchers. Some collaborations have already created significantly expanded data sets (Eltis et al 2000). Major institutions have been developing global collections of data for recent times, relying on efficiency of modern governments in record keeping, and on contemporary categories (World Bank website). To develop equivalent data for the past requires substantial investment and collaborative work.

This is an excellent time to launch this program of global historical database-building. Existing collections on the late twentieth century are strong for economic and demographic data. The improved access to archival collections in Russia and in the countries of Africa, Asia, and Latin America will facilitate the collection of new data. Archival collections in Lisbon, Seville, Amsterdam and in various centers in India and China hold immense quantities of unretrieved data. In addition, the increased ease of photocopying and electronic scanning eases data collection.

Objectives: data for global historical analysis

The World History Databank is to create sets of consistent historical data, readily usable for long-term, trans-regional, and trans-disciplinary analysis. The project addresses and links current debates in world history, and will nurture institutions and collaborative networks for world historical research.

To develop historical data, as contrasted with survey data or experimental data, the researcher must accommodate to the constraints of previously defined data, and then transform them into data sets that are internally consistent and relevant to the problems under study. To illustrate the issues and problems in developing data for world historical analysis, here are some examples.

To assert that over 10 million Africans crossed the Atlantic in slavery — or that over 50 million Europeans crossed the Atlantic from 1840 to 1920, or that more than 40 million people left China and India in the same time period — is to make a broad but imprecise generalization. To provide data for a more useful level of argument, a World History Databank must identify migrants consistently by place of departure and arrival, by time period — and when possible by age and sex and mortality. Similarly, to argue that most silver mined in the seventeenth and eighteenth centuries went to China, or that the volume of wheat on the world market doubled during the nineteenth century and then leveled off, the underlying data must be organized consistently by region, and summarized in consistent weights and measures. Further, to claim that families have

become smaller with the rise of migration and urbanization (despite the decline in mortality), we must have systematic ways to document and interpret the social boundaries of families, and ways to sample the range of family size by region and by time period.

In order to create appropriate global historical datasets, drawing together data from publications and archives of nations, empires, localities, and international organizations, and from records of firms and individuals, the project must complete four major functions: collection, transformation, display, and analysis of historical data. Collection means retrieving and accumulating historical data, more than creation of new data. We will record “raw” data in the form we find them, whether the data are “new” (newly retrieved from archives or newly estimated) or “existing” (data previously published or held in known data collections). Transformation means rendering the data in each category as strictly comparable as possible, by developing conversions in weights and measures, price indices, and by linking data to standardized regions and time periods. An ability to compare family size and precious metals trade in seventeenth-century Philippines and twentieth-century Ghana will greatly strengthen our understanding of world-historical processes.

In making decisions for the implementation of the project, we will emphasize the following priorities:

- *Overcome fragmentation and inconsistency in data sets.* For large political units (U.S., China, India, Brazil, Russia) we will organize data into sub-regions.
- *Address the full range of world regions.* In particular, we will emphasize collecting data on Africa, Asia, and the Americas, as well as on the North Atlantic.
- *Address a long time frame.* To the extent possible, we will extend data sets from the twentieth back through the seventeenth centuries.
- *Seek out links across disciplines and themes.* Data on commerce, migration, and family should be structured to permit links among these themes and also to data on environmental and political, and other issues.
- *Annotate and update data.* All data will be linked to relevant entries in the catalog of sources, and updates will be entered promptly.
- *Ensure ease of access to and commentary on data.* Web display and a detailed site map will make data available with little delay. Project participants and other users will be able to offer comments and corrections, thus increasing the accuracy of the Databank.
- *Allow flexible aggregation.* Analysts working with the Databank will be able to define a wide range of units (regions, nations, continents, ocean basins, or nodes in networks) in which to aggregate the data and conduct their analyses.

- *Stimulate development of new lines of interpretation.* As new global or regional patterns appear in the course of research, the project will adjust its work to elucidate these issues.

