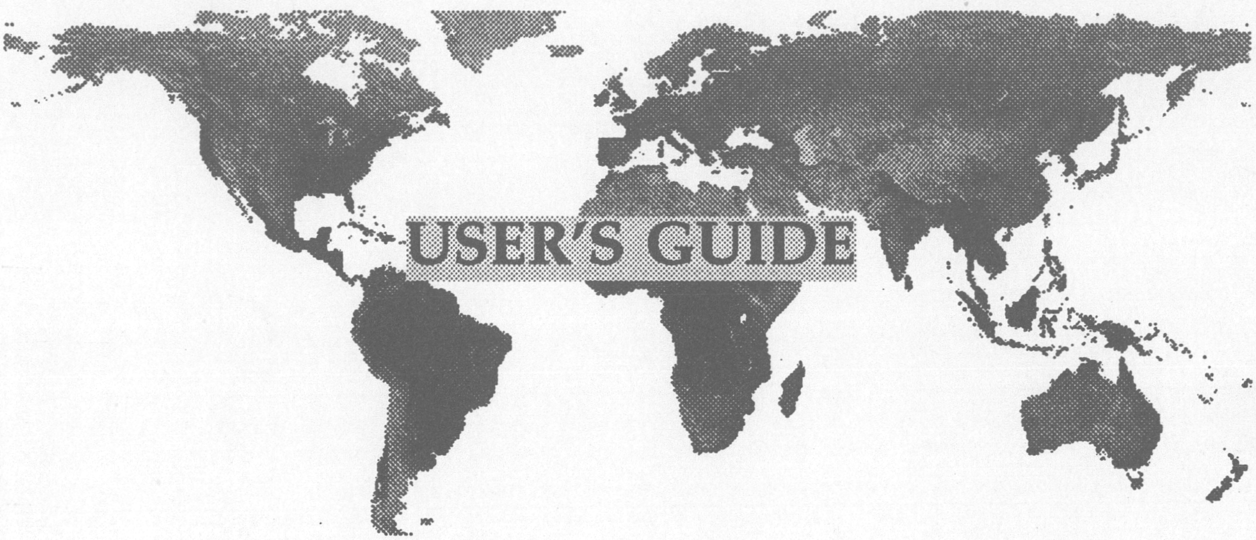
GLOBAL ECOSYSTEMS DATABASE
Version 1.0 (on CD-ROM)

EPA Global Climate Research Program
NOAA/NGDC Global Change Database Program



USER'S GUIDE

EPA/600/R-92/194a
NGDC Key to Geophysical Records Documentation No. 26
Incorporated in: *Global Change Database - Volume 1*

EPA
U.S. Environmental Protection Agency
Environmental Research Laboratory
Corvallis, Oregon

U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Geophysical Data Center and
World Data Center A for Solid Earth Geophysics
Boulder, Colorado

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NOAA
U.S. DEPARTMENT OF COMMERCE

GLOBAL ECOSYSTEMS DATABASE
Version 1.0 (on CD-ROM)

EPA Global Climate Research Program
NOAA/NGDC Global Change Database Program

USER'S GUIDE

EPA/600/R-92/194a
NGDC Key to Geophysical Records Documentation No. 26
Incorporated in: *Global Change Database - Volume 1*

John J. Kineman

June, 1992
(with corrections through September, 1992)

Produced in cooperation with the US EPA under Interagency Agreement (Contract No. DW13934786-01-0) titled "Co-developing data, tools, and methods for characterization and analysis of environmental system patterns to support EPA Global Climate Change Research and Modeling."

United States Department of Commerce
National Oceanic and Atmospheric Administration
National Geophysical Data Center
325 Broadway
Boulder, Colorado 80303

Disclaimer

The information in this document has been funded in part by the U.S. Environmental Protection Agency (EPA) under Interagency Agreement (DW13934786-01-0) to the National Geophysical Data Center (NOAA). It has been subjected to the agency's peer review, and it has been approved for publication as an EPA and NOAA document. While every effort has been made to ensure that the data accompanying this documentation, as well as the documentation itself, are properly represented given the limitations of the original data and the current state of the art in data integration, The U.S. Government cannot assume liability for any damages caused by inaccuracies in the data or documentation, or as a result of the failure of the data or software to fulfill a particular purpose. The U.S. Government makes no warranty, expressed or implied, nor does the fact of distribution constitute a warranty.

Copyright Notice

Trademarks

Mention of a commercial company or product in this document does not imply endorsement by the U.S. Government, any of its agencies, or any sponsor or participant of the project for which this document was produced. All brand or product names are trademarks or registered trademarks of their respective companies.

Trademark information indicated below was derived from various sources and is accurate to the best of our knowledge:

ARC/INFO/GRID, and ESRI are registered trademarks of Environmental Systems Research Incorporated.
DR-DOS is a registered trademark of Digital Research, Inc.
ERDAS is a registered trademark of Earth Resources Data Analysis System, Inc.
IBM, PS/2, and AT are registered trademarks and EGA, VGA, and 8514A are trademarks of International Business Machines Corporation.
IDRISI is a registered trademark of Clark University.
INTEL, 80286, 80386, and 80486 are trademarks of Intel Corporation
GRASS is a registered trademark of the U.S. Army Construction Engineering Research Laboratory (CERL)
GRASP and GRASPRT are trademarks of Paul Mace Software, Inc.
MS-DOS is a registered trademark of Microsoft Corporation.
OSU-MAP-for-the-PC is a registered trademark of Ohio State University.
SPANS is a registered trademark of Tydac Technologies Corp.
UNIX is a registered trademark of Bell Laboratories, Inc.
WINDOWS is a trademark of MicroSoft Corporation

TABLE OF CONTENTS

Preface

The US Environmental Protection Agency (EPA), Environmental Research Laboratory - Corvallis, Oregon (ERL-C), established an Interagency Agreement with the US National Oceanic and Atmospheric Administration (NOAA), National Geophysical Data Center (NGDC) in September 1990. This agreement began a five year cooperative effort to develop a geographic database for modeling terrestrial climate-biosphere interactions in support of EPA's Global Climate Research Program. Although performing specific tasks under contract to the US EPA, NGDC independently operates a Global Change Database Program (GCDP) as part of its NOAA mission. Considerable synergism therefore exists between the tasks performed for the EPA under the "Global Ecosystems Database Project," and other activities supporting NOAA Climate and Global Change Program.

Within this contract the following cooperative goals have been established:

- Assemble and distribute periodic updates of the database in answer to EPA needs assessments, incorporating existing data and priority data developments, compiling appropriate and complete documentation, providing quality assurance and quality control, and ensuring adequate scientific review.

- Assess the data needs for global change modeling and research, and determine future directions for database and software systems developments.

- Pursue cooperative linkages and exchanges with parallel national and international global change database activities.

Because this database was constructed from many pre-existing data sources of varying quality; errors and omissions can be expected. All data are provided for experimental use, and to initiate a process of evaluation and improvement. Caution is advised when applying these data and computer programs. Users should make special note of limitations mentioned in the *Documentation Manual* and peer-reviews published in the *User's Guide*. For example, problems are endemic in the use of uncorrected Normalized Difference Vegetation Index (NDVI), such as the NOAA Global Vegetation Index (GVI), for time-series analysis, and suitable corrections are still being researched. This and similar issues associated with the various data-sets should be the focus of initial study, along with consideration of potential methods for intercomparison (referring here to comparisons between different data-sets), validation, and empirical calibration using Geographic Information System (GIS) tools and multi-thematic analysis.

GED Materials Checklist

Please confirm that you have a complete set of materials that comprise the Global Ecosystems Database. These are:

1. User's Guide (June 1992)
2. GED Version 1.0, CD-ROM Disc A (June, 1992)
3. Disc-A Documentation Manual (June, 1992)
4. DOS Access and Exploration Software (floppy disk), Version 1.0
5. IDRISI-Explorer Technical Reference Manual, Version 1.0

Also available on request:

1. IDRISI 3.0 Headers (supplemental floppy disk)
2. GRASS 4.0 Headers (supplemental floppy disk)
3. Project 3-ring binder

As of this printing, Disc B of GED Version 1.0 was in preparation for pre-release review, and scheduled for public availability December, 1992. Please contact NGDC for information on the content and availability of Disc B or subsequent additions.

Technical memoranda about the database may be issued on occasion. In addition there is an electronic bulletin board for communications about the Global Ecosystems Database Project and peer-review. Contact NGDC for information about electronic bulletin-board and on-line data services.

Citing These Materials

This database and documentation may be cited as a complete set as follows:

NOAA-EPA Global Ecosystems Database Project. 1992. *Global Ecosystems Database Version 1.0.* User's Guide, Documentation, Reprints, and Digital Data on CD-ROM. USDOC/NOAA National Geophysical Data Center, Boulder, CO.

The manuals may be cited individually, as follows:

Kineman, J.J. 1992. *Global Ecosystems Database Version 1.0, User's Guide. Key to Geophysical Records Documentation No. 26.* USDOC/NOAA National Geophysical Data Center, Boulder, CO. 121p.

Kineman, J.J., M.A. Ohrenschall, et al. 1992. *Global Ecosystems Database Version 1.0: Disc A, Documentation Manual. Key to Geophysical Records Documentation No. 27.* USDOC/NOAA National Geophysical Data Center, Boulder, CO. 240p.

You may also cite the individual data-sets themselves. The recommended citation for each data-set is given in the DATA-SET DESCRIPTION section of each chapter in the appropriate *Documentation Manual* for a given CD-ROM disc. Literature citations are also provided in that section.

System Requirements

The Global Ecosystems Database was designed for maximum platform independence using a standard ISO 9660 CD-ROM format and relatively generic file formats. All data files use common data types that are accessible at the operating system level. Nevertheless, there is no universal standard for all systems.

Version 1.0 of the database is structured for compatibility with the IBM-PC/DOS environment. DOS executable software is provided with the CD-ROM on an accompanying floppy disk (but is not required to access the data files). The software provides many convenient exploration functions for DOS users, plus export capabilities for UNIX systems. Direct access from other operating systems is possible, but may require conversion of data storage types for correct numerical interpretation.

The software provided on floppy disk requires an IBM-PC compatible computer running IBM-DOS, MS-DOS, or DR-DOS, with at least 256K of memory and EGA, VGA, or 8514A graphics (IBM 8514A compatible graphics is required for 256 color display). For optimal performance, an Intel 80286, 80386, 80486 or compatible processor with 640K of memory and a graphics accelerator card is recommended. A CD-ROM reader is required to access the disc. Since data access is slow from CD-ROM readers, adequate disk space to download portions of the database is desirable. Various printers are supported.

Technical Support

NGDC staff are available to assist with inquiries about data and software availability, and with any technical problems concerning data obtained directly from NGDC. While project resources do not allow extensive user support for the implementation and use of this database or software systems, NGDC staff are always willing to share their knowledge. Where more extensive services are needed, NGDC staff may be able to refer your inquiry to an appropriate source. Please use the following address for contact regarding this project and it's products:

Global Ecosystems Database Project
National Geophysical Data Center
325 Broadway E/GC1
Boulder, Colorado 80303
Phone: (303) 497-6125
Fax: (303) 497-6513
EMAIL: info@mail.ngdc.noaa.gov

Acknowledgements

Many of the data-sets provided here, as well as the general concept and structure of the project and database, were developed and tested through collaboration between the ICSU Panel on World Data Centres and the National Geophysical Data Center as part of a Pilot Project of the International Geosphere-Biosphere Program (IGBP). That project, known as the Global Change Database (Diskette) Project, provided an experimental database for Africa, integrated with Clark University's IDRISI GIS software. The current effort owes much to this prior development, as reflected in the list of contributors.

The encouragement and support of Daniel Marks and William Campbell, as well as subsequent funding from the USEPA, Environmental Research Laboratory in Corvallis, Oregon, in addition to the support from NOAA/NGDC, was critical in the establishment of this project. Other funds were provided in the early stages of development by NOAA/NESDIS, the ICSU Panel on World Data Centres, the U.S. Interagency Working Group on Data Management for Global Change, and UNEP/GRID.

The collaborative support and encouragement of Ron Eastman and the IDRISI Project of Clark University has been instrumental in making this project feasible within the necessary time-frame and funding limitations.

Much of the current database effort is the outgrowth of planning and proposals initially encouraged by David Clark and David Hastings of NGDC, and Stanley Ruttenberg of the ICSU Panel on World Data Centres. The first draft of the project proposal was prepared in Kenya, during a joint assignment between NOAA/NGDC and UNEP/GRID, with the assistance of Michael Gwynne, Harvey Croze, and Jim Weber, and the special encouragement and assistance of Michael Norton-Griffiths.

Over 100 reviewers received copies of the prototype of this database in 1991 for the peer-review. Their continuing evaluations, as well as the responses received so far, are a great benefit to the improvement and expansion of this database.

