Terra Populus: Integrated Data on Population and the Environment

Matthew Sobek, Steven Ruggles, Tracy Kugler, and David Van Riper
University of Minnesota

Terra Populus (TerraPop) is a project designed to lower barriers to conducting human-environment interactions research. It will do this by making data with different formats from diverse scientific domains accessible and easy to use in combination. TerraPop will develop infrastructure to integrate, disseminate, and preserve data describing population and environment on a global scale over the past two centuries, including data on human population characteristics, land use, land cover, and climate. It will make these data interoperable across time and space, disseminate them to researchers and the public, and preserve them for future generations. The TerraPop data will be available to all users free of cost.

TerraPop is housed at the Minnesota Population Center (MPC) at the University of Minnesota. The project is a collaboration of the MPC with two other organizations at Minnesota and two external organizations: the University of Minnesota Institute on the Environment (IonE), the University of Minnesota Libraries, the Center for International Earth Science Information Network (CIESIN) at Columbia University, and the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. TerraPop is funded by the National Science Foundation Office of Cyberinfrastructure as part of the DataNet initiative. The first phase of the project is funded from 2011 to 2015. A public beta release of the data system is scheduled for late 2013, and the official release in early 2014.

Population Growth and Environmental Change

Many of today's key scientific challenges involve understanding the interconnections between population growth and environmental change in the face of rapid transformations in both. Over the past five decades, the world's population more than doubled, and the spatial distribution of population is also changing dramatically. The rate of population growth is slowing as fertility rates decline, creating a shift in the age composition of the world's population (National Research Council, 2001; United Nations, 2006). Other dramatic demographic trends—rising urbanization and international migration, industrialization of the developing world, and improvements in education and health—are likely to continue or accelerate in coming decades. Economically speaking, world per-capita gross domestic product roughly doubled, but that expansion was uneven, marked by growing inequality in many regions and little convergence in economic development between rich and poor countries (Ferreira and Ravallion, 2009; Lutz and Goujon, 2001).

The extraordinary levels of global demographic and economic growth since the 1950s have had ominous environmental consequences: alarming environmental degradation, resource depletion, and climate change (O'Neill, MacKellar, and Lutz, 2001; World Bank, 2009). In just the last 50 years, food and water consumption roughly tripled, alongside a four-fold increase in use of fossil fuels. These increases in consumption place extraordinary new pressures on global land resources, biodiversity, and "ecosystem goods and services" (Fischlin et al., 2007; Foley et al., 2005). The average global temperature has gone up 0.74˚ Celsius over the past century, and is now rising at an accelerating pace, and new precipitation patterns are becoming more pronounced. Deforestation and pollution are compounding the direct effects of global warming and contributing to the destruction of ecosystems and decline of biodiversity (IPCC, 2007; National Science Foundation, 2009).

The connections between population and environment work in both directions. The key drivers of environmental change—especially fossil fuel emissions and deforestation—are direct consequences of population growth and economic development. Conversely, environmental change has profound implications for demographic behavior. Flooding, erosion of coastal areas, destruction of ecosystems, and drought and degradation of water supplies all have consequences for human populations, such as mass migration, food scarcity, and increased armed conflict.

The Need for TerraPop

Our understanding of the interactions of population and environment has been hampered by the dearth of internationally comparable data. The variety of data needed to study human-environment interactions is scattered across scientific disciplines and is consequently subject to different conventions, terminologies, access mechanisms, and data formats. Simple discovery of the available data sources is difficult for non-specialists of the relevant fields. It is substantially more daunting for a researcher to manipulate those data into a form amenable to the analytical methods of their field. In fact, few researchers have the necessary background to handle the variety of data formats, and the cost of doing so is high even for those with the requisite skills.

Key scientific organizations, including the U.S. National Research Council and National Science Foundation, have emphasized the need for readily discoverable, accessible, and integrated data spanning large spatial, temporal, and topical extents. Scientific advances and policy insights are greatest when users with varying theoretical perspectives and models have access to the same data. The National Research Council (2001) makes a compelling case for the development and use of cross-national and cross-temporal data sources, arguing that "cross national studies conducted within a framework of comparable measurement can be a substantially more useful tool for policy analysis than studies of single countries." The National Science Foundation (NSF) Cyberinfrastructure Council describes "a vision in which science and engineering digital data are routinely deposited in well-documented form, are regularly and easily consulted by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved", and NSF's 2009 report on *Solving the Puzzle* of climate change reinforces the importance of cyberinfrastructure (National Science Foundation, 2007; 2009).

The TerraPop infrastructure will directly address these cyber- and data infrastructure needs. TerraPop will create a system for integrating and disseminating spatiotemporal data. It will preserve and integrate social and natural science data and metadata from across the globe, disseminate them across multiple research communities, and translate these digital data into usable knowledge. The data that will be incorporated into TerraPop have broad spatial and temporal dimensions. Most of the data are global in extent and span the past five decades, and some sources reach back to the nineteenth century.

The closely integrated population and environmental data that will be made available through TerraPop will allow us to describe the unfolding transformation of human and ecological systems. By creating a framework for locating, analyzing, and visualizing data describing the world's population and environment in time and space, TerraPop will provide unprecedented opportunities for investigating the agents of change, assessing their implications for human society and the environment, and developing policies to meet future challenges.

