

---

# Georeferenced Population Data

---

by Hendrik Meij and Robert Chen<sup>1</sup>  
Consortium for Earth Science Information Network

## INTRODUCTION.

The primary mission of The Socioeconomic Data and Applications Center (SEDAC), at the Consortium for International Earth Science Information Network (CIESIN), is to develop new policy oriented applications and information products that synthesize earth science and socioeconomic data. The SEDAC policy applications development effort is the primary means by which the Earth Observing System Data and Information System (EOSDIS) program helps to ensure that the scientific investment embodied in NASA's Mission to Planet Earth (MTPE) program leads to tangible benefits. SEDAC's activities closely link with ongoing activities related to land use and trace gas emissions at other EOSDIS Distributed Active Archive Centers (DAACs) such as EROS Data Center (EDC) and Oak Ridge National Laboratory (ORNL). In addition, SEDAC's efforts also play a critical role in the arena of integrated assessment of global climate change. Its aim is making outputs of the U.S. Global Change Research Program (USGCRP) useful for policy and develop specific tools and mechanisms to enhance the use of integrated assessment models in the policy process.

Population dynamics and distribution have been consistently identified as key elements in understanding human interactions with the environment and in considering possible responses to environmental change. The National Research Council (NRC) has identified (1) population dynamics as one of five priority areas of research for the USGCRP. It also points out the key role of georeferenced social data in two other priority areas:

- 1) improving understanding of land use change, and
- 2) assessing impacts, vulnerability, and adaptation to global changes.

In addition, the report highlights the need to address the full range of energy policy options in relationship to greenhouse gas reduction. The 1992 report of the Human Dimensions of Global Environmental Change Programme (HDP) on Population Data and Global Environmental Change (2) emphasized the importance of georeferenced population data for global change research and applications and recommended development of data bases at several different levels of aggregation and resolution.

## GEOREFERENCED POPULATION DATA.

Georeferenced population data provide a critical link between data on the natural environment and data on human behavior and welfare. Past and potential policy oriented uses of such data include natural resource management, famine early warning and vulnerability assessment, design and planning of sample surveys, damage assessment and associated disaster response, assessment of the impacts of environmental variation and change (e.g., coastal storms and sea level rise), public health and medical service applications, urban and regional planning, and estimation of pollution emissions and land use change.

Spatial and temporal data on population and environment can be useful in analyzing behavioral impacts. For example, energy analysts interested in emission reduction policies may want to consider population location and behavioral patterns in relationship to alternative energy sources, air and water resources, pollution control, work and recreational areas, and transportation infrastructure. Urban planners may want to understand the effects of decentralized versus centralized population growth on energy use and emissions and assess the effects of alternative population and land use policies on population distribution. National policy makers are likely to be interested in the causes and impacts of internal and international migration, especially as they relate to environmental degradation and change in specific regions. Educators may want to introduce such data sources early in the curriculum to provide a base for interaction with national and local processes.

## DATA PRODUCTS.

SEDAC has been developing a number of different georeferenced population datasets:

- o A set of gridded population datasets for more than 120 countries originally developed by the Center for International Research (CIR) of the U.S. Bureau of the Census for the Department of Defense (DoD) and recently made available

through SEDAC. The data files contain both urban and rural population density data at a resolution of either 20 by 30 minutes or 5 by 7.5 minutes. These files may be accessed via anonymous File Transfer Protocol (FTP).

```
ftp ftp.ciesin.org user <ftp> password <email address> cd /pub/data/Global_Population_DB
```

o A suite of high resolution, one tenth of one degree, population data sets for the globe. Two gridded data sets are provided, one with smoothed population counts data, and one with smoothed population density data, both at the 1/10th of a degree resolution. These data are smoothed by applying a mathematical technique (pynophylactic interpolation) for preserving areal data while redistributing such data on a sphere (3). These files may be accessed via anonymous FTP.

```
ftp ftp.ciesin.org user <ftp> password <email address> cd /pub/data/Global_Demog_Project
```

o A georeferenced population dataset for the conterminous U.S. at the square kilometer resolution level. The approach taken in the development of this product requires both input from the U.S. Bureau of the Census 1992 TIGER/Line and Summary Tape File 3A databases. The initial attempt will focus on the transformation of census blockgroup total persons counts and total housing units structures to the pixel based format of a square kilometer. Several disaggregation methods are applied including majority rule, pynophylactic interpolation, proportional allocation, and geostatistical estimations (kriging). Prototype efforts may be found at:

```
ftp ftp.ciesin.org user <ftp> password <email address> cd /pub/census/usa/grid
```