List of Project Contributors

PROGRAM MANAGEMENT

National Oceanic and Atmospheric Administration

Michael A. Chinnery, Director
National Geophysical Data Center
National Environmental Satellite, Data, and Information Service
325 Broadway, E/GC
Boulder, Colorado

Herbert Meyers, Chief
Solid Earth Geophysics Division
National Geophysical Data Center and World Data Center - A
325 Broadway, E/GC1
Boulder, Colorado

David A. Hastings, Program Manager
Global Change Database Program
Solid Earth Geophysics Division
National Geophysical Data Center
325 Broadway, E/GC1
Boulder, Colorado

Environmental Protection Agency

Dr. Peter A. Beedlow, Chief
Terrestrial Branch
U.S. EPA Environmental Research Laboratory
200 S.W. 35th St.
Corvallis, Oregon

David T. Tingey, Program Leader
Global Processes and Effects Program
Global Climate Research Program
U.S. EPA Environmental Research Laboratory
200 S.W. 35th St.
Corvallis, Oregon

Robert K. Dixon, Program Leader
Global Mitigation and Adaptation Program
Global Climate Research Program
U.S. EPA Environmental Research Laboratory
200 S.W. 35th St.
Corvallis, Oregon

PROJECT LEADERSHIP

Principal Investigator - NOAA/NGDC

John J. Kineman
Ecosystems and Global Change
NOAA National Geophysical Data Center
and World Data Center - A
325 Broadway, E/GC1
Boulder, Colorado

Project Coordinators - EPA/ERL-C

Jeffrey Kern, and formerly William G. Campbell
Global Climate Research Program
ManTech Environmental Technology, Inc.
c/o U.S. EPA Environmental Research Laboratory
200 S.W. 35th St.
Corvallis, Oregon

NATIONAL GEOPHYSICAL DATA CENTER and WORLD DATA CENTER - A
NOAA NATIONAL ENVIRONMENTAL SATELLITE AND DATA INFORMATION SERVICE

Data Management

David A. Hastings
Remote Sensing and Data Integration

John J. Kineman
Ecosystems and Global Change

David C. Schoolcraft
Documentation and Geo-reference Data

Mark A. Ohrenschall
Technical Coordinator

Project Team

Susan Boyle
Michael Callaghan
Jeff Colby
Silvana Delamana
Liping Di
Michael Doney
Kevin Gomolski
Doug Green
David Hastings
Steve Hochberg
Warren Holtquist
Joy Ikelman
Greg Johnson
John Kineman
Leah Eicher Lewis
Andrew Locher
Andra Mealey
Lane Middleton
David Mellon
Laura Nigro
Mark Ohrenschall
John Panskowitz
Stuart Racey
Brendon Roake
James Ross
Lee Row
Joe Salazar
Joel Schacter
David Schoolcraft
Paul Weschler

PANEL ON WORLD DATA CENTRES
INTERNATIONAL COUNCIL OF SCIENTIFIC UNIONS

Stanley Ruttenberg, Director

IDRISI PROJECT
GRADUATE SCHOOL OF GEOGRAPHY
CLARK UNIVERSITY

J. Ronald Eastman, Principal Investigator

Lee Thompson, Project Manager

OTHER INSTITUTIONAL COLLABORATORS

Canada Center for Remote Sensing
Centre Suivi Ecologique (Senegal)
Centro Argentino de Estudios de Radio-comunicaciones y Compatibilidad Electromagnetica, Consejo Nacional de Investigaciones Cientificas y Tecnicas
CSIRO Division of Forestry (Australia)
Environmental Protection Council of Ghana
Environmental Systems Research Institute (USA)
Fondacion Tutela (Venezuela)
IBM Corporation (USA)
IGBP Data and Information System
Interagency Working Group on Data Management for Global Change (USA)
International Maize and Wheat Improvement Center (Mexico)
International Council of Scientific Unions
National Academy of Sciences (USA)
National Center for Atmospheric Research (USA)
National Center for Geographic Information and Analysis (USA)
National Climatic Data Center (USA)
Satellite Data Services Division
Sociedad de Especialistas Latinoamericanos en Percepcion Romota
United Nations Environment Program
Global Environment Monitoring System
Global Resource Information Database
United Nations Institute for Training and Research
University Corporation for Atmospheric Research (USA)
US Army Corps of Engineers, Construction Engineering Research Laboratory
USGS EROS Data Center

INDIVIDUAL CONTRIBUTORS

Steven Acker
Johannes Akiwumi
Nassrine Azimi
David Beddoe
David Bickmore
Graham Bland
Donald Block
Trevor Booth
Ann Bottman
Francis Bretherton
Anne Burrill
Julie Church
Josef Cihlar
David Clark
Ellen Cooter
J.T. Coppock
John Corbett
Wolfgang Cramer
Harvey Croze
Robert Cushman
Frank Davis
Richard Dikau
Rusty Dodson
Phillip Dougherty
Thomas Downing
Ronald Eastman
Jack Eddy
Jack Estes
Kian Fadaie
Alex Faizoun
Kevin Gallo
D.A. Gagliardini
Dale Gillette
Michael Goodchild
William Goran
Robin Graham
Alex Guenther
Michael Gwynne
John Hancock
Jeremy Harrison
Michio Hashizume
Ann Henderson-Sellers
Robert Jackson
Roy Jenne
H. Harszenbaum
David Kicklighter
Felix Kogan
Caspar Kundert
Richard Lamprey
Rik Leemans
David Lindbergh
Margaret Livingston
Robert Lozar
Wayne Luscombe
Elaine Matthews
Duane Marble
Kooiti Masuda
Pamela Matson
Michael McCullagh
Mauro Mendoza
Laura Meszaros
Mark Morrissey
Helen Mounsey
M. Norton-Griffiths
Jerry Olson
Jonathan Overpeck
Anthony Palmer
M. Sanchez Peña
B. Pfeiffer
Fred Pospeschil
David Portman
Christopher Potter
Colin Prentice
Yves Prevost
Jeff Price
S.I. Rasool
Rudolf Reitenbach
John Roads
Martin Rusak
Stanley Ruttenberg
Gilbert Saint
John Schaake
Rejean Simard
Timothy Smith
Allen Solomon
John Taylor
Peter Thacher
John Townshend
Bill Turner
Sushel Unninayar
Vincent Van Engelen
Andrey Velichko
Kevin Weinrich
Robin Webb
Jim Weber
Michael Wilson
John Zerbe

DATA-SET INVESTIGATORS, INSTITUTIONS, AND ANALYSTS

See Data-set Documentation Chapters

EXECUTIVE SUMMARY

The Global Ecosystems Database Project (GEDP) is an Interagency project between the National Geophysical Data Center (NGDC) of the U.S. National Oceanic and Atmospheric Administration (NOAA), and the Environmental Research Laboratory-Corvallis (ERL-C) of the U.S. Environmental Protection Agency (EPA). It is a part of NGDC's "Global Change Database Program" (GCDP), which has the goal of providing modern, global and continental scale data (for the entire Earth's surface) needed by the global change research community. This program is also complementary to the NGDC Paleoclimatology database program, which is focused on building a global array of paleoenvironmental data, another major component of NGDC's global change efforts.

The primary objective of cooperative research and development within the Global Ecosystems Database Project is to produce an integrated global database (including time sequences and model outputs) and quality assurance for spatially distributed characterization and modeling support in response to the combined needs of EPA/ERL-C and NOAA/NGDC. The project, at least in its initial stages, concentrates on modern observational data, including remotely sensed data and data from other origins. The full title of the project is "Co-developing data, tools, and methods for the Characterization and Analysis of Environmental System Patterns to support EPA Global Climate Change Research and Modeling."

A prototype of this database was distributed in 1991 on CD-ROM to selected reviewers for quality evaluation, scientific review, and experimentation. The feedback received from this review was used to design the current version of the database. Subsequent review, to take place each year of the project, will be used to guide annual revisions and additions.

The database includes complementary multi-thematic data-sets on compatible grids, registered to a common origin and projection (latitude-longitude). The database has been structured to be operable with several existing Geographic Information Systems (GIS), so that a complete analytical package can be available to reviewers and other scientists for evaluation, experimentation, and further development. The software accompanying this CD-ROM (a subset of the GIS known as IDRISI) was developed and adapted for this project at Clark University. Although compatible with IDRISI, the database is also designed to be easily up-loaded to the GIS known as GRASS, running on UNIX operating systems. The database structure has been kept as system-independent as possible. Data have been ported relatively easily to several GIS and other systems. This global database has been made publicly available by NGDC and it's associated World Data Center-A (WDC-A); however, individual ("source") data-sets, if desired, may also be obtained from their respective distribution points, as indicated in the accompanying *Documentation Manual.*

The GEDP began in 1991 as a 5-year effort to build a reviewed database with an integrated approach to analysis. The project is making advances in integrating database structure and function, and defining how this integrated approach connects with characterization and modeling. Results of this work include the distribution of CD-ROMs containing successive improvements of the database and supporting materials.

The goal of the GEDP is to develop database and methodological support for global environmental and ecological characterization within the U.S. Global Change Program. Ecosystem characterization has been defined since the early 1970's (by the Coastal Ecosystems Project of the U.S. Fish and Wildlife Service) as:

> "a study to obtain and synthesize available environmental data and to provide an analysis of the functional relationships between the different components of an ecosystem and the dynamics of that system...it is simply a structured approach to combining information from physical, chemical, biological, and socioeconomic sciences into an understandable description of an ecosystem." (Watson, 1978 and others - see references in Appendix A)

Such efforts are needed to support global change research because it is impossible to develop accurate prediction without an accurate description and understanding of present conditions. Furthermore, developing models of long-term processes and system response requires an observational testing ground as well as modern analogs to link both past and predicted conditions. The complexity of global change research thus implies not only different kinds of information on many scales (both temporal and spatial), it also implies the combination of various complementary research efforts, including theoretical vs. descriptive, long-term vs. comparative, and basic vs. goal-directed.

Characterization is descriptive, comparative, and goal-directed. It is organized by multi-disciplinary conceptual models of the ecosystem, which provide the scientific context, experimental designs, and information priorities needed to focus the effort. To apply this concept to the global change scientific program, which is also a goal-directed, multi-disciplinary ecosystem study, characterization efforts must be developed as a complementary link to process-oriented research and modeling efforts (i.e., our present theoretical understanding of the Earth system), such as the global change research motivating this project.

In the context of modern databases and GIS technology, we may think of computer database designed for *adaptive characterization*, which can be defined as:

> An analytical process of integrating and re-designing existing multi-disciplinary data for the purpose of representing important spatial patterns and temporal trends, along with capabilities for data intercomparison, empirical testing, data synthesis, and quantitative assessments of confidence based on experimental design information (i.e., meta-data).

The GEDP represents a modern approach to characterization that relies on digital information (both remotely and directly observed) and computer analysis technology. Specifically, the effort has begun by building a characterization database employing and expanding GIS methods. It is presently linked with on-going modeling research at the EPA's Environmental Research Laboratory in Corvallis, Oregon, and links are being formed with research in the NOAA Climate and Global Change Program. This research linkage determines the short and long-term priorities for database development, as well as the functional requirements for analytical tools. An extensive peer-review effort ensures scientific quality and relevance to broadly based research efforts.

PROJECT DESCRIPTION

OBJECTIVES

Although data for global change research are becoming increasingly available to researchers on a disciplinary and institutional basis, carefully managed data integration and synthesis efforts are required to support specific project and programmatic missions within EPA, NOAA, and other agencies. The primary focus of this cooperative research project is to develop an integrated, global database (including time sequences and model outputs) and quality assurance for spatially distributed characterization and modeling support, related to global environmental and ecological change. This is an in-depth integration and quality assurance effort that is driven by specific research priorities for global change, as defined by needs assessments within supporting agencies and research programs. As such it does not attempt to cover the full breadth of global data or replace larger efforts to establish data distribution systems and exchange formats, but rather to be a participant in those activities. The project will develop methods that are applicable to other research-driven integration efforts. Related objectives include:

1) System independent integration of the database with existing GIS software to support peer review and overall GIS compatibility; with enhancements to readily available software to provide needed functionality of the database.
2) Coordination with other national and international data gathering and evaluation activities that may lead to improvements in the database or its usefulness.
3) Delivery of ERL-C data and models to NGDC for pre-release distribution to reviewers and subsequent integration into the database for general distribution.
4) Dissemination of scientific reviews, enhanced documentation, and quality assessments of the database.
5) Quality assurance of all data processing and documentation work.

GENERAL REQUIREMENTS

- The project will be managed as a 5-year effort on a calendar-year work cycle. NGDC, in collaboration with ERL-C will provide project descriptions including a Project Development Plan and Technical Work Plans for each budget cycle. Periodic status reports will be made, including information on data availability, parallel activities, and the review process.

- Needs assessments will be conducted by EPA/ERL-C in order to re-evaluate and prioritize data requirements according to research needs within the sponsoring organizations. These priorities will determine the scientific focus of data integration work and the specific set of variables in the characterization database.

- Additions and revisions to the integrated database and documentation manuals will be released following each work cycle. Each update will improve the database through: 1) inclusion of new data-sets (including ERL-C derived data), 2) better accuracy and/or temporal and spatial resolution, and 3) better quality information and/or documentation, including the results of scientific reviews.

- NGDC, in collaboration with ERL-C, will establish agreements for peer review of the database repeated each year of the project. This will include a multi-disciplinary (extensive) review of the annually-released database products (i.e., CD-ROMs and documentation), and disciplinary (intensive) reviews of individual data-sets to be added to the database during each annual cycle.

- NGDC, through the Inter-Agency Agreement, will be responsible for data distribution to reviewers, in cooperation with ERL-C. NGDC will be independently responsible for public distribution of reviewed data products through its normal channels and Data Center mission; however, such products will appropriately represent institutional cooperation and support.

- CD-ROMs will be the primary distribution media, and data structures will be compatible with the latest advances in GIS technology and characterization methods. As part of this requirement, NGDC will ensure that the database is fully operable in a GIS environment and portable between systems.

- NGDC will implement and maintain quality control procedures to ensure state-of-the-art data processing, data management, and product development.

- NGDC will identify parallel activities in global databases within national and international research organizations and data centers, assess overlaps with the project, identify efforts that may be leveraged with project activities, and pursue collaboration in support of the project.

TASK DEFINITIONS

The Global Ecosystems Project can be divided into three task levels. Each of these levels is further divided into Tasks that range from assessing data needs and availability to applications in global change research. To date, Tasks 1-5 have been established and some projects have been started under Task 6. Research related to Task 7, 8, and 9 has been initiated and will be expanded in 1992. A more complete description of each task is provided in Appendix A.

DATABASE INTEGRATION AND DESIGN
Task 1. Data Availability and Needs Assessment
Task 2. Data Acquisition, Integration, and Archive
Task 3. Database Structure and Function (GIS Implementation)

TECHNICAL DEVELOPMENT
Task 4. Quality Assessment and Documentation
Task 5. Distribution and Peer Review
Task 6. Development and Improvement

SCIENTIFIC SUPPORT AND METHODOLOGY
Task 7. Characterization Methodology
Task 8. Scale Integration and Error Tracking
Task 9. Links to Modeling and other Applications

While a number of data-systems planning efforts are underway within the U.S. Global Change Program, this database effort has been driven by existing data availability and more immediate research needs. The most fundamental level of scientific support provided by NGDC is thus the provision of an operational, commonly usable, state-of-the-art database. That job involves data and software integration, quality assessment, data development, and review activities that ensure scientific credibility. The link with actual research, however, must also be carefully established to ensure "proof-of-concept," and overall relevance of this effort to global change research.