Source Data

By allowing researchers to easily obtain integrated data on both population and environment in formats conducive to their work, TerraPop will constitute a unique resource for investigating changes in the human-environment system. The initial phase of the project focuses on five specific kinds of data:

1) Census and survey microdata describing the characteristics of individuals and their families and households
2) Aggregate census and survey data, describing the population characteristics of places
3) Census and survey data describing land use and land cover characteristics of places
4) Remote-sensing data describing land cover and other environmental characteristics
5) Climate data describing temperature, precipitation, and other climate-related variables

The core of TerraPop's population data is the Integrated Public Use Microdata Series (IPUMS), the world's largest collection of spatiotemporal population data (Ruggles et al, 2008; Minnesota Population Center, 2009a). IPUMS combines census and survey data into one seamless database where all variables are coded consistently across countries and over time. Each record in IPUMS is a person, hence "microdata." IPUMS currently provides individual-level data on 545 million people from 238 censuses in 74 countries since 1960. By 2017, the database is expected to include roughly one billion records (see www.ipums.org). The IPUMS data are complemented by the National Historical Geographic Information System (NHGIS), the most extensive collection of small-area population data for the United States (Minnesota Population Center, 2009b).

TerraPop will also incorporate the world's largest collection of spatiotemporal data on agricultural acreage and yields, the Global Landscapes Initiative (GLI). GLI data include harvested area and yields for 175 crops in the year 2000 and cropland and pasture data back to 1700. These data are derived from census-based land-use and land-cover records in combination with remotely sensed information, enriching both forms of data (Ramankutty and Foley, 1998, 1999; Ramankutty et al, 2008; Monfreda, Ramankutty and Foley, 2008).

TerraPop will make IPUMS and GLI data interoperable with important global data sources on land cover and climate. The initial release of TerraPop will include data from the Global Land Cover 2000 and WorldClim datasets (Fritz et al, 2003; Hijmans et al, 2005). Both of these datasets provide global coverage with approximately 1-kilometer resolution. The GLC2000 dataset is based on SPOT-4 vegetation imagery and uses a globally standardized 22-class classification system based on the FAOs Land Cover Classification System (Di Gregorio and Jansen, 2000). The WorldClim dataset is derived from climate data from nearly 25,000 stations around the world and includes several temperature and precipitation variables.

The datasets being incorporated into TerraPop are obtained in three basic formats:

1) Microdata structured as records of individuals and households
2) Area-level data structured as records of places
3) Raster data structured as values arranged in a spatial grid

The flexibility inherent in population microdata makes them extremely powerful. Microdata records for individuals contain demographic and other variables such as age, sex, occupation, education, and migration. Records of individuals are organized into households, providing information about family structure and variables related to household characteristics such as utilities and dwelling materials. Figure 1 gives an example of census microdata, with persons organized into households (identified by the variable HHID) and having individual-level responses. The availability of data at the level of individuals allows researchers to investigate very specific relationships and to summarize data in ways that may not be possible with area-level data alone. For example, a researcher could analyze age by individual years, whereas area-level data may only be tabulated in five-year categories. The organization of persons into households also allows the creation of novel measures never found in area data, such as children's school attendance by the employment status of one or both parents.

Figure 1. Census Microdata

HHID	PERSON	AGE	SEX	RELATE	MARSTAT	EDUCATION
4023	1	42	Male	Head	Married	Tertiary
4023	2	40	Female	Spouse	Married	Secondary
4023	3	16	Female	Child	Single	Secondary
4023	4	11	Female	Child	Single	Primary
4023	5	2	Female	Child	Single	N/A
4024	1	29	Male	Head	Divorced	Secondary
4024	2	25	Female	Other relative	Single	Secondary
4024	3	2	Female	Child	Single	N/A
4025	1	61	Female	Head	Widowed	Primary
4025	2	12	Male	Grandchild	Single	Basic
4025	3	10	Female	Grandchild	Single	Basic
4026	1	41	Female	Head	Separated	Primary
4026	2	21	Male	Child	Single	Secondary
4026	3	16	Male	Child	Single	Basic
4026	4	14	Female	Child	Single	Basic
4026	5	12	Female	Child	Single	Basic

Although they offer unique analytic possibilities, population microdata do not generally contain information on environmental conditions. Because the data describe actual individuals, they are subject to confidentiality measures restricting geographic specificity in ways that complicate spatial analysis. Moreover, microdata are not available for all countries or time periods, so there are coverage gaps in both spatiotemporal dimensions.

Area-level data are available for both population and environmental characteristics and are commonly used for many types of analysis. Area-level population characteristics may be derived from censuses or surveys, and they are more broadly available than microdata. Figure 2 shows population counts of different fuels used for household lighting by district in 2008 Malawi. For many organizations, the area-level data are the primary product intended for public distribution, and the underlying microdata used to generate the area summaries are lost or deemed not suitable for dissemination. Thus, population data summarized to administrative units are often available for countries and time periods where microdata do not exist. Area data are also highly heterogeneous, including such things as agricultural censuses describing land use, economic censuses, and a host of indicators produced at the national level.