Building on previous projects. This project echoes some large-scale projects launched in earlier years. The work of the National Bureau of Economic Research developed estimates of national product for the U.S. and other nations for the twentieth century and in some cases for earlier times (NBER website). The Human Relations Area Files resulted in compilation and coding of a large amount of ethnographic data, in cross-sectional perspective (HRAF website). The Inter-University Consortium for Political and Social Research became a valuable electronic archive for local and national studies performed in various social sciences (ICPSR website). These structures, created a generation and more ago, have provided immense service to social-scientific research, but remain constrained by the national, ethnic, and disciplinary paradigms in which they were created.

Now, in the era of globalization — with new problems, new theories, new data structures, newly available data, and a developing global perspective — it is wise to invest in creating the institutions to study social-science issues at the global level. Little in the way of research and institutional funding has been invested in global study, especially that with a historical dimension.

While most of the new work of world historians, as cited above, has been conducted by individual scholars, collaborative work and research centers have begun to emerge. The World History Center, founded in 1994, is a substantial research center that maintains contacts with world historians and with globalists in other social sciences. This data-gathering project is intended to link the center to other existing and nascent institutions and facilitate the generation of research at a trans-regional and trans-disciplinary level consistent with the new developments in theory.

Global historical studies have advanced to this point without major funding for research. The experience of area studies research, beginning in the 1950s, suggests that new types of research can develop far more rapidly if some strong centers are given substantial resources. It is unlikely that there will ever be global-studies equivalents to the massive, single-campus area-studies centers of Columbia, Wisconsin, and Yale, a system of national and international collaboration among strong and well-funded nodes of global analysis will accelerate the pace of research and analysis in global studies. The World History Databank, based on the unique development of the World History Center, provides an opportunity for launching such a system.

Organizational and programmatic priorities. This project is *collaborative*: it conducts work that is cross-institutional, cross-national, cross-disciplinary, and utilizes multiple methodologies. The value added by this collaborative model of social science research gives it significant advantages for large-scale research, in contrast with the models of Sole Investigator or Isolated Team. It is *web-based*, emphasizing wide distribution and immediate access to its results.

The work is **inclusive** of many leading research groups, and it is **distributive** in making its results immediately available on the web. Working in this structure yields (1) clear and agreed-upon goals, (2) principles to guide working groups, (3) conscious creation of teams. The virtual working groups can develop close collaboration even when they are not at a single site. The project is **specific** in defining the level and dimensions of all data collected and analyzed. The attention to the varying levels of data and processes for handling data — definition, methodology, collection, management, and analysis — will yield improved clarity and usability of the results. The dissemination of preliminary results during research effort will lead to early correction of errors. The project is **structured** to gain the benefits of coordination at its administrative center and autonomy of the various working groups collecting and transforming data.

The Databank project will function through four **virtual working groups** (for data in economics, demography, family, and conversions), plus the infrastructure group (conducting web display and data analysis) and overall project administration. Each virtual working group includes **staff** (mostly at Northeastern) and **collaborators** (mostly at other institutions in the U.S. and beyond). Staff in the research groups includes a post-doc and one or two doctoral students; staff in the infrastructure group includes professionals in web operation. Those who agree to participate as Collaborators will contribute, discuss, and analyze data. Most discussion will take place on the web and through other group-ware.

Debate and resources in world history.

The field of world history has expanded steadily since the early 1980s, and has now established a firm place in research, employment, and teaching in departments of history. (Globalists also have appointments in departments of sociology and other fields.) More than the addition of another subfield to a large discipline, the rise of world history is central to the current reevaluation of historical studies, in which working across disciplinary and geographical boundaries is now becoming the rule rather than the exception. The *Journal of World History* has become a major publication, the *American Historical Review* now devotes significant attention to world history, and the *H-WORLD* e-mail list has 1500 subscribers. Research programs with doctoral study in world history have grown up at Northeastern, Hawaii, Ohio State, Johns Hopkins, San Diego State, and Minnesota, with several new programs forming, including a consortium at the University of California. The dramatic expansion in teaching world history at high school and college levels has helped stimulate further research. The full scope of world historical studies ranges from earliest times to the present. For the Databank, however, it is appropriate to concentrate on the last four centuries, both to achieve consistency of data and to focus on the three historical issues we have selected.