DATABASE INTEGRATION AND DESIGN includes both data integration and integration with analytical methodology. This level is driven by data availability and external developments in GIS software. Requirements are expressed in terms of database operability and software functionality, using the structure and conventions that are rapidly developing in the GIS concept. Internal quality assurance efforts are extremely important at this level, to ensure minimal loss between original investigations producing data and the integrated package that is ultimately distributed. Although there is much new ground to break, many decades of unique data management and processing experience with physical environmental data provide considerable quality assurance in addressing the new disciplinary requirements of the global change program, combining traditional methods with new.

TECHNICAL DEVELOPMENT includes all efforts to improve the integrated data, through in-house projects, specific contracts, and collaboration with other researchers. This level is driven by feedback from the external peer-review, and may involve internal data improvement projects as well as external contracts or collaborative agreements. Collaboration with ICSU World Data Center projects and other international programs for scientific exchange and data development on a local and regional basis is part of this effort. Certainly, however, the peer-review process established in 1990 has become the major element for quality assurance and improvement of the database.

SCIENTIFIC SUPPORT AND METHODOLOGY covers various scientific issues regarding the design and application of a large database for global change research. This level is driven by the methodological requirements of modeling and characterization research, and involves defining and implementing a characterization database for global change studies. The concept of characterization implies not only a scientific database, but also the correct relationship between the database and conceptual models, for example its relationship with dynamic modeling, including process models and global simulations. Another important issue is the link between studies (i.e., characterization and modeling) at different scales, scaling up to the coarse grids of General Circulation Models (GCM) from site-specific studies and scaling down to site-level prediction from global models. This presents the complementary problems of combining data scales within a useful range, and linking predictive information from whole studies at widely different scale ranges.

5-YEAR DEVELOPMENT PLAN

----TASK----			----SCHEDULE----			
AVAILABILITY AND NEEDS	**90**	**91**	**92**	**93**	**94**	**95**
• data availability inventory (DBMS)		R&D	X>>>>>>>>>>>>>>>>>>>>>>>			
• needs assessment		X	X	X	X	X
• collate reviews, availability, and needs			X	X	X	X
• determine research driven priorities			X	X	X	X
ACQUISITION, INTEGRATION, AND ARCHIVE	**90**	**91**	**92**	**93**	**94**	**95**
• ingest and archive	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• data integration	R&D	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>				
• documentation search	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• publication/distribution clearance	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
STRUCTURE AND FUNCTION (GIS)						
Projection	**90**	**91**	**92**	**93**	**94**	**95**
• lat/lon	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• multiple projections			R&D	X>>>>>>>>>>>>>>>>>		
GIS structure	**90**	**91**	**92**	**93**	**94**	**95**
• raster	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• vector overlay	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• multiple grid convention		R&D	X>>>>>>>>>>>>>>>>>>>>>>>			
• vector/tabular (attribute) data			R&D	X>>>>>>>>>>>>>>>>>		
Software Support	**90**	**91**	**92**	**93**	**94**	**95**
• DOS/IDRISI support	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• Unix/GRASS support				X>>>>>>>>>>>>>>>>>		
• access software		X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>				
• on-line documentation retrieval			X>>>>>>>>>>>>>>>>>>>>>>>			
- improved browse and display interface			X>>>>>>>>>>>>>>>>>>>>>>>			
• multiple export formats			R&D	X>>>>>>>>>>>>>>>>>		
- re-projection capability	R&D	R&D	R&D	?>>>>>>>>>>>>>>>>>		
- improved gridding and contouring	R&D	R&D	R&D	?>>>>>>>>>>>>>>>>>		
- lineage tracking interface				R&D	?>>>>>>>>>>>	
- error tracking and analysis				R&D	?>>>>>>>>>>>	
- collaboration with GIS developers			X>>>>>>>>>>>>>>>>>>>>>>>			
QUALITY ASSESSMENT AND DOCUMENTATION	**90**	**91**	**92**	**93**	**94**	**95**
• processing and lineage documentation	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• version control				X>>>>>>>>>>>>>>>>>		
• data intercomparison		X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>				
• exploratory statistical analysis		X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>				
• data-model intercomparison				R&D	?>>>>>>>>>>>	
DISTRIBUTION AND PEER REVIEW	**90**	**91**	**92**	**93**	**94**	**95**
• prototype CD-ROM		X				
• Pre-Release Review: data-set/quality review		X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>				
• Public Release CD-ROM (DOS operable)		X	X	X	X X	
• Public Release CD-ROM (UNIX operable)				X	X	X X
• Primary Review: Database/scientific review		X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>				
• User's Guide and Documentation Manual updates		X	X	X	X	X X
• digital "on-line" documentation			X>>>>>>>>>>>>>>>>>>>>>>>			
• global data BBS: forum for reviewers			X>>>>>>>>>>>>>>>>>>>>>>>			
DEVELOPMENT AND IMPROVEMENT	**90**	**91**	**92**	**93**	**94**	**95**
Content						
• observational and derived data	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
• synthetic data			R&D	X>>>>>>>>>>>>>>>>>		
• dynamic model outputs and software			R&D	X>>>>>>>>>>>>>>>>>		
- data-set development projects	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					

	Scale	90	91	92	93	94	95
•	global 10min-1deg (16-100km)	X>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
•	global 1-5deg (100-500km)			X>>>>>>>>>>>>>>>>>>>>>>>>			
•	global 2-5min (4-8km)			R&D	?>>>>>>>>>>>>>>>>>>		
•	global ~.5min (~1km)					?>>>>>>>>>>>	
•	continental 2-10min (4-16km)			R&D	X>>>>>>>>>>>>>>>>>>		
•	continental ~.5min (~1km)					?>>>>>>>>>>>	
-	site/regional ~1km				R&D	?>>>>>>>>>>>	
-	site/regional 10m-100m				R&D	?>>>>>>>>>>>	
SCIENTIFIC SUPPORT		90	91	92	93	94	95
•	characterization methods				R&D>>>>>>>>>>>>>>>		
•	scale integration methods				R&D>>>>>>>>>>>>>>>		
-	GIS-modeling linkage				R&D>>>>>>>>>>>>>>>		
-	synthesis methods				R&D>>>>>>>>>>>>>>>		

•	core project activities	R&D	research and development
-	associated activities	X	planned implementation
?	possible implementation	>	continuation

RESEARCH

ERL-C's environmental research program studies and models the climate-biosphere interaction to understand and predict global climate-related change, to assess impacts, and to recommend mitigation strategies. NOAA's environmental research is primarily concerned with modeling climate and ocean systems, including biogenic feedbacks. Project development was driven mostly by the needs, priorities, and anticipated outputs of the Global Climate Research Program at ERL-C, which jointly funds the project; however, anticipation of NOAA needs within the Climate and Global Change Program also figured heavily into the project plans and spawned a host of research projects and technical developments at NGDC, mostly related to developing and integrating the database. Projects related to the GED effort are collectively and individually identified below (reference indicated in parentheses):

RESEARCH AND DATA DEVELOPMENT (EPA/ERL-C)

Mechanistic Continental-Scale Vegetation Models (Ronald Nielson, Oregon State University, Corvallis)
Vegetation Redistribution and Paleoecology (George King, ManTech Env., Corvallis)
Linking Global Vegetation Models with General Circulation Models (Starly Thompson, NCAR-Boulder)
U.S./Global Soil Carbon and Water-holding Capacity (Jeffrey Kern, ManTech Env., Corvallis)
Effects of Climate Change on Habitat (Don Phillips, EPA, Corvallis)
U.S. Potential Evapotranspiration (Danny Marks, ManTech Env., Corvallis)
Columbia River Basin Water Balance (Danny Marks, ManTech Env., Corvallis)
Regional Runoff Modeling (Dennis Lettenmaier, University of Washington, Seattle)
Ecophysiological Processes and Effects (Bruce McVeety, Battelle, Corvallis)
Effects of Elevated CO_2 and Climate Change on forest Trees (Paul Rygiewicz, EPA, Corvallis)

Paddy Rice Ecosystem Modeling and CH_4 (Methane) Production in Asia (Dominique Bachelet, ManTech Env., Corvallis).
Effects of UV-B and Climate Change on the Rice Ecosystem (Ken Fischer, The International Rice Research Institute, Los Baños, Phillippines).
Long-term Climate Variability in SE China (William Chang, University of Michigan).
U.S. Forest Carbon Dynamics (Dave Turner, ManTech Env., Corvallis).
Carbon Cycle Dynamics in the Former Soviet Union (Ted Vinson, Oregon State University and Tatyana Kolchugina, Moscow State University, Moscow).
Global Assessment for Managed Terrestrial Ecosystems to Withdraw CO_2 from the Atmosphere (Richard A. Houghton, Woods Hole Research Center)
Current and Future Global Terrestrial Carbon Pool Distribution (Jerry Melillo, Marine Biological Laboratory, Woods Hole).
Carbon Dynamics of Slash Pine Plantations (Wendall P. Cropper, Jr., University of Florida, Gainesville).
Vegetation Response to Global Climate Change (Allen Solomon, Michigan Tech. University, Houghton).
Implications of Loss of Biological and Genetic Diversity in a Regulatory Context (Robert McKelvey, University of Montana, Missoula).
Land Suitability and Availability for Carbon Sequestration: Tropical Asia as a Case Study (Sandra Brown, University of Illinois, Urbana).
Forest Management Systems to Conserve/Sequester Carbon (Jack Winjum, NCASI).
Global Carbon Dynamics in Response to Climate Change (Jeffrey Lee, EPA, Corvallis).
Effects of Elevated CO2 and Climate Change on Forest Soil Processes (Mark Johnson, ManTech Env., Corvallis).
Rice Responses to UV-B and Climate Change (Dave Olszyk, EPA, Corvallis).
Tree Physiology Responses to Climate Change (Dave Olszyk, EPA, Corvallis).

RESEARCH AND DATA DEVELOPMENT (NOAA/NGDC)

World Ecosystems Classification and Digital Mapping (Jerry Olson and Lee Stanley; NGDC-Boulder).
Classification and Digital Mapping of Global Phytomass (Dmitry Varlyguin and Ron Eastman, Clark University-Worchester).
Improved Corrections and Re-calculation of the NOAA Global Vegetation Index (Samuel Goward, Dennis Dye; Univ. Maryland).
Extended Coverage of NGDC Monthly Generalized Global Vegetation Index Database (1988-present), and Exploration of Empirical Corrections (John Kineman and David Hastings).
Improved Global and Continental Scale Terrain Data (Lee Row and George Sharman; NGDC-Boulder).
Improved Polygonal and Linear Boundary Data from World Vector Shoreline, Relational WDB-III, and Digital Chart of the World (David Schoolcraft, Mark Ohrenschall).
Dust Modeling and Desertification Studies in the Southwestern U.S. (Dale Gillette and Tes Johnson, NOAA/Air Resources Lab-Boulder; David MacKinnon, US Geological Survey-Flagstaff; John Kineman, NGDC-Boulder).
Scale Analysis of Land Use Data and the Simulation of Hydrologic Response (Jeff Colby, NGDC-Boulder).
Adaptive Ecological Characterization for Global Change (John Kineman, NGDC-Boulder;

Jack Estes, UCSB-Santa Barbara).

AVHRR (Advanced Very High Resolution Radiometer) Stereography (David Hastings, NGDC-Boulder; Bill Emery, Univ. of Colorado-Boulder).

Global 1km AVHRR Mosaic from NDVI Local Area Coverage (LAC) (Martin Rusak, Nevin Bryant, Jim Knighton, JPL-Pasadena; David Hastings, NGDC-Boulder; Joanne Heckmann, NGS-New York; Peter Muller, UCL-London).

NASA Topographic Science Definition Team (Miriam Baltuck, NASA-Washington; Sergio Vetrella, Italian Space Agency; David Hastings, NGDC-Boulder).

Modern Analog (Calibration) Data for Paleoclimatology (Robin Webb, Jonathan Overpeck, Dave Anderson; NGDC-Boulder).

Conversion of the UNEP/GRID Africa Database from Miller Projection (Kenneth Strezpek, CU/CADSWES-Boulder; John Kineman, NGDC-Boulder).

SOFTWARE AND TECHNIQUE DEVELOPMENT (NOAA/NGDC)

Data Integration Methods and Technology (Mark Ohrenschall, John Kineman, David Hastings).

GIS Structure and Function for Global Studies - IDRISI (Ron Eastman, Clark University-Worchester; John Kineman, NGDC-Boulder).

GIS-Database Operability -- GRASS (Bob Lozar and Bill Goran, CERL-Champaign; David Hastings, NGDC-Boulder).

Data Browse and Visualization (Allen Hittleman and Ray Haberman, NGDC-Boulder).

Optimizing Integrated Earth Systems Data (David Hastings and Liping Di, NGDC-Boulder).

QUALITY ASSESSMENT AND DISSEMINATION (NOAA/NGDC)

Data Intercomparison and Descriptive Modeling (David Hastings, Liping Di, John Kineman; NGDC-Boulder)

Intercomparison of NDVI Data-Sets (David Hastings, John Kineman, Jeff Colby, NGDC-Boulder; Ron Eastman, Clark Univ.-Worchester; Josef Cihlar, CCRS-Ottawa; John Townshend and Chris Justice, Univ. of Maryland)

UNITAR/IGBP Dakar Data Workshop - September 1990 (John Kineman, NGDC-Boulder; Alex Faizoun, LERTS, Toulouse)

SELPER/IGBP Buenos Aires Data Workshop - April 1991 (David Hastings, NGDC-Boulder)

NOAA Educational Support and Development (David Hastings, David Schoolcraft, John Kineman, NGDC-Boulder; Douglas Brown and John Schmidt, NOAA Educational Affairs Division-Suitland)

DATA ACQUISITION AND DEVELOPMENT

SCOPE

The focus of this project is on characterizing environmental and ecological variables at suitable continental and global resolutions, and on time scales from years to decades. Initially the choice of data-sets was determined by availability at scales between approximately ten minutes and one degree (longitude and latitude). This range was also chosen because it identifies a scale that is intermediate between the more detailed scales required for site-specific process studies, and the coarser scales to which GCMs are presently restricted, thus filling a gap and providing a means for scaling up or down between these extremes. Scale integration, including the relationship between data and application of predictions at different scales, has great scientific value by allowing the intercomparison of predictions at different scales and studies of different kinds of phenomena.