Figure 2. Area Data: Source of Fuel for Lighting, by District (N of persons)

District	Total	Electricity	Paraffin	Candles	Firewood	Other
Kasungu	606,028	18,971	529,669	17,861	17,068	22,459
Nkhotakota	293,452	9,873	266,975	3,489	7,369	5,746
Ntchisi	216,416	4,117	186,720	3,732	9,426	12,421
Dowa	541,227	11,005	480,662	9,654	19,404	20,502
Salima	327,117	13,417	295,616	3,992	7,564	6,528
Lilongwe	1,200,378	16,904	1,122,218	13,275	28,522	19,459
Lilongwe City	647,458	218,064	302,123	121,492	2,732	3,047
Mchinji	442,843	9,786	404,924	8,312	11,833	7,988
Dedza	607,995	11,062	562,890	6,306	18,306	9,431
Ntcheu	457,012	8,243	434,550	4,090	6,928	3,201

Though they are broadly available, area data lack the flexibility of microdata and raster data. There is an extra layer of decision-making built into the tabulation of the data that often cannot be unwound, such as the grouping together of categories or imposition of a particular universe for specific statistics. Those differences can make it difficult or impossible to create fully comparable statistics across times and places. The formats in which area data are produced also pose a serious technical challenge. Many of these data are organized as tables intended for human comprehension, not for interpretation by statistical software.

Many environmental datasets are produced as rasters, which divide a geographical area into grid cells and provide a data value, such as average annual temperature, for each cell. Satellite imagery and data derived from it, such as land cover, are typically produced in raster format. Figure 3 presents a stylized snippet of raster data describing the type of land cover predominating in each grid-cell. Data from point measurements, such as climate stations, may also be interpolated to produce raster data that covers the entirety of a geographical area. Raster data from remote sensing devices lack some of the richness of microdata and area data, providing a single value per grid cell. But the data have wide coverage, are geographically precise, and can be overlaid onto any set of boundaries.

Figure 3. Raster Data: Land Cover

1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	4	4	3	3	3
1	1	1	1	1	4	4	4	3	3
1	1	1	1	4	4	4	4	3	3
1	1	1	1	4	4	4	4	3	3
2	2	2	2	2	4	4	3	3	3
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2

1 = residential 2 = farmland 3 = wooded 4 = water

Data Preparation

Developing a system that can accommodate three distinct data types requires considerable processing and metadata development. Fortunately, TerraPop can build on the technical standards and methods developed for the IPUMS and NHGIS projects at Minnesota. A variety of software tools convert the original metadata into structures understood by the data processing and web software.

TerraPop will be designed at the outset to use the IPUMS population microdata and its associated metadata without modification. Because IPUMS has already harmonized the population data internationally, they come to TerraPop fully documented and well suited to the project's needs.

In the initial project phase, the area data will be tabulated from the IPUMS microdata at the national and subnational levels. New software is required to conduct these tabulations at scale and generate datasets understood by the TerraPop system, and metadata must be written to describe for users how the data were constructed. But the source microdata are fully consistent and documented, minimizing the challenges.

The raster data for the initial release come at different resolutions: 1 kilometer and 10 kilometer grid cells. To increase compatibility and minimize processing demands in the early stages of the project, the higher-resolution data—the climate and land-cover datasets—will be converted to 10-kilometer cells. The metadata accompanying raster datasets tends to be uneven and not consistently organized in a form that software can interpret. Some labor is required to prepare this material for purposes of documenting variables in the web dissemination system, and much must be written from scratch.