1. Economy. Economic growth has returned to the forefront of discussion, this time linked to issues of convergence and divergence among regions and within regions. Thus, debate about the relative rates of growth of Asian and Atlantic economies today is linked to debate on their growth from the sixteenth to eighteenth centuries. Flynn and Giráldez (1997) have emphasized the early linkage of Asian and Atlantic economies through silver flows, while Pomeranz (2000) and Frank (1998) have argued that Asian economies remained stronger than those of Europe through the eighteenth century, though Landes (1998) and others have contested this view. For the nineteenth century, O'Rourke and Williamson (2000) have argued that a wave of globalization brought convergence among the economies of Europe and North America, followed by divergence in the early twentieth-century era of relative autarky.

The debates on these two major and related questions — including the relationships among long-term growth and cyclical economic change — can be addressed by systematic collection of relevant commercial data. The Databank will include data on prices and volumes of key commodities (precious metals, iron, textiles, grains) and wages.

2. Demography. The study of global patterns of migration, and not simply movements at the national and regional level, will reveal patterns in world history that are scarcely suspected. International migration in the late twentieth century may not have reached the levels of the beginning of the century, but the number of people moving to cities (often within their own nations) created a global level of migration that was unprecedented. The “Great Migration” of 1840-1920, in turn, is seen to be more than doubled in magnitude if one adds, to European emigrants, those moving across recognized boundaries in Asia, Africa, and Latin America in that period. The trade in African slaves was the dominant migratory movement of the eighteenth century, but detailed analysis may reveal other important migratory movements in that and earlier times.

What can be learned from studying world migration patterns? A consistent summary of migratory flows in the overlapping eras of global maritime encounters, of slave trade, and of voluntary migration will give a new and more comprehensive view of relations among migration and economic growth. Such data will make it easier to distinguish among past regional and global networks of migration, and their impact on social and economic change. The Databank will include gross and net flows of migrants among regions, and populations within regions, with detail by age and sex where possible.

3. Family. Of the many dimensions of society and social history, the family is perhaps most basic and most important to document. Families have generally been treated as local phenomena, though individual families and the patterns of family life stretch around the globe and have global implications. The results of European family studies summarized by Watkins, Laslett (1968), and Wrigley and Schofield (1981) show the potential of family history for large-scale interpretation. Thus, Watkins' *From Provinces into Nations* (1991) indicates that the inverse correlation of wealth

and fertility, so obvious at the aggregate level, was far less pronounced at the local level. In addition to this discussion in the demographic literature, there are significant debates on family history in the literatures on anthropology, sociology, and history.

Is there a global history of the family waiting to be articulated? Two simple arguments suggest a response in the affirmative. First, the decline in infant mortality over the past century has been so great that the structure of all families and the nature of adult-child relationships has changed greatly in that time. Second, the increase in migration has meant that many more young people have been able to make their own decisions on whom to marry and when, rather than wait until parents and grandparents gave them access to land, housing, and allies. (Manning, 2001) Such early centers of migration as the Caribbean and Southeast Asia may thus have been harbingers in the global transformation of families. The Databank will trace these issues by including qualitative descriptions of family structure and data on family size for each region over time.

For all these issues, the development of a sound body of consistently organized data will help. Even for complex debates, simple and consistent data will help.

Collecting and transforming data

We define our data into four levels: raw data, conversion data, transformed data, and aggregated data. For each level we distinguish data in economic, demographic, and family sections of the project.

Raw data. These are the data collected by project participants, which we distinguish as follows:

Type 1. Data selected from published (print or electronic) data collections.

Type 2. Data tabulated from published official sources.

Type 3. Data tabulated from unpublished official sources.

Type 4. Data collected from private primary sources.

Type 5. Data constructed by estimation from various sources.