RESOURCE MANAGEMENT & SITE STUDIES		GLOBAL CHARACTERIZATION		GLOBAL SIMULATION MODELING	
10-100m	1-4km	4-16km	50-100km	100km	500km
Spot, Landsat	AVHRR LAC/GAC	AVHRR/ GVI	Climate data, classifications, etc.	Climate data, Classifications, etc.	GCM runs

Figure 1

Establishing a reasonable hierarchy of spatial and temporal scales is important because (1) direct comparison of data at different scales that are not too widely separated can be useful for quality assessment and developing classification hierarchies; (2) patterns represented at one scale may not be visible at another; and (3) the study of phenomena and patterns at resolutions of interest must be matched with appropriate sampling (in both time and space). Figure 1 shows where the global characterization effort falls within a very general division of scales.

The project aims to test various ideas about data at this intermediate range of scales, such as:

(1) that some interesting processes and phenomena may in fact scale within this range,

(2) that this intermediate scale database can serve as a link between the two scale extremes, perhaps providing a means to extrapolate process information to global distributions, and/or apply generalized predictions to more localized phenomena,

(3) that the database can provide inputs to coarser-scale models,

(4) that the database can be improved from interaction with local facilities and databases,

(5) that by beginning global analysis with currently available data, we will help to improve analytical methods that may later be applied to, or assist in the development of, more comprehensive data-sets.

The general scope of the data search is expected to widen as work progresses, to include:

(1) Global environmental and ecological data-sets of approximately 5-10 minute (1:10,000,000) scale, up to 1 degree (1:100,000,000), with near-continuous spatial distribution, according to availability. (1990-)
(2) Coarse-scale GCM outputs, from 1 degree (1:100,000,000) to 5 degrees (1:500,000,000). (1992-)
(3) Global statistical (socio-economic) data by country and sub-country units. (1992-)
(4) Dynamic process models (including compiled and/or source code), and associated data-sets. (1992-)
(5) Continental or regional data-sets from 1 km (1:1,000,000) to 5-10-minute (1:10,000,000) scale (planned). (1992-)
(6) Site specific data-sets for case studies. (1993-)

INITIAL DATA NEEDS

Specific data needs were identified by EPA/ERL-C in 1991 in response to their Global Climate Research Program activities (listed above). These were:

- Daily: Potential evapotranspiration, precipitation, solar radiation, snowpack, minimum and maximum temperature.
- Monthly: Precipitation, air and surface temperatures (min, max, avg.), potential evapotranspiration, relative humidity, solar radiation, growing degree days, leaf biomass, day length, wind, vegetation indices from AVHRR.
- Long-term: Soil texture, soil organic matter, growing degree days, coefficient of annual moisture availability, vegetation type, soil classes, soil depth, soil rock fragments by layer, soil organic carbon, soil thickness, soil water holding capacity, topography, surface roughness, erosion factors, leaf area index, forest type, stand age, forest and soil biomass, net ecosystem productivity estimates, forest fire extent and frequency, land use, management practices.
- Historic: 100 years climate data, 1000 years climate proxy records and sediment fossil climate indicators.
- Predicted: Climate change temperature and precipitation scenarios.

These requirements of the ERL-C program were sufficient to organize the initial data search and acquisition effort, especially considering the poor availability of most of the required measures. It was clear, and remains so, that most of the required data or proxies will have to be generated or synthesized from existing data and models, and that the first goal of the database effort should then be to assemble a basic set of immediately available data into a form that lends itself to the necessary characterization and modeling tasks that will follow. That job in itself is not trivial, and is expected to continue for the life of the project, concurrently with specific derivations.

DATA PRIORITIES FOR ERL-C RESEARCH

At the end of 1991 another needs analysis was conducted at ERL-C in response to NGDC's data availability inventory. This resulted in more specific priorities for data in the coming years, as follow (list provided by ERL-C):

First Priority:

- Improved versions of the Global Vegetation Index (GVI) - i.e., 16km. NDVI
- Improved global soils and particle size data
- Improved global elevation
- ISSCP Cloud Distribution
- Major Deltas of the World
- Continental (4-8 km) NDVI, i.e., Global Area Coverage (GAC)
- Global ozone concentration

Second Priority:

- Global wind velocity at .5-degree
- Rivers of the world
- Worldwide population density
- Location of major cities of the world
- FAO Natural Vegetation of SE Asia
- Country level digital soil maps

DATA PRIORITIES FOR NGDC RESEARCH AND DATABASE DEVELOPMENT

- Multiple versions of 16 km GVI data-sets
- Global 8 km Normalized Difference Vegetation Index (NDVI) time series
- Regional 1 and 4 km NDVI for selected arid land sites
- Improved 5 and 10 minute (or finer) global elevation and bathymetry
- Improved Olson World Ecosystem classes
- Global biomass estimates
- Global Socio-economic database
- Coarse scale GCM outputs
- Improved global gridded climatology
- Global heat budget from satellite
- Global aerosols from satellite
- Geographic (unprojected) Africa database (in conjunction with UNEP/GRID)
- Geographic (unprojected) South America database (in conjunction with CIMMYT)
- Geographic (unprojected) Asia database

DISTRIBUTION OF ERL-C DATA AND MODELS

In addition to the data NGDC will acquire or develop in support of this project, ERL-C will be producing data for distribution through the Global Ecosystems Database. These contributions will consist of both data and models developed in the ERL-C research program, listed by ERL-C as follows:

Spatial Data:

- Monthly time series of potential evapotranspiration (PET), air temperature, vapor pressure, and wind for the contiguous USA (Marks)
- Daily time series PET, actual evapotranspiration (AET), air temperature, vapor pressure, wind, and precipitation for the Columbia basin (Marks)
- Global Eagleman's PET, AET, relative humidity, and soil water deficit (Neilson)
- Ultraviolet light levels for Asia (Bachelet)

- Agroclimatic Data for Asia (Bachelet)
- Regional/Global Soil organic carbon content (Kern)
- Regional/Global Soil texture, rock fragment content, soil depth, and extent of organic soils (Kern)
- Regional/Global Soil water holding capacity and hydraulic conductivity (Kern)

Tabular (site) Data:

- Carbon pools and fluxes for the USA (Turner)
- Global database of forest practices (Winjum)

Models/Software:

- Models for generating PET, elevation corrected temperature, humidity, and wind; and the Image Processing Workbench Software (Marks)

QUALITY ASSURANCE

Quality Assurance (QA) is defined as a program for ensuring that data derived from environmentally-related measurements do not exceed acceptable limits within a stated level of confidence. QA encompasses the plans, specifications, and policies affecting the collection, processing, and reporting of data. Correspondingly, Quality Control (QC) is defined as the activities that are conducted to carry out QA, i.e., to reduce error and ensure that data are generated and/or processed within known and acceptable limits. Quality assessment, on the other hand, is an important part of the Quality Assurance plan and is defined as the testing of data-sets to determine their quality for documentation purposes. Since the project integrates existing data-sets as well as new ones, it may not have control over original production in many cases. Quality Assessment, therefore, becomes a retrospective activity that results in new documentation of problems and artifacts (published in the Documentation Manuals).

Quality Assurance objectives are discussed here and in the related Peer Review section that follows. It is also expanded upon in Appendix A, Task 4, 5 and 6. Quality Assurance guidelines for the project are also described in the Project Development Plan.

Quality Assurance, addressing issues of traceability and uncertainty in large spatial data-sets, is a primary goal in building a commonly accessible characterization database. It is complimentary to the goal of scientific support, and is the motivation for much of the cooperative activity associated with the project, and for adopting an approach that employs community-wide participation and methodological conventions, as emerging in GIS and related technology for handling large databases. This project addresses Quality Assurance on two levels, first in regard to specific Quality Control procedures at NGDC, and second in regard to uncertainty in the data-sets themselves and their usefulness to the global data community as a whole.

The first level -- Quality Control procedures at NGDC -- involves internal processing, testing, and documentation procedures, as well as the pre-release review established under the project. These operating procedures are described in various documents within the Data Center, and summarized in the Project Development Plan and Technical Work Plans (not repeated here).

The second level, i.e. the issue of uncertainty in the data-sets themselves and their usefulness for global change research, requires a much broader approach. This is done through efforts to improve the technical and scientific verifiability of the data-sets individually and as an integrated whole. Improved documentation and meta-data contribute to this effort, as do internal quality checks performed during processing. The greatest contribution, however, may be the integration of these data into a completely functional package, with common structures and useful tools for intercomparison and analysis. This step provides a common methodology for verification and improvement of the data. By also establishing a widely-based peer-review effort, it is possible to assess the effectiveness of database design, integration, and documentation efforts, and to incorporate review comments into the documentation itself. Just as global change research must ultimately be a multi-disciplinary community-wide effort, this project approaches the issue of quality by facilitating the widest possible participation in coordinated use and review.

Public distribution and feedback is also considered part of the review process, although without formal control and with less direct feedback. By providing common ground for feedback and exchange within the global change community at large, it is possible to stimulate data improvements, knowledge about new data-sets, reduced duplication and greater compatibility of database efforts, involvement of new disciplines and local expertise from diverse locations, data exchanges, improvements in GIS and related technology, support for educational and policy outreach projects, and general intercomparability of research. Such communication improves quality as data publication and distribution quality standards become established and begin to parallel similar conventions for reviewed literature (i.e., emphasizing verifiability).

EXTERNAL PEER-REVIEW

The peer-review effort was implemented in 1991 and will be maintained throughout the project. A three-stage review process now exists, providing pre-release technical quality review of individual data-sets (targeted to a small number of specialists), multi-disciplinary review of the integrated CD-ROM database (within a network of 100 reviewers), and feedback from unrestricted use within the scientific community. In this way, the CD-ROM products will be in continual review, and continually accessible, with additions of new reviewed data released annually on CD-ROM, and any major structural changes taking place through occasional release of new versions of the entire database. The three stages of distribution and review are:

Stage 1: **Pre-Release Technical Review by Data-set:** Intensive assessment of technical quality of individual data-sets to approximately 3 to 5 disciplinary experts, prior to public distribution.

Purpose: Assess knowledge of technical and statistical nature, sampling design, and derivation method, perform statistical analyses and validation tests, assess technical quality of data, assess accuracy and completeness of documentation, and provide recommendations for use.

Results: Critique of data-set and documentation, suggested or actual improvements, collaboration with parallel data development activities, and improved documentation.

Stage 2: **Primary Technical and Scientific Review of the Database:** Extensive assessment of scientific quality and usefulness within a selective distribution to approximately 100 active researchers, concurrent with public distribution.

Purpose: Evaluate database structure and functionality, determine usefulness for scientific applications, intercompare data-sets and assess their combined value, develop case studies and additional links with global change research, assess overall scientific value, recommend distribution and use.

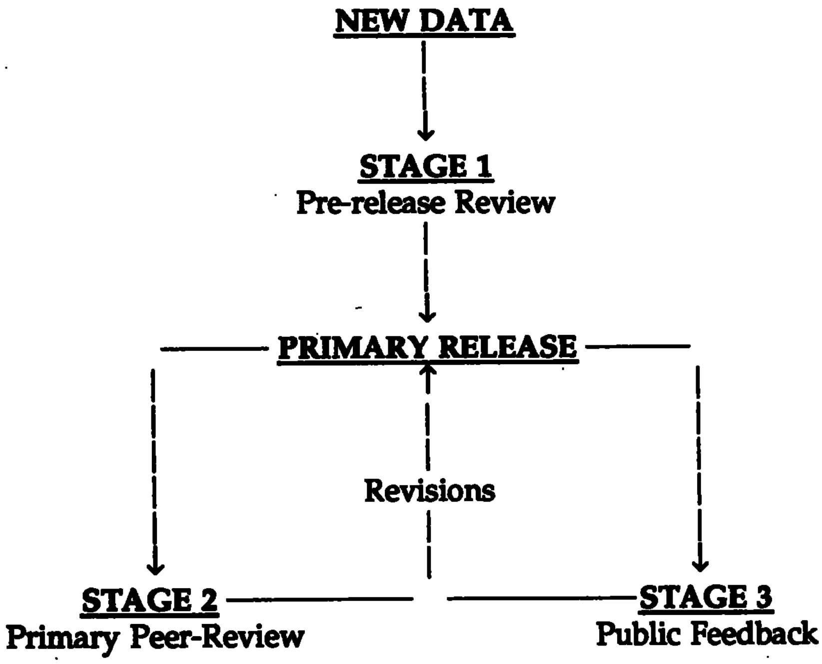
Results: Evaluations of overall database structure and function, feedback on relative value of individual data-sets and their combination, contributions and/or information of new data-sets, links with data centers and other institutions, improved documentation, case study examples of research and characterization efforts, links with and between researchers worldwide, specifications for GIS and other analytical developments.

Stage 3: **Public Distribution and User Support:** Public availability at cost or exchange basis, dissemination to specific groups. Distribution is independently supported by NGDC/WDC-A.

Purpose: Dissemination to the scientific community, common ground for exchanging data and characterization information, methodological development and training, support for developing global change information and assessing usefulness for policy applications and decision-making.

Results: User support and community feedback, basic support for scientific outreach activities, basic support for the development of educational products.

The annual cycle of peer-review and distribution is sketched below:



SUMMARY OF REVIEWS

Out of approximately 100 reviewers of the prototype CD-ROM, there were 32 responses to the questionnaire and 5 letter responses. Representativeness of the 38 responses to the overall community has not been determined, although the original selection of reviewers was targeted to active global change scientists. The reviews were collated for each question and are summarized here. Many of the reviewer's comments and recommendations have been incorporated into the current version. Based largely on this review, some structural and formatting changes were made as well. In all, the review was very supportive of the basic approach and initial effort, and very constructive in criticism. The text of reviewer's actual comments is provided in Appendix B. The following is a very general tabulation of the response to specific questions.

