A vast quantity of raw census microdata for the period since the 1960s survives in machine-readable form. Most of these data, however, remain inaccessible to researchers. This proposal seeks funding to create harmonized and documented samples of 16 countries, in 5 major world regions, totalling more than xx percent of the world’s population. These data will be made available for scholarly and educational research through the web-based data dissemination system, www.ipums.org/international as governed by project protocols endorsed by xx official statistical agencies (Appendix A).

This project leverages previous federal investments in social science infrastructure. Grants from the National Institutes of Health and the National Science Foundation have laid the groundwork for this new Global data series by funding many of the initial costs. Those projects have underwritten the development of data cleaning and sampling procedures, data conversion and dissemination software, and design protocols for data and documentation. We are making arrangements to obtain raw microdata files, internal documentation, and redistribution agreements for over 50 censuses from 16 countries with populations totaling more than 1.xx billion people. As a result, the new global microdata samples will be highly cost-effective.

The following tasks must be carried out to capitalize on these past investments and make these new census samples widely available to researchers: draw new samples of each census; reformat and clean the samples; impose confidentiality protections; recode variables into existing harmonized coding systems and develop new coding designs; allocate missing and inconsistent data values; create a set of consistent constructed variables; develop harmonized English-language documentation; convert all documentation to the Data Documentation Initiative metadata standard; and improve and maintain the web-based data access system.

With over 100 million records spanning as many as four decades, the new database will allow scientists to make comparisons around the globe during decades of dramatic change. Coupled with data from other IPUMS projects, this information will allow innovative comparative research across time and space. The data series will result in a substantial body of new scientific and policy-relevant research on economic transformation, demographic transition and population aging, international migration, and many other topics.

PERFORMANCE SITE(S) (organization, city, state)

Minnesota Population Center
537 Heller Hall
271 19th Avenue South
Minneapolis, MN 55455

Minnesota Population Center
425 20th Avenue South
Minneapolis, MN 55454

KEY PERSONNEL. See instructions. Use continuation pages as needed to provide the required information in the format shown below. Start with Principal Investigator. List all other key personnel in alphabetical order, last name first.

<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Role on Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert McCaa</td>
<td>University of Minnesota</td>
<td>Principal Investigator</td>
</tr>
<tr>
<td>Steven Ruggles</td>
<td>University of Minnesota</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>Miriam King</td>
<td>University of Minnesota</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>Deborah Levison</td>
<td>University of Minnesota</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>Matthew Sobek</td>
<td>University of Minnesota</td>
<td>Co-Investigator</td>
</tr>
</tbody>
</table>
Overview
The goal of this project is not simply to make census microdata available; it will also make them usable. Even where census microdata can be obtained, comparison across countries or time periods is challenging because of inconsistencies between datasets and inadequate documentation of comparability problems. Because of this, comparative research based on pooled census samples is rarely attempted. This project will reduce the barriers to cross-national research by converting census microdata into a uniform format, providing comprehensive documentation, and making the data available without cost to researchers through a web-based access system. The new global census microdata series will be fully harmonized with IPUMS-International, thus facilitating comparison across continents as well as within them.

We anticipate that this new microdata series will include as many as one hundred million person records from sixteen countries encompassing more than 50 censuses, and there is potential to include additional censuses from other countries. Table 1 reports the countries that have been invited to participate, the censuses that are likely to be included, proposed sample densities and a rough estimate of the number of person records. (Please note: As of March 8, this table will be updated to include only those countries where a letter is forthcoming authorizing the project to proceed. Please examine the information for your country and email corrections as your earliest convenience to Robert McCaa: rmccaa@umn.edu)

For purposes of planning and design, we must work simultaneously with the censuses of all participating countries. This will ensure that we accommodate the full range of variation across countries and census years when designing harmonized variable coding systems. During data and documentation processing, however, we will work with batches of four countries at a time. This approach—also used for the first IPUMS-International initiative as well as for IPUMS Latin-America and IPUMS-Europe—allows timely release of samples and avoids the logistical complexity of processing too many censuses simultaneously.