Types 1 and 2 can be considered as “existing” data, while types 3, 4, and 5 can be considered as “new” data. Data for the twentieth century will mostly be “existing” data, data for the seventeenth will mostly be “new,” and data for intervening centuries will be mixtures of the two. All data are entered in terms of the original weights, measures, and currency in which they are listed, and noting the time and place to which each entry applies.

Conversion data. Weights and measures, currency exchange rates, calendar equivalencies — these and other dimensions of data are inconsistent on records of the past four centuries. As a result, one working group and one whole segment of the project is devoted to

collecting dimensional data for the world over four centuries, and for establishing equivalencies among them, building on earlier work of this type (McCusker 1978). This includes the geographic space to which each datum refers, which presents a problem of particular complexity. That is, we cannot take modern national units as unproblematic: they have changed too much to be projected back for centuries. We need more persistent categories and a wider range of categories for data collection. We will seek to identify data by place in three ways — by place as indicated in the raw data, by GIS coordinates, and by standard regions which we will construct (Melo 1999a). We will use the standard regions for regional aggregation, and will use the GIS coordinates for placing data into networks.

Transformed data. Once data have been cleaned and transformed, they will be comparable across regions and across time. It is likely that we will use several levels of transformation. That is, while it will be necessary to develop a four-century set of standard regions in order to compare volumes and values of grain trade over that time, analysts of the eighteenth-century world may prefer to rely on a set of standard regions designed especially for that era.

Aggregated data. Transformed data may be aggregated in various ways: from standard regions to nations, empires, or continents; from annual to decennial totals; from localities to transoceanic networks. Users will also have control of aggregation procedures: wage rates could be averaged over several regions, or they could be calculated as a weighted average based on the population of each region.

Displaying and analyzing data:

Each of the above levels of data will be available for inspection on the Databank website. A direct but sophisticated system of navigation will enable users to locate, download, manipulate, and comment on data at each level. For data display, a detailed and continuously maintained **site map** will provide access to information on all project activities, and to all current data sets. Data at all levels will be accessible on screen, and will also be available for downloading. The website is to be continuously reviewed for errors in format and for errors and updates in data. This will require a system for referencing points (or cells) on screen for commentary, and an e-mail link to get the comment to the webmaster.

Project staff and collaborators will work to maintain the highest level of excellence and accuracy in data entry, data display, data manipulation, and the highest feasible level of computer software and hardware. For data entry, we will establish guidelines for the format of data submitted to the project, to minimize loss or confusion of data.

Data manipulation will take place at several levels. First, project staff will store incoming data, then translate the raw data from stored form to their form on the website. Next, project staff will determine the transformations to be made for each group of data. Then project staff will place

the transformed data on the website; placing aggregated data on the website is really an extension of this same task. Meanwhile, users of the website may manipulate raw or transformed data either on-line or by downloading data and analyzing them on their own system. Further, users may create their own aggregations of data over time and space, in addition to those set up by project staff.

In software, we will seek to use the most recent available versions of web and database programs, though with an emphasis on broad compatibility. In hardware, we rely on two excellent computers — an EMC mainframe for storage and the project web server — and on skilled staff and students available at Northeastern and through our collaborators. Collaborators will be expected to have machines compatible with the project machines and software at Northeastern, though there may be exceptions.

Progress of the Databank project will bring about, for the topics addressed, a steady narrowing of the lacunae and inconsistencies in the data. This will provide us with an increasingly clear picture — region by region and over four centuries — of major patterns in commerce, migration, and family structure. The project will then turn increasingly to analysis, with emphases at three levels:

1. Verifying the quality and consistency of the data. This analysis will focus on documenting conversions — the units of measurements and dimensions of historical data—to permit valid comparison and aggregation.

2. Localized historical analysis. Case studies and comparisons provide the bedrock of social science analysis. The first level of interpretive analysis will focus on localized and comparative analysis of data in commerce, migration, and family structure.