Geographic Harmonization

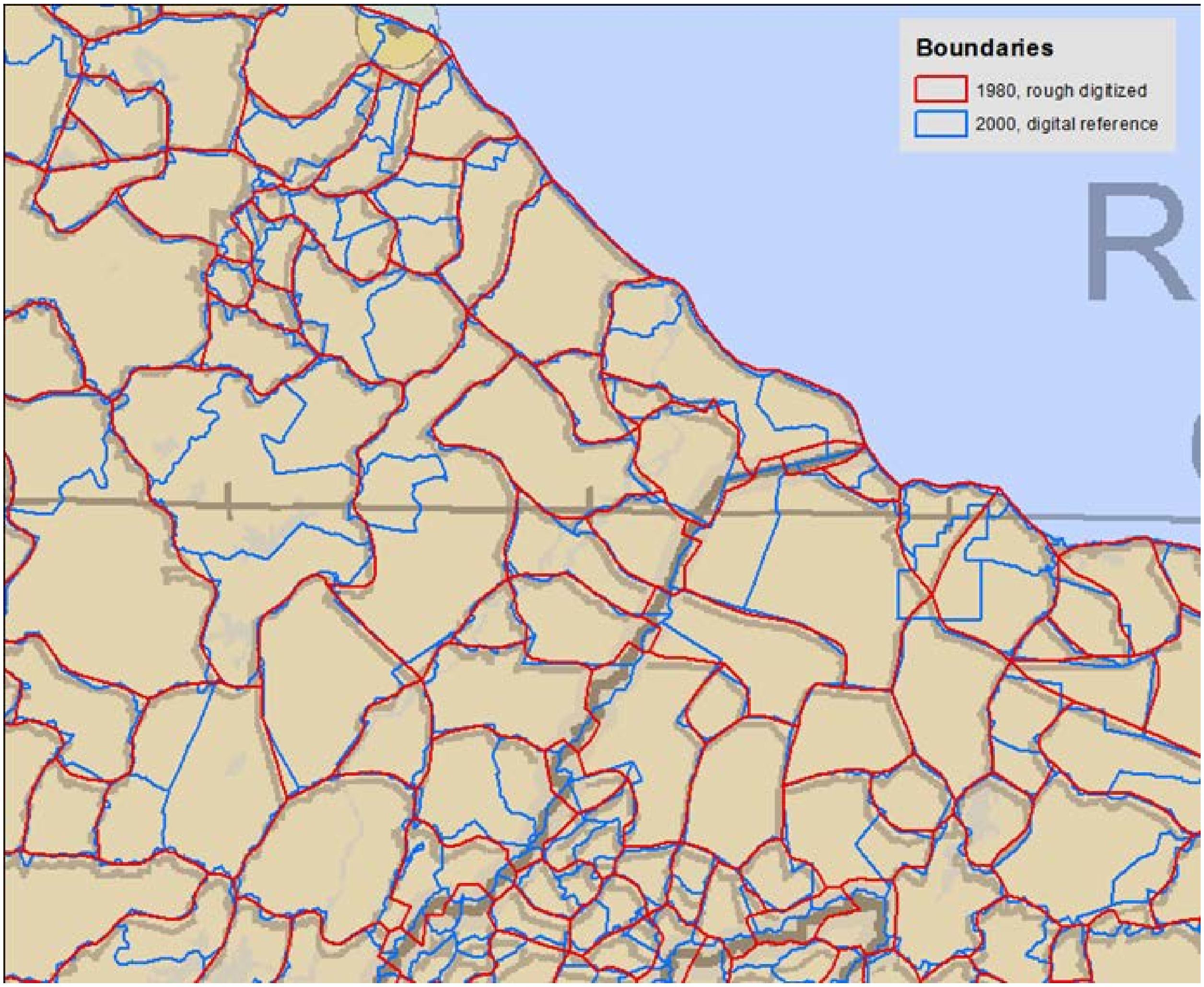
Processing the data types into TerraPop-ready form is an essential but insufficient condition for building an interoperable data system. The lynchpin of interoperability is geographic harmonization. The ability to link data across the three basic formats relies on spatial data delineating the boundaries of administrative and statistical units. Microdata records include codes identifying the unit within which the individual lives, and area-level data include similar codes identifying the unit described by the record. These codes can be used to link microdata and area-level data and connect both types of data to real-world locations through boundary data. The boundaries can then be overlaid on raster data to link all three data formats.

Accurate boundary information is necessary to combine data from different sources. Data derived from statistical agencies and administrative sources may use different spatial units, and codes within the datasets must be harmonized. The most serious challenge for TerraPop, however, is change over time in the boundaries of geographic units. Going back several decades, it is not unusual for governments to create new administrative divisions, for units to merge, or for boundaries to shift. Yet, for most analyses over time, one wants to hold the geographic space constant and study the changes in the attributes of the units, whether population, land cover, or climate.

Geographic harmonization is necessary to ensure that analyses are not subject to artifacts of boundary changes. The first step is to acquire modern digital boundary files that identify the administrative units in the datasets. Boundary files for older data typically do not exist, and it is necessary to digitize maps, using the modern boundaries as a base for editing. Most changes involve the creation of new units (i.e., splitting), which requires only erasing some new boundaries from the modern file to create the original units. But other changes do happen and can involve hand work drawing new lines. The final product of this process is a set of boundaries that match the units in every dataset for a country.

Once the foregoing is complete, one has accurate boundary data for a country, but some of those boundaries define different spaces over time, some units do not exist in all periods, and the names of places may have changed. Year-specific boundaries are sufficient for static applications where time is not an element in the study, but temporal analyses require that the units be harmonized over time. The method for doing this is to overlay the boundaries from different periods, combining units together until the outer boundary of the combined area defines a space that is identical in all times. Thus, if unit A in an early dataset later split into A and B, the harmonized unit would be A+B in all periods. Figure 4 overlays 1980 and 2000 municipio boundaries for a portion of Brazil. In practice, harmonization tends to have the biggest impact in the modern period, forcing the re-aggregation of places that split in the past, which is by far the most common type of change. The full unharmonized geographic detail is still available to researchers studying a single point in time, where boundary changes are irrelevant. Using harmonized TerraPop geography, it is possible to analyze change over sixty years in population characteristics over much of the world at the sub-national level.

Figure 4. Brazil Municipio Boundaries Super-imposed 2000 and 1980



A final element in geographic harmonization addresses complications arising from confidentiality procedures. Population microdata describe real people, and their privacy must be protected. Various measures are taken to ensure this, but the salient one for TerraPop is that no unit with fewer than 20,000 population can be identified in the microdata. Thus, at the first administrative level of, say, province, all units within a country may meet the minimum population threshold. But at the district level, many may not. The small districts need to be combined with other units to meet the 20,000 minimum. IPUMS, the source of the microdata, simply combines all the small units within the province into a single residual category; but the resulting aggregations may not be contiguous, and they can create oddly shaped aggregations that themselves are sometimes large enough to be subdivided, if one had a scientific basis for doing so. TerraPop addresses this problem by using a program called REDCAP to combine small units based upon empirical criteria (Guo, 2008). Preliminary study suggests population density and a rule forcing contiguity yield promising results for constructing sensible regions.