We will process as many batches as possible within the five years of this project. Based on our experience with China, France, Kenya, Vietnam, Mexico, Colombia, and Brazil, we hope to complete work on four batches—sixteen countries and fifty datasets—within the time frame of the present project. Depending on the extent of data format and consistency problems we encounter, however, that number could change. We will establish a priority sequence based on intellectual salience, census quality, technical characteristics, and the release schedule for the 2000 round of census data for each country.

Foreign Expenses
We are negotiating licenses and fees for the dissemination of microdata for each country, at a rate of US$5,000 per census microdataset provided. In year one of the grant, one-half of the total fee for each country will be paid to license the complete set of microdata and documentation for all census rounds and microcensuses before 2000. The remainder will be paid upon receipt of microdata and metadata for the 2000-round census. In addition, there will be a budget for each country, equal to the amount paid to the official census authority, which will be used to compensate expert consultants. This is very cost-effective, since the fees cover the cost of supplying the data, and where necessary, translating essential source documentation, and technical support by national experts (e.g., drawing samples from the complete data as needed, supplying necessary documentation, and answering questions). There are several earlier censuses for which the readability of old census tapes cannot be confirmed until the license agreements are executed, including India 1971/81, Italy 1981, Philippines 1960, Sudan 1973, and Zambia 1969. Funds are also requested for the recovery of these data and others that may be discovered during the course of work. If any of the datasets do not materialize, we will use the savings to negotiate licenses for censuses from other countries. We also have budgeted a substantial sum in the final project year to cover the costs of developing additional censuses that become available during the course of the project.
Table 1. Countries invited to participate, proposed sample density (%) and sample sizes (millions)
(bold census year = existence of microdata confirmed)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Algeria</td>
<td>31.4</td>
<td>10</td>
<td>1998</td>
<td>2.4</td>
<td>1987</td>
<td>2.0</td>
<td>1977</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>128.1</td>
<td>5</td>
<td>2001</td>
<td>6.5</td>
<td>1991</td>
<td>5.0</td>
<td>1981</td>
</tr>
<tr>
<td>* Canada</td>
<td>30.7</td>
<td>5</td>
<td>2001</td>
<td>1.5</td>
<td>1991</td>
<td>1.3</td>
<td>1981</td>
</tr>
<tr>
<td>* Egypt</td>
<td>68.3</td>
<td>5</td>
<td>2006</td>
<td>3.4</td>
<td>1996</td>
<td>2.6</td>
<td>1986</td>
</tr>
<tr>
<td>* Germany</td>
<td>82.2</td>
<td>1</td>
<td>2000m</td>
<td>1.2</td>
<td>1997m</td>
<td>1.8</td>
<td>1987</td>
</tr>
<tr>
<td>Ghana</td>
<td>19.5</td>
<td>2000</td>
<td>?</td>
<td>1984</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Hong Kong S.A.R.</td>
<td>7.0</td>
<td>2001</td>
<td>0.6</td>
<td>1981</td>
<td>0.5</td>
<td>1971</td>
<td>0.4</td>