3. Global historical analysis. Aggregations of the data will permit the description and analysis of large-scale patterns in commerce, migration, and family structure. While the aggregations will address large-scale, contiguous nations and continents, they will not be limited to such units. It will also be possible to compare and analyze such multi-regional entities as the commercial networks of the Dutch, Portuguese, and Gujarati, and the migratory networks of people from South China, South India, West Central Africa, and the Eastern Mediterranean. Project findings at all levels, but particularly at this level, will have immediate impact in either confirming or redirecting the efforts of researchers in world history.

4. Analytical techniques. Analyses will address, nominal, ordinal, and interval data, beginning with tabulations, cross-tabulations, and regressions, and may extend to more elaborate techniques

Virtual working groups: staff and collaborators

The senior staff of the project include the PI, the co-PI, two Faculty Associates at Northeastern, and a Faculty Associate with a subcontract at the Pacific World History Institute in Stockton, California. In the early stages of the project, the senior staff will have responsibility for leading in defining the data structures within their groups, especially at the project-wide conference in the first year. In subsequent stages, the group leaders are to ensure that collaborators are working within the adopted procedures, and that submission, display, and transformation of the data take place appropriately. In two subsequent years, the leader of each group will organize a conference to discuss the main issues for the working group. Group leaders will also make their own contributions of data to the Databank. Post-doctoral appointees, selected from qualified applicants, will work for periods of two to three years with the group leaders, and will work both in directing the efforts of that working group and in conducting Databank research. Further, PhD candidates in history or in other social sciences will work at Northeastern or at collaborating institutions, under the direction of each of the group leaders.

The **collaborators** will participate in the initial discussion, to design the data sets. Thereafter they will contribute data, design criteria for transforming data, discuss the results at each stage, and participate in the triennial conferences on each topic. Because of the varying nature of the data and the interpretive issues in the three research areas, we expect that the structure and practice of collaboration within each area will evolve in its own direction. The Databank project relies on collaboration of leading groups worldwide, whose work is relevant to global historical studies. Most collaborators are university-based, but the list includes private-sector firms (Digital Learning Group for web technology and Gale Group for electronic document collections) to insure that we have access to the latest technology. Our streamlined process should bring broad and interconnected results.

Senior Staff:

Patrick Manning: Principal Investigator, with responsibility for directing Economics and Conversion working groups, and for overall Project Administration.

- Director of World History Center; Professor of History, African-American Studies and Education, Northeastern University. Experience: publications in economic, demographic, and social history, website development, direction of large projects in social history research, demographic simulation, Migration CD-ROM production, and world history PhD program.

Adam McKeown: Co-P.I., with responsibility for directing Demography and Family working groups.

- Assistant Professor of History, Northeastern University. Experience: research and publication in global history of migration, program chair of World History Association international conference, 2000.

Jeffrey Burds: Faculty Associate, with responsibility for Infrastructure group.

- Associate Professor of History, Northeastern University. Experience: research and publication in migration, economic, and social history of Russia and Ukraine; editor of major documentary collections from Soviet archives; website development.

Dennis O. Flynn: Faculty Associate, director of sub-contract on Economics at the Pacific World History Institute, Stockton, California.

- Professor of Economics, University of the Pacific, and Director of the Pacific World History Institute. Experience in research and publication on silver and monetary flows in world history.

Additional staff. Two post-docs, one working on Economics and Conversions in association with Manning; one working on Demography and Family in association with McKeown. Two doctoral candidates working on Economics, and one each working on Conversions, Demography, and Family. Working with Burds on Infrastructure will be one professional webmaster and two full-time undergraduate co-op web-assistants. A sub-contract at the Pacific World History Institute, led by Flynn, will also include one history faculty member (Arturo Giraldez) and one computer science post-doc (James Sobredo). Working with Manning in project administration will be one administrative/financial assistant and one full-time undergraduate co-op graphic artist.