1) OVERALL USEFULNESS OF THE INTEGRATED DATABASE:

Category	Number of responses
0 no answer	5
1 not useful	0
2 maybe/minimal	1
3 useful	10
4 quite/very useful - major step forward	9
5 nearly essential or extremely useful	7

Overall Ranking: (3.8) Very useful, major step forward

1a) SOFTWARE SYSTEMS MOST USEFUL WITH THE DATABASE:

This table counts the number of mentions of a given software package. Multiple categories per reviewer were counted.

Category	Number of responses
no answer	3
unable to access	1
GRASS	4
UNIX	13
IDRISI	18
DOS	22
User developed software	6
NCSA, Datascope, XIMAGE	1
SPANS	2
ARC/INFO	4
Laserscan Horizon	1
BIOGIS	1
DbaseIV, SPSS/PC, PC/SAS, Grapher/Surfer	1
IPW	1
MIPS	1
Text editor	1

Overall preference: 1-DOS/IDRISI, 2-UNIX/various systems

1b) IMPORTANCE OF LINKAGE BETWEEN THE DATABASE AND GIS:

Category		
0	no answer	4
1	not important	4
2	marginal	3
3	important/desirable	7
4	necessary/very important	8
5	essential/extremely/absolutely	7

Overall Ranking: (3.4) Important/Very Important

1c) IMPORTANCE OF "LOWEST-COMMON-DENOMINATOR" (GIS) APPROACH:

Category		
0	no answer	6
1	not important	5
2	marginal value	4
3	valid approach/desirable	5
4	correct/very/necessary approach	12
5	essential	1

Overall Ranking: (3.0) Valid approach/desirable

Notes: General agreement on importance of portability and need for supporting upload and exchange formats as an absolute minimum. Point of disagreement was if Lowest Common Denominator implies less portability and system independence.

2) ACCESSIBILITY OF DATA, STRUCTURES, FORMATS, MEDIA, PROJECTION:

Since some reviews gave a divided opinion between the various items in the question. To represent this evenly, two votes per reviewer were counted.

Category		
0	no answer/unable to evaluate	9
1	no - terrible	1
2	poor/not easy	8
3	fair - workable/useful	7
4	yes - good/easy/fine/no problem	33
5	yes - excellent/very accessible	7

Overall ranking: (3.7) Good, but some problems

Notes: The main criticisms were support for upload/exchange formats, lack of projection capability, need for better hardcopy output support, and need for distribution with full GIS software.

3) ADEQUACY FOR RESEARCH NEEDS.

Category		
0	no answer/unable to evaluate	8
1	no - inadequate	2
2	poor - not entirely/unknown	4
3	fair - some data-sets only	6
4	yes - good/useful	7
5	yes - excellent/very useful	6

Overall ranking: (3.4) Limited, fair to good
Notes: Main issues were scope, resolution, verifiability, and documentation. There was general agreement that while this is a good start, considerable improvement and expansion is needed.

3a) LIMITATIONS OF THE DATABASE.

The major issues, in approximate order of importance, were:

1. disciplinary scope and completeness
2. scale integration
3. resolution
4. quality of data-sets and geographic consistency
5. temporal coverage
6. documentation of derivation methods
7. interpretation
8. multiple platform support

Notes: See specific comments in Appendix B.

3b) PRIORITIES FOR IMPROVEMENT:

The following items, in random order, were mentioned:

System priorities:
- access and multiple platform/systems support
- geographic reference
- subsetting vector files
- ASCII file format
- include hardcopy of images for comparison

Data development priorities:
- documentation - add scope and limitations of each data-set
- error analysis
- accessibility of raw data
- geographic evenness (quality, scale, coverage)
- temporal evenness and longer time sequences
- finer resolution / especially from remote sensing

Additional/improved data-sets and/or derivations

improved coverage of temperature, rainfall and FAO soils
more detailed topography
updated land-use coverage
soil attribute data layers
improved NDVI derivations
improved World Ecosystems
improved (corrected) topography and bathymetry
new soil texture database
agricultural practices
Earth Radiation Budget Experiment data
more current climate data
Sea Surface Temperature
sources and sinks of greenhouse gases
radiative forcing indices
volcanoes, large faults, historic landslides
dams, mines, cities, observatories
population density
GCM climate change scenarios
general geological/geochemical data
air pressure, air quality
drainage connectivity array consistent with terrain data
derived indices of moisture demand; PET, AET, etc. (weekly or daily)
monthly evapotranspiration, moisture indices (Thornthwaite)
Kuchler's potential vegetation for North America
USGS DLG and DEM data for North America
paleoclimate data-sets, 6000 and 18,000 B.P. - e.g., Dept. of Energy lake level changes
paleo-environmental change last 130,000 years B.P.
atmospheric chemistry data (NO_x, SO_2, etc.)
socio-economic data (population, land use, agricultural zones, cities)
analytical routines (aridity index, agroclimatic zones, vegetation biomes)

4) DATA POTENTIALLY AVAILABLE FOR CONTRIBUTION FROM REVIEWERS:

Atmospheric trace gases
Aggregated Holdridge Life Zone Classifications for current and doubled CO_2 scenarios.
BIOME model vegetation-climate classification
Agroclimatic zones, length of growing season, start of growing season, other variables.
Natural emissions of methane, non-methane HC, nitrogen and sulfur
Improved Micro World Data Bank II
Soil Nitrogen mineralization and trace-gas fluxes based on global models
Water budget of the Nile basin (precip., potential evaporation, and runoff)
EPA ecological and derived climate change scenario databases (ARC/INFO format)
Individual monthly raingauge totals from Pacific atols, 1971-1990.
Global distribution of 20 Most Important Agricultural Crops.
Vegetation, NPP, Net nitrogen mineralization.
Improved IIASA Database and agroclimatic indices from BIOME model
Hydrometeorological data-sets

T & P for Europe (.5x1 degree)
GCM scenarios
Paleoenvironmental database

5) OTHER KNOWN/RECOMMENDED SOURCES FOR DATA:

Complete model runs with all fields from GCM runs (NCAR 1XCO_2 and 2XCO_2 runs).
GCM scenarios (GISS, GFDL, CCC, etc.)
NCAR GCM temperature and precipitation runs
CDIAC Trends '90 data?
Improved FAO soil map of the world (in production)
Climatic trends and series (monthly), (see reference in Appendix B, V:#3)
Monthly interpolations for solar radiation, (see references in Appendix B, V:#5)
Agricultural models, e.g., International Benchmark Sites Network for Agrotechnology Transfer, and DSSAT model system (Dept. of Agronomy, Univ. of Hawaii)
World Weather Disc CD-ROM (Weather Disc Associates, 4584 NE 89th, Seattle, WA)
Tropical Forest Depletion - FAO, Rome.
Digital Chart of the World - Defense Mapping Agency
Geophysical hazards - World Data Center
Weekly GVI, e.g., Gutman, Univ. of Maryland
Int. Satellite Cloud Climatology Project - NASA
Soviet Global Land-Use - Univ. of New Hampshire
Digitized Global Vegetation Maps (see references in Appendix B, V:#21)
Numerical weather prediction, 3-D atmospheric structure - NCAR, NMC/CAC
Outgoing long-wave radiation - NCDC
SPOT and LandSat data
CIA World Data Bank II (full resolution)
Sea Surface Temperature time series from AVHRR
UK Met. Office database
Total water vapor (SSM/I) - NASA
Higher resolution NDVI - NOAA/NASA
Sea Ice - NOAA
Coastal Zone Color Scanner - NASA

6) IS THE YEARLY RELEASE SCHEDULE REASONABLE:

Category		number of responses
0	not answered	3
1	too long	1
2	just right	25
3	too short	2

Overall ranking: (2) just right

7) HELPFULNESS AND COMPLETENESS OF DOCUMENTATION:

Category		number of responses
0	no answer/unable to evaluate	5
1	no - inadequate	1
2	poor - needs a different approach	3
3	fair - needs significant improvement	3
4	yes - good/useful, some problems	18
5	yes - excellent/very satisfactory	3

Overall ranking: (3.7) good, but has problems
Notes: Needs better installation and operation procedures, standardized data-set documentation formats with more author input, better lineage and quality information.

8) DATA QUALITY:

Category		number of responses
0	no answer	8
1	no - inadequate	4
3	fair - needs significant work	5
4	mostly - some problems	7
4	yes - adequate	8

Overall ranking: (3.3) fair to good, but needs improvement
Note: Most of the lower scores here refer to the quality and scope of the original data, not the quality of the integration work. The poorly documented quality of existing data-sets is a primary motivation for this effort.

9) RESEARCH ACTIVITIES OF THE REVIEWERS:

(See specific responses in Appendix B)

10) POTENTIAL APPLICATIONS OF THE GED:

(See specific responses in Appendix B)

11) PROBLEMS IN EDUCATIONAL USE:

(See specific responses in Appendix B)

12) PROBLEMS IN DISSEMINATION TO SCIENTIFIC GROUPS:

(See specific responses in Appendix B)

13) COMMENTS ON INDIVIDUAL DATA-SETS:

(See individual data-set documentation chapters in the *Documentation Manual*)

WHAT IS A GIS?

GIS stands for Geographical Information System, which is a particular form of software applied to georeferenced, spatially distributed data in a uniform structure. This technology is relatively new. It has its roots in computer cartography and image processing, and the invention in the 19th Century of overlays to study spatial relationships.

A GIS integrates spatial and other kinds of information within a database structure and provides software tools that can be used to manipulate and display geographical data objects. Most GIS are graphically oriented, with display and map output capabilities ranging from cartographic displays, spatial imagery, and 3-D overlays, to graphs and histograms of data or statistical investigations. Additionally, the integrated data structure and standardized "tool kit" of GIS functions provides useful capabilities to aid in data exploration, data intercomparison, overlay studies, and more complex analyses of multiple data-sets in space and time. One of the most important current areas for growth in GIS is error analysis - an opportunity that is well supported by the multiple data-sets compiled here.

The GIS is a working environment and analytical philosophy that is established by the combination of geographical data objects (i.e., the database) and an organized set of data management, analysis, exploration, and descriptive modeling tools. The introduction of GIS has revolutionized the way the scientific world views and manipulates geographical information. Of particular interest here are the advantages that GIS methods may provide for quality assessment and analysis of observational data, and for the derivation of useful information related to global change.

SYSTEM COMPATIBILITY

It is not necessary to use the software provided for IBM-PC compatible microcomputers to access the data files on the CD-ROM, however numerical type conversions may be necessary depending on the particular operating systems and software being used. The conversion options provided with the DOS compatible software should be able to handle most requirements, including conversion to other GIS formats and some common data exchange formats. As of this release, however, the conversion software was not available for other operating systems.

UNIX systems can access the CD-ROMs directly, although, the interpretation of numerical values may be altered due to different conventions between UNIX and DOS operating systems. There are several issues to consider. First, the ordering of bytes in two-byte integer data may be reversed between systems. This can usually be corrected during the UNIX copy operation ("dd"), however, for signed integers the situation may be more complicated. Signed numbers in DOS are coded using a 2's compliment notation. This may be the case in other systems as well, however the UNIX/GRASS GIS, for example, requires signed numbers coded only with a sign bit (2's compliment notation changes all the bits to their binary opposite, then adds a sign bit).

The incompatibilities between systems would mount further if we had to consider floating point formats, and thus the decision was made to use ASCII notation for all data requiring floating point representation. There is an IEEE 4-byte floating point "standard" (which is implemented in the new version of IDRISI, for example), however this is not yet universally implemented on all computers. Also, four-byte integer data has fewer compatibility issues than real data types, but the range of numerical values allowed by 4-byte integers, though great, is still not enough to handle all data-sets. It is a major requirement of this project, sometimes at odds with integration, to avoid degrading a data-set for the sake of numerical representation. Surely the long-range solution is in better implementation of 4-byte integer and real standards, however until the volume of data-sets requiring real data types becomes great, the ASCII format can serve as the compromise.

Finally, it has been suggested that providing all data in ASCII form would provide maximum compatibility. While probably true, the large volume of data required, especially in the near future, not to mention the degradation in access times and loss of "off-the-shelf" functionality, makes it essential to pursue other solutions.

IDRISI (DOS) COMPATIBILITY

The software provided with the database gives the user basic access, display, exploration, and export/conversion capabilities, but does not provide GIS functions that would produce derivations from the database, neither in single nor multi-thematic operations. This limitation of the distribution software is balanced by the decision to maintain a generic database structure within an existing GIS environment, thus providing the most needed initial capabilities while ensuring compatibility with GIS in general.

The database is fully operable within the IDRISI GIS (Version 4.0), and the CD-ROMS as provided can be used directly as a resource disk. IDRISI operability provides a common denominator to facilitate the review process and dissemination to the widest possible group of users. This approach provides sufficient common ground (with a minimum of cross-training) to facilitate technical communication that is critically important in the peer-review process and for validating and improving the database. At the same time, full support is provided for upload to other systems.

The 1991 peer-review of the prototype CD-ROM confirmed that, while many global change researchers will use higher performance work stations (many with user-developed software), the need to ensure compatibility with GIS processing, support for diverse user applications, a broadly based peer-review process, and the opportunity for direct interaction with data developers, makes a "common denominator" system very important for building the database. This means that in addition to having a default system, the structure must be generic enough to allow compatibility across a range of available systems. The scientific and technical issues of integrating disparate data on varying themes and varying sampling schemes, presents challenges in maintaining both compatibility and operability. However, retaining both of these goals in actual implementation at this generic level provides a practical test for the database, and results in further specifications for both data and software.

Implementation of GIS satisfies important quality, operability, and review requirements, it is also designed to be as non-restrictive as possible for other platforms and software. Since it is not feasible to support even a majority of available GIS packages directly, the "common-denominator" approach reduces much of the difficulty in implementing the database with any system the user may obtain. To the extent that this can be done at the "low-end" of computer requirements, without compromising the information content or structure of the database, much wider use can be supported.

In fact, the current implementation provides extensive export and conversion capabilities for other systems, and special versions of the database are planned to overcome specific platform independence problems (for example, between DOS and UNIX).

This is something that cannot be accomplished by common data exchange formats alone, unless they also address the issue of common, comparable data structures, including data object definitions and geographical conventions that are characteristic of the GIS environment. Also, common exchange formats alone do not ensure suitable documentation and functional support for the wide range of applications desired of such computer databases.