Data Access System

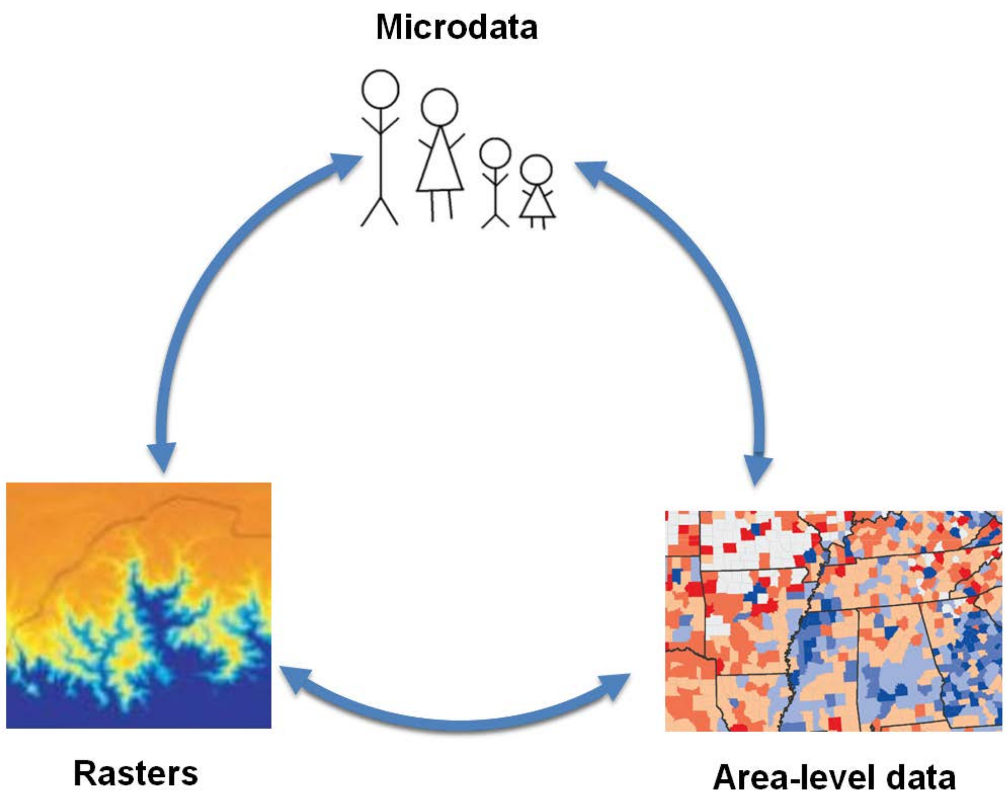
TerraPop must enable researchers to identify the data they need, choose a data format in which they want to work, and convert those data—whatever their original type—into the desired format on demand. The goal is to make it easy to construct datasets that combine data derived from all three formats:

(1) Microdata describing the characteristics of individuals, families, and households merged with variables describing the characteristics of the physical environments and social contexts in which they live

(2) Area-level data describing population and environmental characteristics of spatial units, including summaries of variables from both microdata and raster data

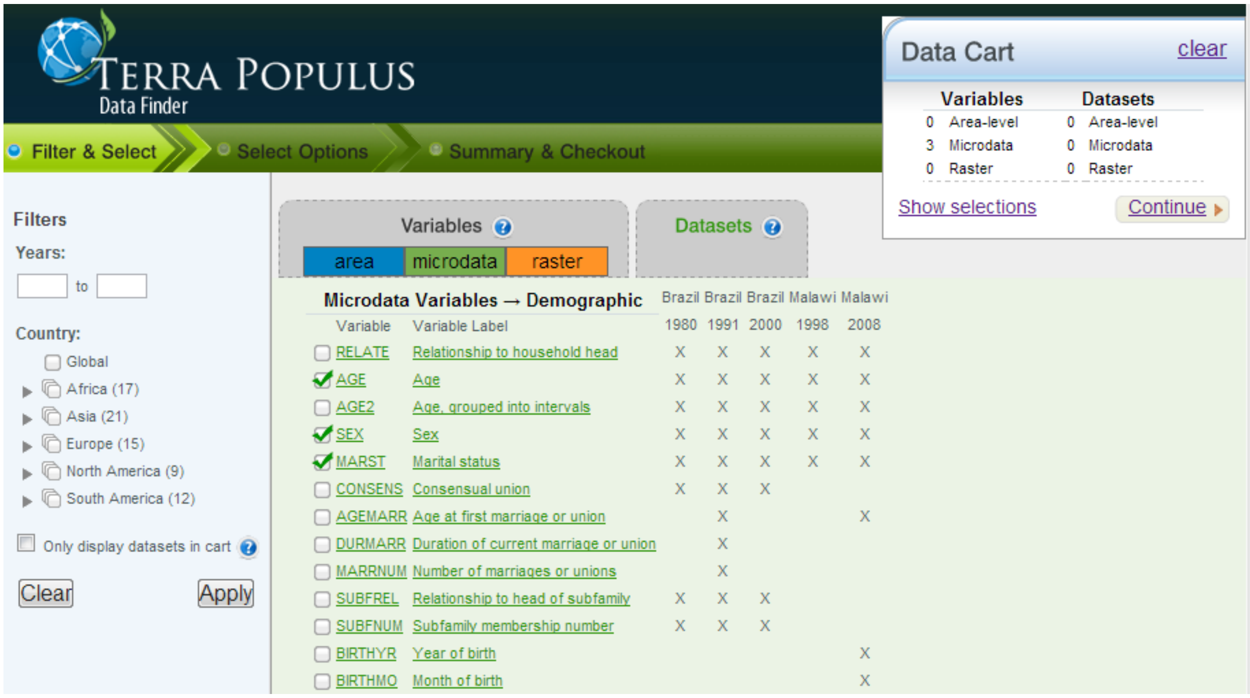
(3) Raster data that provide estimates of both environmental and population attributes