Collaborators. Some collaborators have confirmed their participation; others have expressed interest. We have found wide interest in this project, and expect that many major figures in world history will participate as collaborators. A full list of those invited to be collaborators is available on the Databank website, at www.databank.neu.edu. Included are the following. **Economics.** Collaborators will include Shahid Alam, Ballard Campbell and Jeffrey Burds at Northeastern; the Pacific World History Institute; Prasannan Parthasarathi at Boston College, Patrick K. O'Brien and others at the London School of Economics, Sanjay Subrahmanyam at EHESS (Paris); and groups at Moscow State University, the University of Leiden, the University of California World History Consortium, and the World Bank. **Demography.** Collaborators will include the Population Studies Center at the University of Pennsylvania (especially its African Census Project); James Lee at Caltech, the WEB Du Bois Institute at Harvard, and the Gale Group, a leading publisher of document collections. **Family.** Collaborators include the Chinese Academy of Social Sciences. Susan Watkins, and the Gale Group. **Conversions.** Collaborators include John J. McCusker. **Infrastructure.** Collaborators include James Sobredo, Pacific World History Institute; and Digital Learning International, a leading firm in instructional websites.

The World History Center

The World History Center provides a platform for global historical studies. It has grown at a university in which social science research and collaborative approaches have thrived in several

centers during the 1980s and 1990s. The Department of History and the World History Center have developed a track record in world history — a decade of systematic building and active collaboration, revealing an ability to coordinate a complex range of activities, to coordinate and collaborate within the university, and to attract increasing levels of outside support. Its accomplishments include establishing a successful doctoral program, a World History Seminar drawing many noted speakers, multimedia production, and substantial publication in global studies. The center is thus a strong base for a demonstration project on national and international collaboration in world history. More precisely, its resources for this project include

- Economy. Seminar series on Europe and Asia in the World Economy.
- Demography. (1) A web-based demographic simulation of migration — Manning. (2) a database on 20th-century Latin American population and migration — Hector Melo.
- Family. Development of theses on global family history out of Migration CD-ROM.
- Conversions: a database on administrative units of Latin America in the national period — Hector Melo.
- Doctoral research in world history. 17 current students, 4 graduates.
- Bibliography. A substantial collection of bibliographies in world history.
- Communication network, notably through H-WORLD, edited at Northeastern and University of California – Irvine.
- Organizational experience: Migration and other multimedia projects, major conferences (WHA, World History Symposium), collaborative projects, professional development.
- Links to teaching. Two contracts with College Board on AP World History.

Time frame of the work

Year 1, 2001-2002.

- Agenda: establish collaborations, define data sets (economy, demography, family), define criteria for transforming data (economy, demography, family).
- Conferences and publications: Data Definition — project-wide.
- Appointments: 1 post-doc on demography, 3 doctoral students, staff and faculty researchers

Year 2, 2002-2003.

- Agenda: define data sets (economy, family), define criteria for transforming data (economy, demography, family), collect, transform and display data (economy, demography), analyze data (demography).
- Conference and publications: Demography.
- Appointments: 3 post-docs, 6 doctoral students, staff and faculty researchers.

Year 3, 2003-2004.

- Agenda: define data sets (family), collect data (economy, demography, family), transform and display data (economy, demography, family), analyze data (demography, economy).
- Conference and publications: Economic change.
- Appointments: 3 post-docs, 9 doctoral students, staff and faculty researchers.

Year 4, 2004-2005.

- Agenda: collect data (economy, demography, family), transform and display data (economy, demography, family), analyze data (economy, demography, family).
- Conference and publications: Family

- Appointments: 3 post-docs, 12 doctoral students, staff and faculty researchers.

Year 5, 2005-2006.

- Agenda: collect data (economy, demography, family), transform and display data (economy, demography, family), analyze data (economy, demography, family). Apply for project renewal for next five years.
- Conference and publications: Summarize project results
- Appointments: 3 post-docs, 12 doctoral students, staff and faculty researchers.

