

Terra Populus: Integrated Data on Population and the Environment

Steven M. Manson, Associate Professor of Geography, Environment, and Society, University of Minnesota

Tracy Kugler, Research Associate and Terra Populus project manager, Minnesota Population Center

David Van Riper, Spatial Analysis Core director, Minnesota Population Center

The goal of Terra Populus (TerraPop) is to lower barriers to conducting interdisciplinary human-environment interactions research by making data with different formats from different scientific domains easily interoperable. TerraPop is developing organizational and technical infrastructure to integrate, preserve, and disseminate data describing human population and the environment over time. The project has been funded for 2012 through 2018 with \$8 million (US) as part of the National Science Foundation's DataNet program. TerraPop uses location-based integration to combine information from individual-level and aggregate census data from 238 samples in 74 countries along with global datasets describing land cover, land use, climate, hazards, and other environmental variables. In addition to offering a large amount of data, TerraPop is developing new methods and open-source software for integrating data from different domains and translating across data structures based on spatio-temporal linkages among data contents.

The infrastructure created by TerraPop enables scholars and policy makers to identify and merge data from heterogeneous sources to study the relationships between human behavior and the natural world. TerraPop invites users to utilize data originating in formats that would normally be unfamiliar. Users are able to obtain datasets in familiar formats that fit into their existing work flows, yet combine information from microdata, area-level and/or raster datasets. For example, sociologists may want census microdata with contextual information about land cover attached to each person-record, while geographers may want aggregate census data along with land use information summarized for administrative units, or ecologists may want census data represented in a raster format to use in environmental models.

TerraPop's data access system is flexible enough to accommodate a broad variety of needs but comes at the cost of having to deal with the substantial costs associated with managing and processing terabytes of data. TerraPop therefore provides a strong use case for many CyberGIS capabilities. For example, constructing a customized dataset may involve performing zonal operations over dozens of rasters for thousands of polygons. In addition, TerraPop will serve as a mechanism through which a broad audience of researchers can access CyberGIS capabilities without requiring specific expertise in GIS or cyberinfrastructure. In addition to providing a brief overview of TerraPop, we describe critical technological challenges and solutions within the context of the broader realm of CyberGIS research.