Figure 5. TerraPop Data Format Conversions



The TerraPop web system will allow variable browsing across the three data formats. Users select datasets to identify the places, times, and topics of interest to them. The system will restrict the variables displayed to only those selected datasets, preventing overloading the user with extraneous information. (It is also possible to view everything without filtering.) A drop-down menu organizes the variables by topic within each of the three data types. At a click, researchers have access to full metadata for each variable, as applicable to its data type: variable description, codes, frequencies, questionnaire text, how the data were constructed, etc.   Users click checkboxes next to the variables in the list to send them to the "data cart". Researchers can click on the cart at any time to see the variables and datasets they have selected. If they are satisfied with their choices in the cart, they click “continue” to initiate the steps necessary to finish defining their data extract.

Figure 6. TerraPop Data Browsing and Selection

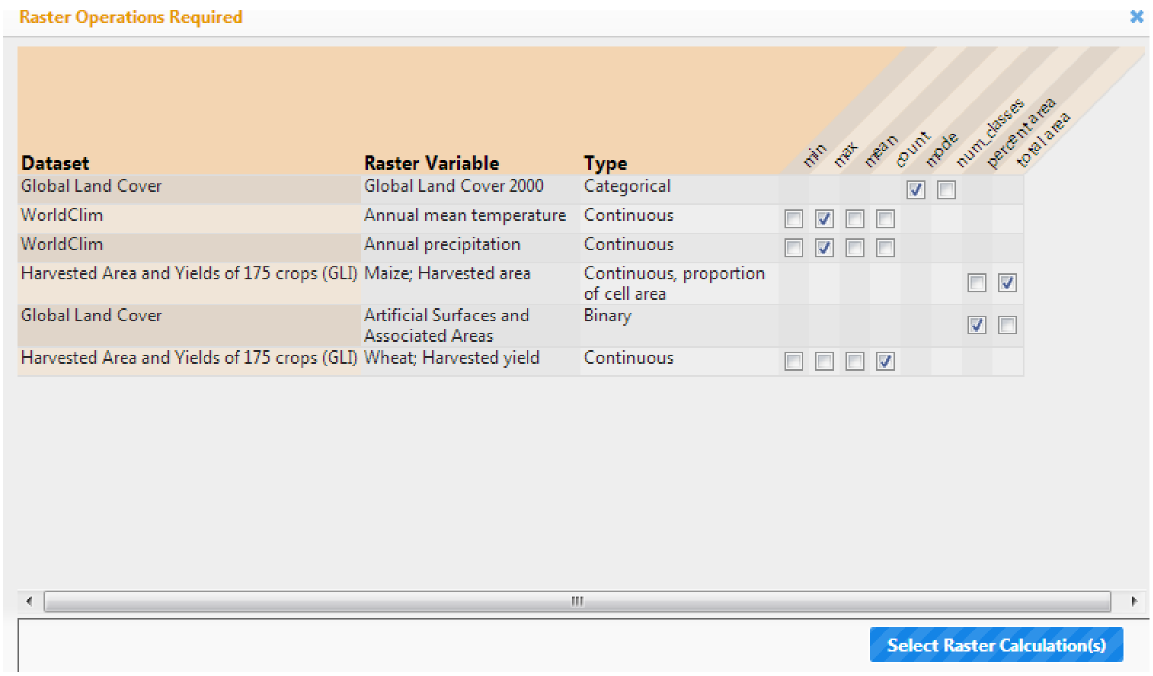


The first decision in the extract check-out process is the most fundamental: whether the user wants data output as microdata, area data or rasters. Subsequent steps depend on that choice and on the types of variables in the cart that need to be transformed. For most transformations, the user needs to specify the level of geography at which data need to be linked and/or summarized: country, first administrative, or second administrative level. Data are held in their native formats behind the scene and are only processed for data extracts upon request. There are six possible data conversions.

> 1) Area to microdata. No user specification is needed. Area-level variables, such as percent of workers engaged in agriculture, are simply appended to the microdata records that apply to their geographic space and time period. They become contextual variables for subsequent microdata analysis.
>
> 2) Area to raster. No user specification is needed. A raster grid is superimposed on the area boundaries, and each raster falling within a particular geographic unit inherits its area-variable value. In the initial release, all rasters within a given area will receive the same value. In future years, the method will be refined to consider factors like elevation, water features, and built-up areas in distributing area values such as population counts or agricultural production.
>
> 3) Raster to area. The raster cells falling within each area need to be summarized into a single area-level value. The user must specify the type of computation. Possible procedures vary by the nature of the raster data—categorical, binary or continuous—and

include sum, mean, mode, maximum, etc. The newly calculated variable, such as percent of land area in forest, is provided as a new variable on each area record. Figure 7 shows the kind of interface envisioned for this step.