Anticipated Project Results: documentation and interpretation

The first result of the project will be the launching of organized research, rather than just individual efforts, into global history — initiating the experience of collaboration, within and across institutions, among historians exploring global dimensions of our past. Development of a collaborative style among historians, and addressing the known and unknown obstacles to such a style, is a necessary concomitant to establishing institutions for advanced research in world history. Further results of the project fall into three categories, as follows:

Data sets. The databases themselves will be the most tangible results of the project. More than an index of localized databases, this will be a database on a new level, and will lay the groundwork for creating broader such global databases. Aggregations of transformed data will probably be the most useful form of data for historians, in that they are most parallel to the data under debate in the literature, which focus on national comparisons. But analysis of transformed data by standard (and usually sub-national) regions may be especially useful for correlation analysis and for study of networks.

In this large and wide-ranging project, not all data will be of equal precision. Yet to have comprehensive data, known to be valuable within a certain tolerance, will bring a substantially positive impact on interpretation. For instance, Philip Curtin's estimates of the volume of the Atlantic slave trade, assessed as valid within a tolerance of 20%, were sufficient to bring a major reinterpretation of Atlantic history. Those same figures, used in association with a demographic simulation, have been sufficient to demonstrate with virtual certainty that the population of West and Central Africa declined during the century of largest-scale export slave trade (Curtin 1969, Manning 1990a). While the project may be hindered by the difficulty of locating or transforming certain data, staying in touch with active scholars will help us to locate other and unexpected data.

Research findings and debates. Debates will take place during the development of the datasets, not just at the conclusion of the project. Continuing discussion on the web — with conferences and proceedings to provide periodic summaries — will highlight emerging debates on definition of the data, interpretation of the data, formulation of historical proxies for theoretical variables, the causes and impacts of long-term trends, and the nature and causes of interactions in economy and society. As with any active and productive field of study, world history needs its debates, and needs to provide a dependable forum at which differing opinions may be expressed.

Infrastructure for World History. By strengthening an administrative node and a broad collaborative net, the project will become a model for social science research in the information age. Rather than construct a vast academic center at a single university, in the manner of Area Studies Centers of the postwar era, the project will create an academic infrastructure of far greater reach without having to displace its participants. Of this general infrastructure, the World History Center is one substantial element. We will create a World History Gateway, linking all major sites of world-historical interest, and annotating them according to standards set by an advisory board. We will also develop a global historical bibliography, annotated for the purposes of world historians, as part of the general infrastructure. The center administration, through its coordination of a range of interactive elements in global historical studies, will gain experience in the conduct of research at this level of breadth. The doctoral program in world history, which will grow in strength through its association with this research project, will provide a stronger base for global research projects on other topics.

This network of collaborations among global and area-studies scholars will benefit national historians as well as world historians. We emphasize that, in both research and teaching, the study of world history will be as a complement to and not a substitute for historical studies at national and area-studies levels. The datasets themselves, since they will be presented in disaggregated as well as aggregate form, will be of use to both world and national historians. The creation of annotated bibliographies and registers of archives draws on the accumulation of work by national and area-studies historians, yet reformulates it for the use of world historians. The transformation and analysis of data, necessary to permit aggregation at a global scale, will also facilitate comparison between pairs of national units.

Comprehensive and interactive study. Recent analyses of global economic, demographic, and social history in the modern period have focused on large states and powerful economies (Pomeranz, O'Rourke and Williamson, Lee, Watkins). While the emphasis on these important cases is surely appropriate, such studies tend to entail an implicit assumption that they may be taken as a proxy for the world as a whole. The purpose of this project is to test this assumption by developing observations on the entirety of the global system, to see how its dynamics have differed from those of its most prominent subsystems. The basic data provided by this project, because of their comprehensiveness, will be of importance even in complex issues of family dynamics and economic growth.

Evaluating and extending the work of this project

The first five years of work under this project will transform substantially the basis of global historical studies addressing the modern period. The substantial investment of NSF in this coordinated program of research will clarify the possibilities of global historical research. The data

assembled and analyzed will lead to the resolution of some debates and the specification of others. The concentration of discussion and the preparation of a substantial cohort of young world historians will expand the discourse in world history, and link it more effectively to the rest of historical studies and the rest of social sciences. The model established by this project, including the research undertaken by its doctoral students, will strengthen the Northeastern Ph.D. program itself, and will encourage the development of other groups of collaborators in world history. It seems logical to conclude that world history will grow to be a substantial field of study, and many research groups will have to develop.

