

Integrated Ground-based and Remotely Sensed Data to Support Global Studies of Environmental Change¹

R. J. Olson, R. S. Turner, and C. T. Garten
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37931 USA
September 15, 1994

RECEIVED
MAY 01 1996
OSTI

Data centers routinely archive and distribute large databases of high quality and with rigorous documentation but, to meet the needs of global studies effectively and efficiently, data centers must go beyond these traditional roles. Global studies of environmental change require integrated databases of multiple data types that are accurately coordinated in terms of spatial, temporal and thematic properties. Such datasets must be designed and developed jointly by scientific researchers, computer specialists, and policy analysts. The presentation focuses on our approach for organizing data from ground-based research programs so that the data can be linked with remotely sensed data and other map data into integrated databases with spatial, temporal, and thematic characteristics relevant to global studies. The development of an integrated database for Net Primary Productivity (NPP) is described to illustrate the process.

DAACs

The United States Government has initiated the U.S. Global Change Research Program to develop a predictive understanding of the global environment. A pivotal part of the program is the National Aeronautics and Space Administration's (NASA) Mission to Planet Earth with the Earth Observing System (EOS). The NASA Earth Observing System Data and Information System (EOSDIS) manages data from satellites and field measurement programs. EOSDIS is composed of nine Distributed Active Archive Centers (DAACs) that provide data to the global change research community, policy makers, educators, and interested members of the public. Each DAAC focuses on an Earth science discipline and specializes in one or several types of data and data products, which are often associated with specific NASA flight missions. The DAACs became operational in July 1994 to form a physically distributed archive and distribution system linked through a system-wide information management system.

ORNL DAAC

The mission of the Oak Ridge National Laboratory (ORNL) DAAC is to archive and distribute data pertaining to the Earth's biogeochemical dynamics, specifically data acquired from ground-based measurements of biological and chemical interactions among the elements that comprise the Earth system.

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

¹ Research sponsored by the National Aeronautics and Space Administration under Interagency Agreement DOE No. 2013-F044-A1 under Martin Marietta Energy System, Inc., contract DE-AC05-84OR21400 with the U.S. Department of Energy.

The ORNL DAAC provides (1) an information management system that is integrally linked to the other DAACs, (2) a data archive and distribution system, (3) a user services function, (4) value-added products that summarize and synthesize biogeochemical dynamics data from around the globe, and (5) the capability to provide guidance and support for data management, data quality assurance, and data integration for field experiments. Preparing data entails acquisition, quality assurance, documentation, and archiving to produce complete packages of data and metadata for distribution.

Data available through the ORNL DAAC include data from ground-based, NASA-funded projects as well as data generated by other agencies in the United States and other countries. Current and future holdings include data from the First ISLSCP Field Experiment (FIFE), the Oregon Transect Ecosystem Research (OTTER) Project, the Boreal Ecosystem-Atmosphere Study (BOREAS), and the Carbon Dioxide Information Analysis Center (CDIAC). The FIFE project collected data to understand the biophysical processes controlling the exchanges of radiation, moisture, and carbon dioxide between the land surface and the atmosphere; to develop and test remote-sensing methodologies for observing these processes at a pixel level; and to help understand how to scale the pixel level information to regional scales commensurate with the modeling of global processes. The OTTER project estimated major fluxes of carbon, nitrogen, and water in forest ecosystems using an ecosystem-process model driven by remotely sensed data. The BOREAS project investigates the interactions between the boreal forest biome and the atmosphere. CDIAC provides access to information related to atmospheric trace-gas concentrations and global climate change. The preponderance of the CDIAC data deal with historic and atmospheric carbon dioxide and methane concentrations and historic weather and climate readings from throughout the world.

Integrated Databases

Environmental data are collected by a variety of agencies and organizations for specific mission-oriented requirements. To be useful for assessing global environmental issues, it is necessary to assemble and organize selected data into integrated databases. Scientists use integrated data to study across systems, parameterize models, verify model output, and conduct regional assessments. Decision makers require ready access to information on a variety of topics to develop policy and often do not have the time to wait for the completion of definitive scientific studies. Educators need easy access to information for teaching. To meet these user needs, the ORNL DAAC goes beyond the traditional data center role to design and compile integrated databases that meet the needs of these users.

Integrated databases are packaged datasets with an organizing framework; that is, each component dataset conforms to a common set of characteristics, including:

- Spatial** - both extent of coverage (e.g., Europe) and resolution (e.g., countries or common pixel size),
- Temporal** - both extent of time period (e.g., 1980s) and resolution (e.g., annual averages or totals), and
- Thematic** - general level of detail (e.g., if land cover is classified into a few major types then soils would be aggregated to major categories).

Defining the framework involves working with the data user community to incorporate their needs for the integrated database. The iterative design process must often reflect diverse user needs, unknown future needs, and data availability. Although the design and resulting database may not meet all users' needs, it usually will provide users with an initial data resource that can be expanded for more specific needs.