</tr>
<tr>
<td>India</td>
<td>1,002.1</td>
<td>5</td>
<td>2001</td>
<td>25.0</td>
<td>1991</td>
<td>38.5</td>
<td>1981</td>
</tr>
<tr>
<td>Indonesia</td>
<td>212.2</td>
<td>5</td>
<td>2000</td>
<td>10.6</td>
<td>1990</td>
<td>8.2</td>
<td>1980</td>
</tr>
<tr>
<td>Iran</td>
<td>67.4</td>
<td>5</td>
<td>.</td>
<td>1996</td>
<td>2.6</td>
<td>1986</td>
<td>2.1</td>
</tr>
<tr>
<td>Iraq</td>
<td>23.1</td>
<td>5</td>
<td>.</td>
<td>1997</td>
<td>0.9</td>
<td>1987</td>
<td>0.7</td>
</tr>
<tr>
<td>* Israel</td>
<td>6.2</td>
<td>5</td>
<td>.</td>
<td>1995</td>
<td>0.2</td>
<td>1983</td>
<td>0.2</td>
</tr>
<tr>
<td>+ Italy</td>
<td>57.8</td>
<td>2</td>
<td>2001</td>
<td>1.2</td>
<td>1991</td>
<td>0.6</td>
<td>1981</td>
</tr>
<tr>
<td>Jordan</td>
<td>5.0</td>
<td>2004</td>
<td>0.3</td>
<td>1994</td>
<td>0.2</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>Korea, Republic of</td>
<td>47.2</td>
<td>5</td>
<td>2001</td>
<td>2.4</td>
<td>1990</td>
<td>1.8</td>
<td>1980</td>
</tr>
<tr>
<td>* Malawi</td>
<td>10.3</td>
<td>10</td>
<td>.</td>
<td>1998</td>
<td>0.8</td>
<td>1987</td>
<td>0.6</td>
</tr>
<tr>
<td>* Malaysia</td>
<td>23.2</td>
<td>2</td>
<td>2001</td>
<td>0.4</td>
<td>1991</td>
<td>0.3</td>
<td>1980</td>
</tr>
<tr>
<td>Morocco</td>
<td>28.7</td>
<td>10</td>
<td>2002</td>
<td>2.9</td>
<td>1994</td>
<td>2.2</td>
<td>1982</td>
</tr>
<tr>
<td>Nigeria</td>
<td>123.3</td>
<td>5</td>
<td>2002</td>
<td>6.2</td>
<td>1991</td>
<td>4.7</td>
<td>.</td>
</tr>
<tr>
<td>Pakistan</td>
<td>150.6</td>
<td>10</td>
<td>.</td>
<td>1998</td>
<td>11.6</td>
<td>1981</td>
<td>9.4</td>
</tr>
<tr>
<td>* Palestine Territory</td>
<td>3.1</td>
<td>20</td>
<td>.</td>
<td>1997</td>
<td>0.5</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>* Philippines</td>
<td>80.2</td>
<td>5-10</td>
<td>2000</td>
<td>?</td>
<td>1990</td>
<td>6.0</td>
<td>1980</td>
</tr>
<tr>
<td>+ Russia</td>
<td>144.4</td>
<td>5</td>
<td>2002</td>
<td>7.2</td>
<td>1994</td>
<td>7.2</td>
<td>1989</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>21.6</td>
<td>5</td>
<td>2002</td>
<td>1.1</td>
<td>1992</td>
<td>0.8</td>
<td>.</td>
</tr>
<tr>
<td>+ South Africa</td>
<td>43.4</td>
<td>10</td>
<td>2001</td>
<td>4.4</td>
<td>1996</td>
<td>3.4</td>
<td>.</td>
</tr>
<tr>
<td>Sudan</td>
<td>29.4</td>
<td>10</td>
<td>2003</td>
<td>2.9</td>
<td>1993</td>
<td>2.3</td>
<td>1983</td>
</tr>
<tr>
<td>* Thailand</td>
<td>62.0</td>
<td>1-2</td>
<td>2000</td>
<td>?</td>
<td>1990</td>
<td>0.5</td>
<td>1980</td>
</tr>
<tr>
<td>Turkey</td>
<td>65.3</td>
<td>5</td>
<td>2001</td>
<td>3.3</td>
<td>1990</td>
<td>2.5</td>
<td>1980</td>
</tr>
<tr>
<td>* Uganda</td>
<td>23.3</td>
<td>10</td>
<td>2002</td>
<td>2.3</td>
<td>1991</td>
<td>1.8</td>
<td>1980</td>
</tr>
<tr>
<td>Zambia</td>
<td>9.5</td>
<td>10</td>
<td>2000</td>
<td>1.0</td>
<td>1990</td>
<td>0.7</td>
<td>1980</td>
</tr>
</tbody>
</table>

Note: Countries for which participation is confirmed as of March 8, 2004 are noted in bold.
+ = technical recommendation to participate; signing requires approval by ministry or other authority
Census year is NOT noted, where a census was NOT taken or microdata representative of the
entire population no longer survive.
Appendix A Memorandum of Agreement.