Please email [ciesin.info@ciesin.org](mailto:ciesin.info@ciesin.org) for up to date information.

o SEDAC has identified available sub national administrative boundary files for most countries of the world and is developing an integrated product using a public domain version of Digital Chart of the World (DCW) at a scale of 1:1,000,000. Such boundary files are critical in making linkages between socioeconomic data (e.g., on population, land use, and energy production and consumption) and natural science data (e.g., on land cover, vegetation change, and pollutant levels).

## **ACCESS PRODUCTS.**

A series of activities are underway to make data from the U.S. Bureau of the Census, and possibly other national census takers, more widely accessible and usable. This includes development of:

o An interactive tabulation generator publicly accessible over the Internet. This program enables rapid access to very large census datasets for user defined cross tabulations generated from microdata samples. Currently, the U.S. 1% Public Use Microdata Samples (PUMS) files for 1980 and 1990 are available. Results may be save and retrieved via Kermit, FTP or email.

Please email [ciesin.info@ciesin.org](mailto:ciesin.info@ciesin.org) for up to date information.

o An interactive extraction generator publicly accessible over the Internet. This program enables the user to define, interactively, custom extractions to be performed on very large census datasets such as the U.S. 5% Public Use Microdata Samples (PUMS) files for 1980 and 1990. A custom extraction file is generated with a custom codebook and data dictionary. Retrieval mechanism include only FTP.

Please email [ciesin.info@ciesin.org](mailto:ciesin.info@ciesin.org) for up to date information.

o Creation of an “ Archive of Census Related Products” involving the generation of usable products from pre tabulated datasets, such as the U.S. Summary Tape File 3A. Boundary files (from Tiger/Line 1992). Demographic data and boundary files are accessible via anonymous FTP. The demographic data records are uniquely linked to the appropriate area entities in the boundary files. The current Archive contains 11,300 retrievable files spanning 4,000 MB.

ftp ftp.ciesin.org user <ftp> password <email address> cd /pub/census <indent> Informational ‘readme’ files are echoed to the screen to guide the user in navigating this Archive. WWW browsers, such as Mosaic and Netscape clients, can enter the Archive via:

```
http://www.ciesin.org:/datasets/us demog/us demog home.html
```

Look for the “/pub/census” hypertext link.

o The ability for visual display, mapping, of census data by coverages, for browsing/ inspection purposes before retrievals. <indent2> <http://www.ciesin.org:2222/map.html>

o The generation of “Dataset Guides” for informational purposes. <indent2> [http://www.ciesin.org/datasets/us demog/us demog home.html](http://www.ciesin.org/datasets/us%20demog/us%20demog%20home.html)

#### **SUMMARY.**

The development of these products, providing rapid access to large demographic databases, combined with visual presentation of the materials, might enable more focused response measures to catastrophic events and others. It is hoped that in the short term, users will express and identify databases of interest for merging into the prototype systems under development.

#### **LITERATURE CITED.**

1. Paper presented at IASSIST95 May 1995 Quebec City, Quebec, Canada. Consortium for Earth Science Information Network 2250 Pierce Road University Center, MI 48710 (hmeij@ciesin.org, bob.chen@ciesin.org)

2. National research Council (NRC). 1994. Sciences Priorities for the Human Dimensions of Global Change. National Academy of Sciences Reports.

3. Human Dimensions of Global Environmental Climate Change Programme (HDP). 1992. Population Data and Global Environmental Change.

4. Tobler, W. 1979. Smooth pycnophylactic interpolation for geographical regions. Journal American Statistical Association, 74(367): 519-536.