In addition to annual internal evaluation of the project, users of the website will be prompted to provide their evaluation of the Databank.

Conclusion: creating a basis for long-term global analysis

This project will consolidate basic historical data on commerce, migration, and family into an accessible database that will support the expansion of solidly grounded analysis in world history. It will move beyond the Sole Investigator and Isolated Team approaches, will develop among historians the collaborative approach known in other fields.

For migration, the Databank will assemble numbers of persons estimated to have moved from place A to nearby place B or distant place C, for each place and time period. These data will come from shipping data, census and other enumerations, and tax records. Aggregating these results will yield, for instance, links and parallels of African migrations 19th and 20th centuries to those of eastern and southern Asia. It will also document patterns of domestic and international urbanization in 19th and 20th centuries. The Demography portion of the Databank will also yield an updated and systematic estimate of world and regional population since 1600.

For families, the Databank will assemble descriptions of family structure and estimates of family size. The descriptions will have to account for differences in the sources: family structures are described differently in census returns, in descriptions by travelers, in tax records and in religious records. Family size, similarly, is recorded in various ways, for instance as the number of co-resident adults, or the number of children per mother, or the number of people per dwelling. Despite the variety in the sources, a thorough job of data collection and a careful typological review will enable the project staff to develop and display a useful range of estimates of family size and structure, as these have changed over time.

For commerce, the Databank will mostly display quantities and prices of goods flowing from place A to place B. These will include such commodities as precious metals (silver and gold), textiles (cotton, silk, wool, linen, other), grains (wheat, rice, barley, millet, sorghum, maize), metals (iron, copper), and wages (for construction, transport, manufacturing, or agricultural workers). The

data will be retrieved from shipping data, customs and other tax data, merchant records, and inventories.

For the overall Databank, if we created an overview in which we set data into 200 standard regions, and recorded observations every ten years from 1600 to 2000, we would have 8,000 cells into which to place data on commerce, migrations and family. The 200 standard regions would have an average population of 2.5 million for 1600, and an average population of 30 million for 2000.

If we took Angola in 1800 as one cell, the data to be entered would include the following. For migration: in-migrants, out-migrants, population (preferably with age and sex breakdown). For family: estimated family size, description of family structure. For economy: quantities and values of commerce in silver, gold, cotton, wool, silk, wheat, rice, maize, and iron, plus wage rates. For conversions: currency exchange rates, weights and measures, current demographic categories, and regional boundaries. The Databank website will show all of these as raw data (as given in each source document, along with its citation). More usefully, the data will be transformed to fit into standard regions and time periods, and will include other relevant corrections to show the number of migrants, regional population, family size and structure, and quantities and values of commerce in several goods. The values of commerce may be deflated by any of several price indices, depending on the nature of the comparison intended.

For the purpose of more large-scale analysis, the Angolan data might be aggregated with other regions of southern Africa for an estimate of regional trade and migration. Or Angola might be aggregated with regions of Brazil, Portugal, and the Indian Ocean coast for study of the Portuguese empire.

While the total size of the Databank will be smaller than many current survey data sets, it will have numerous dimensions because of the range of factors under consideration. For this reason it will take a powerful server, and the site map will have to be skillfully designed

Expect to begin with 10 collaborators, have that rise to 20 in the first year, with the number rising to 60 as the number of international collaborators increases.

In the first year or so, the website will be of interest primarily to project staff, and will receive perhaps a hundred hits a day. At a next stage it will be of interest to scholars generally, and may receive two hundred hits a day. At a further stage it will be ready for use by teachers and students, and usage will rise to perhaps 500 hits a day.

Approximate division of effort among the segments of the project will be: Economics 35%, Demography 15%, Family 10%, Conversions 10%, Infrastructure 15%, Administration 10%, Contingency 5%.