Figure 7. Raster Operation Screen



4) Raster to microdata. The same steps are required as raster-to-area, then the resulting values are distributed to all of the microdata records within each area.

(The final two conversion possibilities will not be possible in the initial release of TerraPop, because they require a new tabulation system that can accept complicated user-defined parameters. This is a major project deliverable for the future.)

5) Microdata to area. If a user specifies area output, microdata variables need to be tabulated to the desired geographic level: for example, percent of population who are literate at the second administrative level. The TerraPop interface must give researchers the flexibility to combine multiple microdata variables to define the measures they want, and with the desired denominators. They need to be able to name the variable as well, so that proper metadata can be generated, such as labels for statistical software. The system must then be able to rapidly calculate the result for each geographic unit and attach it to the appropriate person records as a new person variable. The tabulating system is a substantial software challenge in terms of user interface design and calculation engine.

6) Microdata to raster. This conversion requires the same steps as microdata to area (i.e., first defining the desired measure at the area level). Once the summary area measure is defined, a raster grid is superimposed, and each raster falling within a particular geographic area receives its value. Because this conversion requires the same tabulator as microdata-to-area calculations, it is the other type of transformation not available in the first TerraPop release.

After researchers specify the necessary data transformations, they submit their extract request, and the TerraPop backend software proceeds to construct the dataset. Behind the scenes, the data are held in their native formats, and this is what the TerraPop system works from. When processing is complete, the system sends an email to the user indicating that their data are ready to download. Microdata and area data extracts come as a single file that pools records across time and space. Each raster variable is produced as its own file, as is the norm for that format. The TerraPop system will produce files in CSV, SAS, SPSS and STATA formats. It will also provide GIS boundary files corresponding to the geography in the extract.

The TerraPop data system will support a wide array of research topics. The following few examples offer a small sample of what we envision.

- Analyze health outcomes in sub-Saharan Africa at the district level. Calculate rainfall and temperature for areas from the WorldClim raster data, and measures of child mortality, socio-economic status, and sewage disposal from the IPUMS microdata.
- Model deforestation and agriculture in the Yucatan, Mexico. Calculate area in forest and crops from the GLC2000 land use dataset at the municipality level. Combine that with climate raster data from WorldClim. Summarize microdata at the area level indicating the types of industries in which people are employed, as well as schooling and other socio-economic factors.
- Study the effect of specific crops on school attendance patterns in Latin America. Use population microdata for school attendance, parental employment, and other household factors. Add contextual variables calculated from the GLI crop raster data, indicating the intensity of various staple crops in the local area.

After the Prototype

The foregoing has described the core features of the TerraPop system intended for release in late 2013-2014. The current project period extends an additional three years, offering time to make many improvements. Some of those are planned, others will be in response to user feedback – both spontaneous and actively solicited from the TerraPop Development Community.

The most important planned addition is the aforementioned tabulator that lets users define new summary variables from microdata in a flexible and intuitive way. Another improvement mentioned above is a smarter way of converting area data to rasters that does not simply distribute the same value across all cells within the unit.

Geographic harmonization will continue through the life of the project, and the key to that work is acquiring or making digitized boundaries. Harmonizing the first administrative level within each country over time is manageable. The great challenge going forward is to harmonize the second level, which is much more difficult. There are far more units at that level, and changes are much more frequent over time. We have developed methods using second level Brazil and Malawi as case studies, but getting the raw materials to map the world over many decades is a bigger job than TerraPop can fully accomplish on its own in the near term. It will remain a work in progress for the foreseeable future.

It is also important that the project develop methods for processing the summary data produced by national statistical offices. Those are generally distributed as tables when countries publish their census results, and they exist in myriad formats not conducive to easy data manipulation.

The U.S. census area data have already been transformed by the Minnesota Population Center into a useful form, and adding those NHGIS data to TerraPop will be a priority, as there is considerable research interest. The tools developed for NHGIS will also be useful for processing some digital summary data distributed by national statistical offices. Some area data, especially older data, may eventually need to be scanned and/or manually entered, because they were never produced in digital form. The data are essential, however, because microdata will never exist for many countries and time periods.

We expect to add many more raster data in the coming years, including additional land cover, historical climate, elevation, and gridded population datasets. The range of potential area data is even greater, and could include topics such as economics, health, politics, and even laws: anything that can be associated with particular administrative areas or countries.

Summary

TerraPop is intended to remove obstacles to spatiotemporal research, especially those involving population and the environment, which have grown into an area of intense scholarly interest in recent years. Researchers across many disciplines could benefit from each other's data collections if a finite set of technical challenges can be met. We believe we have identified the key problems that must be solved, and TerraPop is the platform on which we intend to implement the best solutions we can devise. This will be a difficult and somewhat iterative process, but the rewards for success should be considerable for the research community.

The Minnesota Population Center has developed numerous data infrastructure projects, but the TerraPop system is the most complex we have undertaken. Despite its ambitiousness, in less than two years we have gone from nothing to a functional, if limited, prototype system that has many of the key functionalities we envision. And within months we intend to make the system available to external users. In our experience, that is quite rapid software development for a project of this scope, and it encourages us that we have a sound approach if, still, many challenges ahead.