GRASS (UNIX) COMPATIBILITY

Many global change scientists prefer to use a workstation environment, if available. Earlier versions of the database were successfully converted and used in the GRASS GIS at NGDC and ERL-C, and instructions were developed for this conversion process. As workstation users often have access to sophisticated technical expertise, many are able to perform the necessary format conversions "in-house." Nevertheless improved support for the workstation environment, beginning with GRASS, continues to be a high priority. A special version of the database for GRASS on UNIX systems is planned.

There are currently two ways for UNIX/GIS users to access the database. The first is by access and export from a CD-ROM equipped DOS workstation, using the software provided on floppy disk. This software (i.e., the SUBSET program), will allow the user to select a geographic window and choose the desired variables, or to select all data at full coverage, if desired, and create a GIS structured database on fixed disk in the correct format for direct use in a specified system. As of this release, only the GRASS format was supported, however additional formats are planned in subsequent updates.

The second option is to access the CD-ROM directly from within the UNIX operating system and run UNIX software to perform the necessary conversion.

See APPENDIX D: GRASS IMPLEMENTATION, for further details.

ARC/INFO/GRID COMPATIBILITY

An agreement was established with ESRI in 1991 for mutual evaluation of database operability within ARC/INFO/GRID. As of this release the evaluation has not been completed, however procedures are being tested for importing the database into ESRI's GRID software.

OTHER SYSTEMS

While specific systems have been selected to ensure "co-development" of the database and GIS operability, cooperative arrangements are also possible with other GIS developers. Interested parties are encouraged to evaluate this database from the standpoint of GIS operability, and to pursue interactions that may be mutually beneficial.

DOCUMENTATION

Documentation is provided for each CD-ROM in the form of a separate ***Documentation Manual***. The Manual contains descriptive information about each data-set and exact file descriptions for each element in the data-set. The data-set descriptions are structured according to a precise format that provides detailed information on the Source of the data used for the Global Ecosystems Database, the Original Design of the data from a statistical viewpoint, description of the Integrated Data-Set, Primary References and Additional References. This information is followed by information on Data Integration and Quality, and detailed File Descriptions for each data Element. The format of this information is carefully defined so as to standardize the information and provide a guideline for future contributors (see APPENDIX E).

In addition to their printed versions, the *User's Guide* and *Documentation Manuals* are provided in computer-readable form (bit-mapped image format) on the CD-ROMs. ***Reprints*** of the Primary References, including published journal articles, if applicable, are provided separately from the *Documentation Manual* as scanned image files on the CD-ROM. Reprints are assembled from available reports and publications, with permission from the authors and publishers. Only materials that are directly relevant to documenting the data-sets are reproduced.

There is also an ***IDRIX Technical Reference Manual*** that accompanies the IDRIX software from Clark University. This manual is provided in bit-mapped image format (in addition to it's printed form) on the software floppy disk. The IDRIX manual provides detailed descriptions of the software functions and operations available in IDRIX, many of which are extracts from IDRISI.

All document files have a ".PCX" file name extension (indicating PCX format). File names for the *User's Guide* begin with the letters "UG" followed by the corresponding page numbers, including the cover pages (UG_C1 and UG_C2), the preface pages (UG_I - UG_XI), and the main document pages (UG_001 - UG_112). The file names for the *Documentation Manual* begin with the abbreviation "ADOC" (or "BDOC"), corresponding to the CD-ROM Disc label, followed by the corresponding page numbers as above. Each reprint file is named with a prefix that is similar to the data file names in the corresponding data-set. The prefix is followed by a sequential article number, and finally a page-file number, e.g., "MGV1_01.PCX". The page-file numbers start over with page 1 for each article (but do not necessarily correspond to the original numbering of the article - they are sequential numbers starting with the first page-image-file). The reprint files are stored in numbered sub-directories, corresponding to the data-set chapter numbers in the *Documentation Manual*. Where readability is a problem due to a poor original or small print, a higher resolution image accompanies the default image. These are named identically to the corresponding low-resolution image, except that an "X" follows the page number (e.g., MGV1_01X.PCX").

The bit-mapped images are scaled for EGA/VGA display at 640 columns screen width. They are not intended for reproduction as printed documents. Software for search and display of these files is described in the section below.

IDRIX SOFTWARE (J. Ronald Eastman, Clark University)

Software is provided on floppy disk and distributed with the CD-ROMs for convenient data access, display, exploration, and export operations. This software, and its associated *Technical Reference Manual*, may be updated separately from the database and other documentation. For this reason, the user should refer to release notes supplied with the software for any modifications to the procedures presented below.

The database software is a special subset of the IDRISI Geographic Information (Analysis) System, developed by Clark University, linked by a special menu system aimed at facilitating data exploration, subsetting and downloading. This portion of the software is called IDRIX (for "IDRISI Explorer"), and has been developed partially under contract and through cooperative research with Clark University. Unlike IDRISI itself, which is distributed under a software license on a non-profit basis, IDRIX may be distributed and used freely in conjunction with the CD-ROMs in this database. IDRIX runs under MS-DOS. Similar software for UNIX systems is planned for future versions of the database.

The IDRIX access and exploration software can help the user find data relevant to one's interests, explore relationships and statistical properties, and structure subsets of the database for further analysis. It is not, however, a fully functioning GIS (such as IDRISI or GRASS). However, the software will prove to be invaluable in the initial exploration of data-sets and the subsequent uploading of information for use with a variety of GIS software systems.

The instructions that follow cover the installation and setup of IDRIX for use with the Global Ecosystems Database. Many users familiar with GIS (and many even who aren't) will find the system to be immediately useful. However, specific details about its operation can be found in the *Technical Reference Manual* (from Clark University).

INSTALLATION AND SET UP

First confirm that your CD-ROM reader is properly installed and that the CD-ROM drive has been initialized for access by DOS. Also note which drive letter has been assigned to your CD-ROM reader (e.g., "R:" in the examples below). You might test the reader, after inserting one of the GED discs into the drive, by executing a DOS directory list (DIR) command. This should give the following results:

```
DIR R:

 Volume in drive R is  GED-1.0-A
 Directory of R:\

DOCUMENT      <DIR>       04-01-92  12:00p
GLGEO         <DIR>       04-01-92  12:00p
SOURCE        <DIR>       04-01-92  12:00p
        3 file(s)             0 bytes
                              0 bytes free
```

IDRIX and the accompanying software must be installed onto your hard disk for use. To do so, insert the software floppy disk into the appropriate floppy drive (e.g., 1.2MB 5 1/4"; or 1.44MB 3 1/2" drive), and type:

A:\INSTALL or B:\INSTALL

The installation program will then prompt you to indicate the name of the floppy disk drive which contains the IDRIX distribution disk (e.g., A: or B:) and the name of the directory into which the program modules should be placed (e.g., C:\GED). The installation software will check to see if the destination directory exists, and if it doesn't, it will be created before copying begins. The installation program does not alter your AUTOEXEC.BAT or CONFIG.SYS files.

GETTING STARTED

Before running IDRIX, be sure to activate the computer's mouse, if one is installed. Also, it is advisable to establish a special DOS APPEND path to your working directory to gain access to files that may be created on your computer's hard disk, in addition to those on the CD-ROM (see explanation in Appendix C, pg. 103).

To run IDRIX, change to the GED directory on the drive where you installed the IDRIX software, e.g.,

```
C:
CD \GED
```

Then enter the name of the software menu program, e.g.,

```
IDRIX
```

The IDRIX menu system will then be run. To operate the menu system use either the up and down arrow keys or the mouse to move the menu highlight bar. Pressing return will select the highlighted option. Pressing return for any of the main menu items will produce a submenu. Selecting items from the submenu is the same as for the main menu. To exit a submenu, either select the "exit" option, or press the "ESC" (escape) key. To exit the menu system entirely, again select the "exit" option, or press the "ESC" (escape) key.

The first thing that should be done when using IDRIX is to examine the data directory paths that have been set up, and to modify them as necessary. This is done by choosing the first option on the main menu (**"Show/Set Data Directories"**). Then choose the suboption "Show Current Directories." Consistent with the structure of the CD-ROMs in this database, the software requires that you specify the paths (disk drives and directories) for the following directories on the CD-ROM:

1. The raster data files (e.g., R:\GLGEO\RASTER)
2. The vector data files (e.g., R:\GLGEO\VECTOR)
3. The attribute values files (e.g., R:\GLGEO\VALUES)
4. The meta-data files (e.g., R:\GLGEO\META)

In addition, you will need to specify an output directory for any files that may be created in your use of IDRIX (such as temporary work files and downloaded or subsetted data files). This must be a directory on your hard disk since the CD-ROM is a "read-only" device, e.g.,

5. Output data files (e.g., D:\DATA)

IDRIX will not create this directory, so please ensure that it does exist before using it. To change these directory path specifications, use one of the "set" options provided in this submenu. NOTE, however, that IDRIX will not access files created on the output directory for subsequent display or analysis unless the necessary directory settings in 1 through 4 are set to the output directory (e.g., on your computer's hard disk). To gain full functionality of all the features of IDRIX, however, access to files created and stored in the Output directory (on your computer's hard disk) is required in addition to CD-ROM access. This is because some functions create files in the working directory that you may wish to display or use for display (as in the case of modified color palettes or windowed, SUBSETed, or DOWNLOADed images, vector overlays digitized on-screen, etc.). One may use the DOS APPEND command to gain access to both the CD-ROM files, as set above, and the output directory files (i.e., any files that may be created by IDRIX). See Appendix C and your DOS manual for instructions on using the DOS APPEND command.

After setting all the directory paths, review them with the "Show Current Directories" suboption. Note that these settings will remain in effect until specifically altered, even if the computer has been off in the interim. Thus in your next use of IDRIX you may not need to reset these entries (however, you will have to re-establish any desired APPEND paths each time, from DOS before activating IDRIX).

Note that you will only be able to access one CD-ROM directory (e.g., GLGEO) at a time (unless you use APPEND). To access the SOURCE data-sets on the CD-ROM, re-set these directory path specifications accordingly.

Now that the directories have been set, you can start to explore the database. Try the "list" options to see what is on the CD-ROM. Then use the "describe" options to view the meta-data (documentation) concerning any specific file. Then use the "display" options to examine the files on screen. The "statistical exploration" menu provides some additional query tools while the "file management" menu will allow you to download, subset and alter file storage characteristics.

IMAGE DISPLAY

IDRIX is supplied with two color display modules -- one for standard VGA (and EGA) graphics adaptors and one for 8514/A compatible adaptors (with the appropriate driver loaded!). The VGA driver can display only 16 colors simultaneously (although those specific 16 may be chosen from any of 262,144 possibilities -- the specific 16 colors used at any one time being called a "palette") The 8514/A driver can display 256 colors simultaneously (again with palettes being composed from a total of 262,144 possible colors). Many Super-VGA (SVGA) adaptors can emulate the 8514/A standard, and this is highly recommended, especially for the satellite data. To establish 8514/A emulation, however, the graphics adaptor must have 1 megabyte of display memory and a memory resident driver known as the "applications interface." This driver is supplied by the manufacturer of your graphics adapter, or a third party, and may be known by names such as "HDILOAD.EXE" (IBM systems), "RIXAI8.EXE" (most Tseng and Trident based video cards), or others depending on your specific graphics adapter. Check the documentation for your graphics adapter for details on how to load this driver. Note also that XGA graphics adaptors, as well as many "accelerator" cards, can also emulate the 8514/A graphics standard.

Images may be viewed either "unscaled" or "autoscaled." With unscaled images, the display drivers assume that numeric values stored in the image can be directly interpreted as color codes. Thus, for example, a data value of 8 would be given color 8 (color 0 is also used for the background). If it should happen that there are not enough display colors to show the full sequence, only a subset of the data values will be displayed with instructions to press the "PgUp" and "PgDn" keys to see other subranges ("Home" and "End" keys move you to the subranges at the extremes). In other cases, such as with un-classed scalar data, you may wish to see the entire data range on the screen at once. In these instances the image must be "scaled" to fit the data range within the more limited range of available colors. This is called "autoscaling." Autoscaling is also required for viewing real number images (although the software recognizes this and will select autoscaling automatically if the image contains real numbers or negative numbers in the data range).

There is a broad range of interactive display capabilities with the COLOR display modules, including cursor/query, vector overlay, on-screen digitizing, zoom and windowing, palette change, edit, and saving, and others. Once a COLOR display option has been selected, the following key-commands can be activated:

c - cursor-query function with row/column readout
x - cursor-query function with lat/long readout
w - window/zoom feature
v - vector file overlay
p - select alternate color palette
k - edit color palette
s - saves screen or windowed image (to output directory)
d - on-screen digitizing
l - edit legend (will save to output directory)

DOWNLOAD AND SUBSET

"Downloading" refers to the transfer of the CD-ROM data to your system. While it is possible to use the CD-ROM directly as a resource directory for your system, it will often be desirable to stage your data onto a hard disk first. A hard disk provides much faster data access than a CD-ROM. IDRIX provides a special facility for selecting a set of files to be downloaded as a group. In the process, both the data files and their associated meta-data files are copied into the designated output directory. These files are copied directly with no changes to their internal content. "Subsetting" refers to the extraction of only a portion of the raster image. IDRIX provides a facility for subsetting whole groups of raster images simultaneously. The resulting sub-images are placed in the designated output directory. Note that vector files do not need to be subsetted to match up with subsetted raster images -- they may be downloaded directly.

OTHER FUNCTIONS

In addition to the COLOR display and SUBSET/DOWNLOAD functions, which provide the most important access functions, there are a number of other display and exploration functions selectable from the menu system, including statistical queries, histograms, 3-d plots, space or time profile graphs, printer outputs, data conversion, and other functions. See the ***Software Technical Reference Manual*** for details on these and other functions.