Integrated Public Use Microdata Series International
and [National Statistical Agency of Country X]

Purpose: The purpose of this letter is to specify the terms and conditions under which metadata and microdata produced by the [National Statistical Agency of X] shall be distributed by Integrated Public Use Microdata Series International of the University of Minnesota.

1. Ownership. The [National Statistical Agency of X] is the owner and licensee of the intellectual property rights (including copyright) in the metadata and microdata of [X] acquired by the University of Minnesota to be distributed by Integrated Public Use Microdata Series International. This agreement explicitly authorizes release to the University of census microdata of [X] that may be in the possession of third parties. The University is obligated to provide to the [National Statistical Agency of X] timely notice of any such acquisitions and, upon request and without cost, provide copies of same.

2. Use. These data are for the exclusive purposes of teaching, scientific research and publishing, and may not be used for any other purposes without the explicit written approval, in advance, of the [National Statistical Agency of X]. A copy of both the original census microdata and integrated samples will be deposited with the World Health Organization, Geneva Switzerland for the exclusive research needs of that institution.

3. Authorization. To access or obtain copies of integrated microdata of [X] from Integrated Public Use Microdata Series International, a prospective user must first submit an electronic authorization form identifying the user (i.e., principal investigator) by name, electronic address, and institution. The principal investigator must state the purpose of the proposed project and agree to abide by the regulations contained herein. Once a project is approved, a password will be issued and data may be acquired from servers or other electronic dissemination media maintained by Integrated Public Use Microdata Series International, the [National Statistical Agency of X], or other authorized distributors. Once approved, the user is licensed to acquire integrated metadata and microdata of [X] from Integrated Public Use Microdata Series International or other authorized distributors. No titles or other rights are conveyed to the user.

4. Restriction. Users are prohibited from using data acquired from the Integrated Public Use Microdata Series International or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise.

5. Confidentiality. Users will maintain the absolute confidentiality of persons and households. Any attempt to ascertain the identity of a person, family, household, dwelling, organization, business or other entity from the microdata is strictly prohibited. Alleging that a person or any other entity has been identified in these data is also prohibited.

6. Security. Users will implement security measures to prevent unauthorized access to microdata acquired from Integrated Public Use Microdata Series International or its partners.

7. Publication. The publishing of data and analysis resulting from research using metadata or microdata of [X] is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite [National Statistical Agency of X] and Integrated Public Use Microdata Series International as the sources of the data of [X], and to indicate that the results and views expressed are those of the author/user.

8. Violations. Violation of the user license may lead to professional censure, loss of employment, and/or civil prosecution. The University of Minnesota, national and international scientific organizations, and the [National Statistical Agency of X] will assist in the enforcement of provisions of this accord.

9. Sharing. Integrated Public Use Microdata Series International will provide electronic copies to the [National Statistical Agency of X] of documentation and data related to its integrated microdata as well as timely reports of authorized users.

10. Jurisdiction. Disagreements which may arise shall be settled by means of conciliation, transaction and friendly composition. Should a settlement by these means prove impossible, a Tribunal of Settlement shall be convened which will rule upon the matter under law. This Tribunal shall be composed of an (1) arbitrator, which shall be elected by lot from the list of Arbitrators of the Chamber of Commerce of Paris. This agreement shall be governed by, and construed in accordance with, generally accepted principles of International Law.