With TerraPop, data producers can reach a wider audience for the product of their labors. It will be easier for population and environment researchers to work collaboratively across disciplines, and to leverage each other's data collection efforts. With less time manipulating data into amenable formats, fewer resources will be expended in the least productive areas of fledgling research projects. Researchers can focus on the substantive analytical issues rather than logistics. Most importantly, we hope that new kinds of science can be carried out and new discoveries made.

References Cited

Di Gregorio A. and Jansen L.J.M. *Land Cover Classification System. Classification Concepts and User Manua*l. 2000. FAO. Rome. http://www.fao.org/docrep/003/x0596e/X0596e01.htm

Ferreira, F. and M. Ravallion, Poverty and Inequality: The Global Context, in *The Oxford Handbook on Economic Inequality*, W. Salverda, B. Nolan, and T. Smeeding, Editors. 2009, Oxford University Press.

Fischlin, A., G.F. Midgley, J.T. Price, R. Leemans, B. Gopal, C. Turley, M.D.A. Rounsevell, O.P. Dube, J. Tarazona, and A.A. Velichko, Ecosystems, Their Properties, Goods and Services, in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden, and C.E. Hanson, Editors. 2007, Cambridge University Press, Cambridge.

Foley, J.A., R. DeFries, G.P. Asner, C. Barford, G. Bonan, S.R. Carpenter, F.S. Chapin, M.T. Coe, G.C. Daily, H.K. Gibbs, J.H. Helkowski, T. Holloway, T. Howard, E.A. Howard, C.J. Kucharik, C. Monfreda, J.A. Patz, I.C. Prentice, N. Ramankutty, and P.K. Snyder, Global Consequences of Land Use. *Science*, 2005. 309: 570-574.

Fritz, S. et al. *Global Land Cover 2000*. European Commission Joint Research Centre. 2003. http://bioval.jrc.ec.europa.eu/products/glc2000/products.php. Accessed 4/6/12.

Guo, D. Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP)". *International Journal of Geographical Information Science*. 2008. 22: 801-823.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. Very High Resolution Interpolated Climate Surfaces for Global Land Areas. *International Journal of Climatology* 2005. 25: 1965-1978. http://www.worldclim.org/. Accessed 4/6/12.

IPCC, Summary for Policymakers, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller, Editors. 2007, Cambridge University Press Cambridge, United Kingdom and New York, NY, USA.

Lutz, W. and A. Goujon, The World's Changing Human Capital Stock: Multi-State Population Projections by Educational Attainment. *Population and Development Review*, 2001. 27(2): 323-339.

Minnesota Population Center, *Integrated Public Use Microdata Series–International: Version 5.0.* 2009, Minneapolis: University of Minnesota. https://international.ipums.org/international/.

Minnesota Population Center, *National Historical Geographic Information System.* 2009, Minneapolis, MN: University of Minnesota. http://www.nhgis.org/.

National Research Council, *Preparing for an Aging World: The Case for Cross-National Research*. 2001, Washington D.C.: National Academy Press.

Monfreda, C., N. Ramankutty, and J.A. Foley, Farming the Planet: 2. Geographic Distribution of Crop Areas, Yields, Physiological Types, and Net Primary Production in the Year 2000. *Global Biogeochemical Cycles*, 2008. 22: GB1022, doi:10.1029/2007/GB002947.

National Science Foundation, Cyberinfrastructure Council. *Cyberinfrastructure Vision for 21st Century Discovery*. 2007; Available from: http://www.nsf.gov/pubs/2007/nsf0728/index.jsp. Accessed 5/10/09.

National Science Foundation, *Solving the Puzzle: Researching the Impacts of Climate Change Around the World.* Released May 14, 2009. Available from: http://www.nsf.gov/news/special_reports/climate/pdf/NSF_Climate_Change_Report.pdf.

O'Neill, B., C.F. MacKellar, and W. Lutz, *Population and Climate Change*. 2001, Cambridge: Cambridge University Press.

Ramankutty, N. and J.A. Foley, Characterizing Patterns of Global Land Use: An Analysis of Global Croplands Data. *Global Biogeochemical Cycles*, 1998. 12(4): 667-685.

Ramankutty, N. and J.A. Foley, Estimating Historical Changes in Global Land Cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles* 1999. 13(4): 997-1027.

Ramankutty, N., A.T. Evan, C. Monfreda, and J.A. Foley, Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000. *Global Biogeochem. Cycles*, 2008. 22: GB1003, doi:10.1029/2007GB002952.

Ruggles, S., M. Sobek, J.T. Alexander, C. Fitch, R. Goeken, P.K. Hall, M. King, and C. Ronnander, *Integrated Public Use Microdata Series: Version 4.0* [Machine-Readable Database]. 2008, Minnesota Population Center [producer and distributor], Minneapolis, Minnesota. http://usa.ipums.org/usa/.

United Nations. *World Population Prospects.* 2006; Available from: http://www.un.org/esa/population/publications/wpp2006/WPP2006_Highlights_rev.pdf. Accessed 5/14/09.

World Bank. *World Development Indicators* 2009: Washington D.C. http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:21725423~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html.