ON-LINE DOCUMENTS DISPLAY SOFTWARE (NGDC)

A simple menu-driven program called, MANUAL.EXE, for displaying the on-line documents is also provided with the access and exploration software. This program is invoked by typing the following, in your GED software directory (as established above):

MANUAL

The software will present you with a directory path specification that indicates the location of the on-line documentation. These files originally reside on the CD-ROM under the DOCUMENT directory, with the reprint files in numbered sub-directories corresponding to the data-sets. You will be asked to confirm or correct the root directory path for all documentation before proceeding. It may be particularly desirable, however, to copy the documentation files from CD-ROM to a hard-disk directory to improve the speed of access, using XCOPY /S to transfer all files and sub-directories under the DOCUMENT directory (the directory structure must be maintained as it is on the CD-ROM). If you have not copied the files to a hard disk directory, your directory path specifications should point to the CD-ROM, as shown below:

PATH TO DOCUMENTATION FILES : R:\DOCUMENT

Once your directory path is correctly specified, the software will proceed to display a main menu, from which you may make appropriate selections. Depending on your selection, the software will display one of the following: (a) a sub-menu for more detailed selection, (b) a sequence of pages from the manual, reprints, or accompanying files. Each page is a bit-mapped image or ASCII text corresponding to your menu selection. For bit-mapped images, one can toggle between the default screen-size and high-resolution images (for some hard-to-read pages) using the +/- keys (however, note that memory requirements restrict this feature -- see implementation notes accompanying the software). Previous or subsequent pages can be viewed using PgUp and PgDn keys, and the high-resolution displays can be panned using standard cursor keys. Escape returns you to the previous menu. The bit-mapped images are displayed using GRASP software (GRASPRT), developed for public distribution by Paul Mace Software, Inc.

DIRECTORY STRUCTURE AND FILES

The DOS CD-ROM database is organized in standard files and directories, accessible by the operating system or applications software. The directories are structured as follows:

```
DISC A, B, etc.
 |
 |--DOCUMENT :   User's Guide, Documentation Manual, and Reprints
 |      |-- USER
 |      |     {ug_*.pcx}    :     User's Guide PCX page-image files
 |      |-- ADOC
 |      |     {adoc_*.pcx}  :     Documentation Manual PCX page-image files
 |      |-- A01 :    Reprint sub-directories;
 |      |-- A* :     (additional directories for each data-set chapter)
 |            {*.pcx}       :     Reprint PCX page-image files
 |
 |--GLGEO :      Global, geographic (lat/lon) database, Version 1, Disc A or B
 |      |
 |      |-- META :          Header and supplemental meta-data files (IDRISI version 4x)
 |      |     {*.doc}       :     ascii raster documentation files
 |      |     {*.dvc}       :     ascii vector documentation files
 |      |     {*.dvl}       :     ascii values documentation files
 |      |     {*.pal}       :     ascii color palette files
 |      |     {*.ts}        :     ascii time-series list files
 |      |-- RASTER :        Raster data files
 |      |     {*.img}       :     raster image files
 |      |-- VALUES :        Attribute values files
 |      |     {*.val}       :     attribute values files
 |      |-- VECTOR :        Vector data files
 |            {*.vec}       :     vector data files (point, line, or polygon)
 |
 |--SOURCE :     Source versions of data-sets with special integration problems
        |
        |-- META :          Header and supplemental meta-data files (IDRISI version 4x)
        |     {*.doc}       :     ascii raster documentation files
        |     {*.dvc}       :     ascii vector documentation files
        |     {*.dvl}       :     ascii values documentation files
        |     {*.pal}       :     ascii color palette files
        |     {*.ts}        :     ascii time-series list files
        |-- RASTER :        Raster data files
        |     {*.img}       :     raster image files
        |-- VALUES :        Attribute values files
        |     {*.val}       :     attribute values files
        |-- VECTOR :        Vector data files
              {*.vec}       :     vector data files (point, line, or polygon)
```

DOCUMENTATION FILES

The **DOCUMENT** directory contains this *User's Guide* and the *Documentation Manual* for the corresponding disc, in digital form. The digital files are stored in a PCX bit-mapped format for on-line display. Software is provided on the accompanying floppy disk for IBM compatible DOS machines. (See ACCESS AND EXPLORATION SOFTWARE, ON-LINE DOCUMENTATION)

GLOBAL GEOGRAPHIC (LAT/LON) DATABASE

The **GLGEO** directory contains the main nested global database in a Geographic (i.e., unprojected, latitude/longitude) coordinate reference system. All raster (cell-based) data files have a ".IMG" file extension and are stored in the **RASTER** directory. All vector (point, line, or polygon) data have a ".VEC" file extension and are stored in the **VECTOR** directory. All attribute values files, used to reclassify either raster or vector data layers, have a .VAL file extension and are stored in the **VALUES** directory.

All header and ancillary meta-data files, since they are more subject to software requirements and version changes, are stored in a separate directory called **META**. These files conform to the format of IDRISI Version 4.0, described above, and contain the necessary coded information (meta-data) to properly interpret, display, or analyze the data-sets. If a different version of the software is employed, the user will most likely need only to convert the files in the **META** directory, which represent a much smaller volume than the data files themselves.

SOURCE (NON-NESTED) DATA FILES

Integration of source data-sets into the GED convention of nested grids occasionally requires interpolation and re-gridding. Examples of each data-set that presented special problems in integrating with the GED nested-grid convention, are provided in the **SOURCE** directory. The sub-directories have the same structure as the GLGEO directory described above.

All integration methods are described in the *Documentation Manual.* These source data examples are provided for experimentation with integration methods, and for comparison with the GLGEO files. For example, what method of interpolation is suitable to match grid registration with the GED (e.g., the Legates and Willmott climatic data, originally cell-center registered rather than cell-edge registered)? What does it mean to aggregate classed data to a coarser grid for use with other data (e.g., gridded FAO soils data) -- do the class meanings change? How should the database represent data provided on irregular grids, such as a lat/long grid of 2-degree by 2.5-degree cells (e.g., many data-sets designed to interface with global models)? What is the appropriate method of temporal compositing for time-series data, re-projecting to different grids, and handling quality information embedded in a data-set as a mask (e.g., the EDC-NESDIS Experimental Global Vegetation Index, and NGDC Generalized GVI data-sets)? Each of these questions presents unique problems for the GIS user who wishes to compare or combine information from data-sets that are not geographically compatible (see NESTED GRIDS AND REGISTRATION CONVENTION, below).

FILE NAME CONVENTION

In this version of the database (Version 1, for DOS systems) the filenames conform to the IDRISI name convention, whereby the file type is identified by file extensions (defined above). These file extensions are intended to be somewhat intuitive. For example raster data files are called "image" files in IDRISI terminology and thus have an ".IMG" file extension after the name.

The file names are standardized to aid in searching the database or copying files using wildcard file specifications. There is a convention for labeling time-series data-sets, whereby after the descriptive name, the next two numerical characters refer to the year and the last two numerical characters (before the file extension) refer to the month. This is shown below:

```
MGV  87  06  [.IMG]  [.DOC]
 |    |   |    |       |
 |    |   |    |     .DOC => ascii header file
 |    |   |  .IMG => data file
 |    |  06 = month (e.g., June)
 |   87 = year (e.g., 1987)
MGV = data-set name (e.g., "Monthly Generalized Vegetation index")
```

Data-sets that provide only average (typical) months will simply show the month, in which case the file name may extend to six characters. For data-sets that do not have a time dependency, the data-set name may occupy up to 7-characters. The 7-character limit allows one-character user prefixes or suffixes for files that are created using the SUBSET program. A precise convention for all data-sets is not possible because of the restrictions of a 7-character file name; however, each data-set is named so as to allow a unique wildcard specification for all the files in the data-set (usually the first two or three letters of the file names).

GLOBAL COORDINATES

The global database called GLGEO on the CD-ROM consists of a collection of raster (cell) and vector (point, line or polygon) data-sets (with file extensions ".IMG" and ".VEC," accordingly), each of which represents a global digital map (spatial distribution). All data are provided in a Geographic (Longitude/Latitude) projection (i.e., "unprojected").

The global grids and vector data-sets in this database all have a common origin and geographic window. The western and eastern edges of the grid and vector data-sets fall on the International Date Line, or 180 degrees East/West longitude (+/- 180 degrees). The Greenwich (Prime) meridian thus bisects the window in the east-west direction. Similarly, the northern and southern edges of the grid and vector data-sets are located at the poles, or 90 degrees North latitude (+90) and 90 degrees South latitude (-90); and the equator bisects the map in the north-south direction. The boundaries for this global window are thus:

Maximum longitude (X)	:	+180 degrees (East)
Minimum longitude (X)	:	-180 degrees (West)
Maximum latitude (Y)	:	+90 degrees (North)
Minimum latitude (Y)	:	-90 degrees (South)

This information appears in the header (.DOC and .DVC) files stored in the META directory for use by the GIS software (described below). Note that the geographic limits indicated in these header files refers to the bounding rectangle of the image/map, not the centroid of raster cells (as required in some GIS window specifications, e.g., GRASS), nor, in the case of vector data, does it refer to the actual geographic limits of the data.

By convention, the software provided with the database counts row and column coordinates starting with zero (0). Using this convention, the corner pixels of a ten minute data-set (such as the Global Vegetation Index) are:

	(column,row)
North West corner:	(0,0)
North East corner:	(2159,0)
South West corner:	(0,1079)
South East corner:	(2159,1079)

Similarly, the corner pixels in a one degree grid are:

	(column,row)
North West corner:	(0,0)
North East corner:	(359,0)
South West corner:	(0,179)
South East corner:	(359,179)

RASTER GRID (IMAGE/MAP) DATA FILES

Raster grid data are structured in a basic "image" structure, with sequential values in the file corresponding to a row-wise structure, beginning with the cell at the upper-left (North-West) corner of the geographic area. In the case of geographic (latitude/longitude) data in the GLGEO directory, this corresponds to the cell with its north-west corner on the International Date Line, at the North pole. Rows are filled with sequential values from the data file up to the row size (specified in the documentation file, described below), repeating this pattern for each row from top (North) to bottom (South). Clearly, an artifact of this projection is that the extreme North and South rows are artificially expanded longitudinally. In reality, they represent pie-shaped cells in a circular area around the pole with a radius of 1 pixel height. At the equator, the latitude/longitude projection results in cells that are very nearly rectangular in their true ground distance. Such uneven sampling across the globe is a strong argument for using other tessellations, however to date, most global data have been produced in latitude/longitude coordinates. In future versions of the database, other projections may be employed, beginning with the polar regions.

The number of values in a raster data file will equal the number of rows multiplied by the number of columns. All raster data are stored as binary files, as one or two-byte integer data types or 4-byte IEEE floating point data type. The byte count for each file can be calculated by multiplying the number of values (Rows x Columns) by the number of bytes per value (1, 2, or 4). The sequencing of values in the "image" is shown in the following diagram (for a five column image):

1	2	3	4	5
6	7	8	9	→
→	→	→	→	→
→	→	→	→	→
→	→	→	→	→

NESTED GRIDS AND REGISTRATION CONVENTION

A convention of commonly registered "nested" grids has been adopted whereby various grid sizes may be used, all of which are integer multiples or divisions of each other. The possible grid cell sizes in this scheme are 2-degree, 1 degree (60 minutes), 1/2 degree (30 minutes), 1/6 degree (10 minutes), and 1/12 degree (5-minutes), 1/60th degree (1-minute), and 1/120th degree (30-seconds). This system captures most of the commonly available grids, and allows intercomparison of data-sets at any nested scale, without altering the numerical or classed values of the data or introducing geographic aliasing effects from "nearest neighbor" resampling. The analyst must understand, of course, that the comparison is between data of different scale, and that special considerations must be made in interpretation. Nevertheless, statistical comparisons between, for example, soil or vegetation classes at .5 degree, and similar classifications at 1 degree are useful for quality assessment purposes and to develop derivation or integration methods prior to using data in analysis or modeling. Similarly, comparisons between classed data and finer grid satellite data can be extremely useful for validation and quality control/assessment.

Data on other "non-nested" grids, as discussed earlier, present a special problem because they require re-sampling before they can be geographically compared to the other data. While there are many methods for doing this, they all introduce statistical issues that would best be evaluated by the original investigators, after careful analysis. Examples from data-sets which presented integration problems are provided on the CD-ROM. For example, methods may vary for comparing data on a 2-minute grid with data on a 5-minute grid. Or for comparing data centered on 1/2 degree latitude and longitude meridians, with similar thematic data, identically scaled but edge-matched on 1/2 degree meridians. Or for comparing data with rectangular 2.5 X 2 degree cells with regular grids at 1 degree.

Different assumptions must be made in the process depending on the nature of the data. The necessary re-sampling for such comparisons will always involve greater uncertainty than re-producing the differently scaled or registered grid from original data.

It is clear that the problem is best solved by working with the principal investigators to re-produce the data on one of the recommended grids. This would ensure the ability to carefully document and defend resampling and interpolation methods. Nevertheless, this is not always practical and user feedback may be important for establishing such priorities. While it is possible that aggregation of these data to coarser scales may relieve much of the concern in a given study that requires only coarser information and has developed adequate aggregation methods; the need for quality assessment requires intercomparison with other data at the full sampling resolution. This is a central issue in establishing "nested-grids," and it is important to know if the proposed grid convention can adequately deal with the problem. Improved versions of the GED may encourage data developers to support the nested grid structure, to facilitate intercomparison for quality assessment and validation purposes. In the future, coarse grids from GCM model outputs in the 2 to 5 degree range may be added to the database, and a decision will have to be made on their common representation.

In cases where the original data-set does not extend to full global coverage, or has missing values, the grid was filled or padded with appropriate "no data" values to complete the global grid. Padding to a larger window than the original data-set is theoretically unnecessary for software systems that have full georeferencing (e.g., GRASS and IDRISI, although IDRISI 4.0 does not yet fully utilize it's georeference information). Nevertheless, padding allows the specification of more explicit "no data" flags, preventing ambiguity in any case and simplifying overlay in non-georeferenced systems.