Integrated databases usually reside in a single database management system with tools for retrieval, report generation, analysis, display, geographic information systems (GIS), and export functions. Metadata document the source of each data component, describe the data collection and processing, and define data characteristics. Compiling the integrated database may require an extensive effort for each component, including acquiring data, reformatting, performing QA checks, converting to common units of measure, assigning common code values, resolving problems with the data generator, aggregating or extrapolating data to common spatial and temporal units, and documenting sources of data and processing (see Carter and Diamondstone 1990). The processing and packaging of the data add value to the original component data sets in terms of consistency, completeness, documentation, and availability for secondary use of the data.

NPP Database

Data on spatial patterns of terrestrial net primary production (NPP) and on rates of carbon accumulation are essential for a fuller understanding of the potential extent and impacts of global changes. NPP is the amount of organic matter synthesized per unit of area of the Earth's surface per unit of time. The spatial pattern of mean NPP varies greatly over the Earth from 0 g dry mass/m² per year in extreme deserts to 3000 g dry mass/m² per year in swamps and marshes. Uncertainty in the current estimates of carbon storage in the Earth's vegetation and soils may account for the carbon that currently is unaccounted for in the global carbon budget.

Potential uses of a NPP database are the calibration and validation of remotely sensed patterns and model predictions of terrestrial productivity associated with global change. Successful remote sensing of terrestrial carbon dynamics requires the development of methodologies and models that relate the spectral properties of plant canopies to ecosystem processes like NPP. Ground-based measurements of NPP are essential for meeting the challenge of scaling up from small ecosystem studies to large regional scales. To meet these needs, the ORNL DAAC is working with several groups to design and develop an integrated NPP database, including the ORNL DAAC User Working Group and modelers participating in the 1994 Model Inter-Comparison Workshop for Global Terrestrial Net Primary Productivity. A major recommendation from that workshop was to develop common datasets to parameterize the global models.

The NPP database will provide in situ measurements from terrestrial ecosystems for merging with remote sensing measurements from aircraft and satellite platforms. For example, the Pathfinder Advanced Very High Resolution Radiometer (AVHRR) Land Data Set at Goddard Space Flight Center contains NDVI (normalized difference vegetation index) data from satellites. A recent study used a parametric approach to combine estimates of canopy absorption efficiency of incoming solar radiation from NDVI data with estimates of conversion efficiency to calculate the production of organic matter (Ruimy et al. 1994). The estimates of conversion efficiencies for broad biomes were obtained from the

literature. Although the modeled results generally agreed with previous estimates of NPP, significant differences appeared in the spatial patterns. The authors identify the lack of available data as one source of uncertainty in their results. In recognition of the need for this type of data, the International Union of Forest Research Organizations (IUFRO) has initiated an international cooperative project to coordinate data formats of studies of woody plant productivity.

The NPP database will be compiled from the scientific literature and from regional thematic maps. Information in the database will include (when available from published sources) the following: physical location, time of study, elevation, annual precipitation, plant community type, biome, country or political division, principal investigator or author, reference or publication, plant biomass, NPP, associated measurements (e.g., leaf-area index, solar radiation, plant nitrogen), soil type, and funding agency. Distinctions will be made between aboveground, belowground, and total plant carbon stores. The initial emphasis will be on acquiring data from major terrestrial biomes from published compilations such as Cannell (1982) and DeAngelis et al. (1981).

The NPP database will be available from the ORNL DAAC. Scientists needing data sets for global analysis, methods development, or simulation modeling are encouraged to make full use of resources available from the DAAC. In addition to data that are being gathered for specific EOSDIS goals, environmental scientists also have an opportunity to archive their data in the ORNL DAAC. The staff at the ORNL DAAC will work with scientists entering data and metadata into the system. Contributors will be acknowledged as data base "authors." Individuals wishing to contribute data or to learn more about the data center can inquire by e-mail at ornldaac@ornl.gov, by phone at 615-241-3952, or by FAX at 615-574-4665.

References

- Cannell, M.G.R. 1982. *World Forest Biomass and Primary Production Data*. Academic Press. London.
- Carter, G.C. and B. I. Diamondstone. 1990. *Directions for Internationally Compatible Environmental Data*. Hemisphere Publishing Corp., New York, 401 pp.
- DeAngelis, D.L., R.H. Gardner, and H.H. Shugart. 1981. Productivity of Forest Ecosystems Studied during the IBP: The Woodlands Data Set. In: D.E. Reichle (ed.): *International Biological Programme 23, Dynamic Properties of Forest Ecosystems*. Cambridge University Press, Cambridge London, pp. 573-672.
- Ruimy, A., G. DeDieu, and B. Saugier. 1994. Methodology for the Estimation of Terrestrial Net Primary Production from Remotely Sensed Data. *J. of Geophysical Research*, 99:5263:83.

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.