Date: ________________________________________
Signed: ________________________________________

Regents of the University of Minnesota
By: Kevin J. Mc Koskey, Sponsored Projects Administration
Date: ________________________________________
Signed: ________________________________________

Rev. Aug. 1, 2003
Table 2. IPUMS-International Partners  
(Limited to agencies which have officially endorsed the letter of understanding,  
“Appendix A”, as of January 1, 2004)

<table>
<thead>
<tr>
<th>Agency Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Institute of Statistics and Censuses of Argentina</td>
</tr>
<tr>
<td>Statistics Austria</td>
</tr>
<tr>
<td>Ministry of Statistics and Analysis of Belarus</td>
</tr>
<tr>
<td>Brazilian Institute of Geography and Statistics</td>
</tr>
<tr>
<td>National Statistics Institute of Bulgaria</td>
</tr>
<tr>
<td>National Institute of Statistics of Chile</td>
</tr>
<tr>
<td>National Bureau of Statistics of China</td>
</tr>
<tr>
<td>National Administrative Department of Statistics of Colombia</td>
</tr>
<tr>
<td>National Institute of Statistics and Censuses of Costa Rica</td>
</tr>
<tr>
<td>Central Statistical Office of the Czech Republic</td>
</tr>
<tr>
<td>National Statistics Office of the Dominican Republic</td>
</tr>
<tr>
<td>National Institute of Statistics and Censuses of Ecuador</td>
</tr>
<tr>
<td>General Directorate of Statistics and Censuses of El Salvador</td>
</tr>
<tr>
<td>National Institute of Statistics and Economics Studies, France</td>
</tr>
<tr>
<td>Ghana Statistical Service</td>
</tr>
<tr>
<td>National Statistical Office of Greece</td>
</tr>
<tr>
<td>National Institute of Statistics of Guatemala</td>
</tr>
<tr>
<td>National Institute of Statistics of Honduras</td>
</tr>
<tr>
<td>Hungarian Central Statistical Office</td>
</tr>
<tr>
<td>Central Bureau of Statistics of Israel</td>
</tr>
<tr>
<td>Central Bureau of Statistics, Kenya</td>
</tr>
<tr>
<td>Latin American and Caribbean Institute for Economic and Social Planning (CELADE)</td>
</tr>
<tr>
<td>National Institute of Statistics of Madagascar</td>
</tr>
<tr>
<td>National Statistical Office of Malawi</td>
</tr>
<tr>
<td>National Institute of Statistics, Geography, and Informatics, Mexico</td>
</tr>
<tr>
<td>Netherlands Institute for Scientific Information Services</td>
</tr>
<tr>
<td>National Institute of Statistics and Censuses of Nicaragua</td>
</tr>
<tr>
<td>Palestinian Central Bureau of Statistics</td>
</tr>
<tr>
<td>Census and Statistics Directorate of Panama</td>
</tr>
<tr>
<td>General Directorate of Statistics, Surveys, and Censuses of Paraguay</td>
</tr>
<tr>
<td>National Institute of Statistics and Informatics of Peru</td>
</tr>
<tr>
<td>*Central Statistical Office of Poland</td>
</tr>
<tr>
<td>National Institute of Statistics of Portugal</td>
</tr>
<tr>
<td>National Institute of Statistics of Romania</td>
</tr>
<tr>
<td>*Russian State Committee for Statistics</td>
</tr>
<tr>
<td>Statistical Office of Slovenia</td>
</tr>
<tr>
<td>National Institute of Statistics of Spain</td>
</tr>
<tr>
<td>State Committee of Statistics of Tajikistan</td>
</tr>
<tr>
<td>National Institute of Statistics and Information of Turkmenistan</td>
</tr>
<tr>
<td>Uganda Bureau of Statistics</td>
</tr>
<tr>
<td>United Kingdom Office for National Statistics</td>
</tr>
<tr>
<td>National Institute of Statistics of Venezuela</td>
</tr>
<tr>
<td>General Statistics Office of Vietnam</td>
</tr>
<tr>
<td>*pending ministerial approval</td>
</tr>
</tbody>
</table>