When re-gridding from fractional grid dimensions was required to represent the data within the nested-grid convention, the decision to aggregate or expand to one of the common grids was determined by knowledge of spatial variability. Simple areaweighted averaging (using the percentage of overlap of output cells with the originals) was used in some cases, if the spatial variability was known to be on the order of a pixel. This method is spatially un-biased but has the effect of "smoothing" the data. For example, the NGDC Monthly Generalized GVI was produced by aggregating in this way to 10-minutes cells from approximately 8.6-minute cells of the source "plate carreé" grids. An alternative method, given similar spatial uncertainty, is to re-sample by a "nearest-neighbor" method, which has the advantage of preserving the original data values, but at the cost of some aliasing and spatial distortion (thus implying that spatial accuracy is limited by the cell size). In other cases, where the statistical validity of such averaging cannot be assessed, a more conservative approach may be taken of re-sampling the data to a finer grid in the nested structure, to preserve the original variability or actual data values within the new grid. This allows intercomparison of the original and interpolated values with other data in the nested-grid structure, and may be a reversible process: The cost in this case is storage space for the replicated data. In all other cases, where no geo-referencing problems existed, the grid representation in the database reflects the original geographic sampling.

The cell size in a given data file is determined by the global window and the number of rows and columns (not the "resolution," which refers to the resolution of the original data sampling, not necessarily the current grid). Since cell sizes can vary between raster data layers according to the nested-grid convention (even multiples or divisions of cell size), all overlay and intercomparison operations must expand or contract the maps accordingly. At present, there are few systems which do this automatically (e.g., GRASS, and planned for future versions of IDRISI). In all other cases, the user must expand or contract the maps in a separate operation prior to overlay or intercomparison (see implementation section for specific operations in IDRISI and GRASS). The nested-grid convention allows this to be accomplished by simple pixel replication, without interpolation. While allowing direct intercomparison, this method does not change the original spatial definition of the data, which must always be taken into consideration. (Decimation -- i.e., choosing every nth pixel -- or aggregation -- e.g., averaging or reclassifying -- to coarser grids must be done more carefully to ensure valid geographic and numerical definitions.)

RASTER DOCUMENTATION FILES

Each data-set is accompanied by an ASCII documentation file which contains the ASCII header information necessary for the software to correctly interpret the corresponding data file. The following shows an example of a documentation file, with explanations in brackets.

SAMPLE.DOC

Name of the spatial data element	file title	: Sample data file header
Byte, integer, or real	data type	: byte
File storage method (ASCII or binary) ..	file type	: binary
Number of vertical columns in grid ...	columns	: 2160
Number of horizontal rows in grid	rows	: 1080
Georeference system (projection)	ref. system	: lat/long
Georeference basic unit of measure	ref. units	: deg
Georeference unit multiplier	unit dist.	: 1.0000000
West-most geographic coordinate	min. X	: -180.0000000
East-most geographic coordinate	max. X	: 180.0000000
South-most geographic coordinate	min. Y	: -90.0000000
North-most geographic coordinate	max. Y	: 90.0000000
Spatial location uncertainty	pos'n error	: 0.1666667
Sampling interval of ORIGINAL data ..	resolution	: 0.1666667
Minimum data value in the data-set ...	min. value	: 0
Maximum data value in the data-set ...	max. value	: 32
Data values units of measurement	value units	: scalars or classes
Estimated error of data values	value error	: estimate
Special value to flag non-data	flag value	: 255
Definition of flag	flag def'n	: no data
Number of legend categories	legend cats	: 3
Category 0	category 0	: Water
Category 1	category 1	: Podzol Soils
etc.	category 2	: Brown Podsolic Soils
Lineage line (as many as needed)	Lineage	: see documentation
Comment line (as many as needed)	Comment	: notes as needed

SPECIAL NOTE:

(1) The grid (image) data and documentation files conform to the IDRISI Version 4.0 GIS format, which contains full georeference information.

VECTOR DATA FILES

Vector data are provided in an ASCII file type only, in the following formats:

```
{ID#}          {#Points}
xxx.xxxxxx     yy.yyyyyy
xxx.xxxxxx     yy.yyyyyy
{ID#}          {#Points}
xxx.xxxxxx     yy.yyyyyy
xxx.xxxxxx     yy.yyyyyy
[etc.                  ]
[etc.                  ]
[etc.                  ]
0       0
```

The first line is a point, segment, or polygon label. In the label line, the first number (ID#) is a feature identification number, and the second number (#Points) is the number of coordinate pairs that follow to define the feature. On each line following the header are the coordinate pairs that make up the feature. The first number (xxxx) is the "X" coordinate (longitude) and the second (yyyy) is the "Y" coordinate (latitude). Each feature (point, line, or polygon) begins with a new label line, followed by it's coordinate pairs for as many points that define the feature. The last line of each vector file consists of two zeros, separated by one or more spaces (this terminates the listing of features). Coordinates are in the same units as the bounding rectangle specified in the vector documentation (.DVC) file, with positive values indicating East (X) and North (Y), and negative values indicating West (X) and South (Y). In this format, point data files would have only one coordinate pair defining each feature (point), line data files can up to 750 coordinate pairs for each feature, and an unlimited number of features. Polygon data files are similar to line data files except that each feature is closed (i.e., the starting point and ending point are identical). The units and reference system for coordinate values used in the vector files are defined in their corresponding documentation files.

VECTOR DOCUMENTATION FILES

Name of the spatial data element	file title	: Sample vector file
Data type for feature labels (integer) ...	id type	: integer
File storage method (only ASCII used) .	file type	: ascii
Feature (point, line or polygon)	object type	: line
Georeference system (projection)	ref. system	: lat/long
Georeference basic unit of measure	ref. units	: deg
Georeference unit multiplier	unit dist.	: 0.0166667
Minimum x coordinate of study area ..	min. X	: -10800.0000000
Maximum x coordinate of study area ..	max. X	: 10800.0000000
Minimum y coordinate of study area ..	min. Y	: -5400.0000000
Maximum y coordinate of study area ..	max. Y	: 5400.0000000
Spatial uncertainty	pos'n error	: unknown
Mean point spacing (sample resolution)	resolution	: 0.01666667
Comment line (as many as needed)	comment	: notes as needed

SPECIAL NOTES:

(1) The bounding rectangle corresponds to the global window as described above (not the actual range of values in the file), in the spatial units specified.

(2) The vector data and documentation files are compatible with the IDRISI version 4.0 GIS, which is fully georeferenced.

ATTRIBUTE VALUES FILES

Attribute files contain tabular data used for re-coding an existing spatial data element, or attaching additional labels to coded features. They are used most commonly with vector data-sets that define a set of boundary features for which there are numerous kinds of data (e.g., political boundaries with various statistics for each polygon, or streams with various measurements for each coded portion of the stream). Such data would be described as having only one spatial "distribution" but many "variables". There is no technical reason that attribute files cannot also be attached to raster data-sets, where each data value has a unique translation in the attribute table. Attribute files, whether attached to vector or raster data elements of a data-set, are stored as ASCII files with the following formats:

1	570	Morrison	etc...
2	860	Dakota Ridge	
3	120	Blue Mountain	
etc.			

These "flat files" may contain variable width columns of attribute data up to a total record length of 255 characters. Each attribute column may contain integer values, real (floating point) values, or character strings.

ATTRIBUTE VALUES DOCUMENTATION FILES

Name of attribute data element	file title : Example Attribute Doc.
File storage method (ASCII only)	file type : ascii
Number of values or codes in spatial file	records : 3
Number of attribute fields	fields : 2
Feature ID field (matches spatial file) ..	field 0 : identifiers
Type of field (integer, real, string)	data type : integer
Field format code (unused)	format : 0
Field definition	field 1 : biomass
Type of field (integer, real, string)	data type : integer
Field format code (unused)	format : 0
Minimum value in column (null=0)	min. value : 100
Maximum value in column (null=0) ...	max. value : 2532
Units for field values	value units : gms./sq.m.
Class or value uncertainty	value error : unknown
Special flags in field data	flag value : 255
Flag definition	flag def'n : no data
Field definition	field 2 : site names
Type of field (integer, real, string)	data type : string
Field format code (unused)	format : 0
Minimum value in column (null=0)	min. value : 0
Maximum value in column (null=0) ...	max. value : 0
Units for field values	value units : classes
Class or value uncertainty	value error : unknown
Special flags in field data	flag value : none
Flag definition	flag def'n : none
[repeat previous 9 lines for each field ..	[add'l fields up to 20]
[up to a record length of 255 characters]	["]
Number of legend categories	legend cats : 3
Category 0 (alternative legend)	category 0 : class def. zero
Category 1 (alternative legend)	category 1 : class def. one
Category 2 (alternative legend)	category 2 : class def. two
Comment line (as many as needed)	comment : notes as needed

SPECIAL NOTES:

(1) The bounding rectangle corresponds to the global window as described above (not the actual range of values in the file), in the spatial units specified.

(2) The vector data and documentation files are compatible with the IDRISI version 4.0 GIS, which is fully georeferenced.

DATA STORAGE FORMATS

Raster data files are stored using either byte binary, integer binary, or real (floating point) binary numbers. Byte data are coded in 8 binary bits, thus allowing numbers from 0 to 255 (decimal) to be stored. Integer data are coded as two 8-bit bytes (16 bits), in "little endian" format, i.e., with the most significant byte on the right. Since one bit is used for the algebraic sign, numbers from -32,768 to +32,767 (decimal) can be stored as integer data. The order of the two bytes used for integer data conforms to conventions used by IBM (DOS) and DEC. Some systems (e.g., GRASS/UNIX) require data in "big-endian" format, with the most significant byte on the left (i.e., the order of the data must be reversed, or "byte-swapped," to be correctly interpreted), and representation of negative numbers may also vary between systems (see SYSTEM COMPATIBILITY). Binary real data types (floating point), because of compatibility issues, are used in the database only where it is impossible to represent the full range of the original data-set with integers. An IEEE standard 4-byte real data type is used (this may not be readable on all machines, but is common in DOS). The numerical type for each data file is indicated in their associated documentation files, which are stored in ASCII format. Vector data files (and their associated documentation files) are all stored in ASCII format.

The ASCII format on the Version 1 CD-ROMs uses the standard DOS convention for terminating a line with two bytes, a carriage return (hexadecimal 0D) followed by a line feed (hexadecimal 0A). In DOS, ASCII files are commonly ended with a DOS end-of-file mark (hexadecimal 1A), but this EOF mark is not required by most software, including that provided with the CD-ROM, and is thus not used to avoid incompatibilities with other systems. The binary data files (.IMG) also do not have an EOF mark. File size is thus an exact byte count of the data.

APPENDIX A -- TASK DESCRIPTIONS

John J. Kineman
Global Ecosystems Database Project
National Geophysical Data Center
Boulder, CO

INTRODUCTION

This paper provides a description of the Global Ecosystems Database Project, divided into 3 major Task Levels (Database Integration and Design, Technical Development, and Scientific Support and Methodology), which are further divided into 9 Tasks. These tasks reflect actual operational divisions to a large degree, but also represent distinct philosophical divisions and levels of scientific support. The last section, Extension to Education, Outreach and Policy Support, is the goal for many people interested in this database. The following discussion may help clarify the magnitude of the work needed to pursue that goal in a credible manner.

DATABASE INTEGRATION AND DESIGN

TASK 1: DATA AVAILABILITY AND NEEDS ASSESSMENT

To assess data and information needs, there must be an in-depth evaluation of data availability, including as much information on data design and usability as possible. The project thus maintains and updates a meta-database for global change data-sets to improve documentation and support needs assessment. Acquisition and processing priorities are established on an annual basis, in response to the needs and availability analysis from NGDC, needs analyses at ERL-C and NGDC, and the results of the previous year's review. Needs are defined by EPA and NOAA requirements for research; however, it is clear that many of these requirements are shared with the global change community in general.

Priorities thus established may range from acquisition of known data-sets that require little processing, to recommendations on derivation methods or specific algorithms to produce the desired data from existing sources. The needs assessment also distinguishes between data requirements that can be generalized for wide application, versus unique requirements of the supporting agencies and their immediate research targets.

TASK 2: DATA ACQUISITION, INTEGRATION, AND ARCHIVE

The project interfaces with the normal data management operations of NGDC, and also pursues new activities. These activities focus on GIS and related methods for the reprocessing and integration of individual data-sets received in diverse forms, into the common GIS structure established for the database. The project also aims to develop useful variables required for research support. This involves up to four levels of work:

(a) Data Management:

Data are acquired through various channels and made available in integrated form on CD-ROM. The data acquisition effort seeks source data-sets that are as close to original investigations as possible, or are especially appropriate for specific requirements. The integration effort seeks to represent these data-sets in as close to their source condition as possible, given the requirements of integration. Since error estimates are typically lacking for most data, it is especially important to preserve full representation of quality information within or accompanying each data-set. Similarly, independently derived, multiple data-sets are included to increase intercomparison opportunities between similar data within the database, for quality and error estimation.

As data are ingested into the NGDC system, all versions of the data-sets are archived, from the original source through various modifications produced for the purpose of integration. Documentation is archived with the data, and is annually compiled into the database *Documentation Manual.*

The existing database (established in the 1990-1991 effort) will be augmented by new data-sets acquired from external sources or developed at NGDC or ERL-C, with publication clearances for both data and documentation. All portions of the database, with the exception of re-printed journal articles and custom software, are placed into the public domain. Many individual source data-sets are not distributable by NGDC due to various restrictions placed on them by the investigators or source institutions; or more often, in deference to other distribution agreements. Derivations from such source data-sets, however, as part of an integrated and operational database, are developed and distributed as different products from these source versions.

(b) Data Processing:

Once data are in the system, the data are processed for data management purposes, to inspect quality and content, to verify documentation, and to determine the optimal data structure for integration into the main database. Depending on the nature of the data-set, integration may involve anything from simple media transfers and format conversions, to re-structuring operations such as geographic registration, grid re-sampling, gridding from point data, re-projecting, vectorizing, rasterizing, tabularizing, and/or other forms of interpolation. This work incorporates variously compatible data-sets into a common data structure based on geographical objects and community-wide GIS conventions.

The most common changes are conversions between numerical types (i.e., real, integer, ascii, binary, etc.). Actual re-classifications are avoided unless absolutely necessary due to extremely unusual or cumbersome classification schemes in the source data, in which case every effort is made to involve the principal investigator(s).

(c) Data Derivation:

A number of data-sets represent "derivations" from raw observational data, such as NDVI derivations from the Advanced Very High Resolution Radiometer (AVHRR) satellite sensor data, or monthly composites using a variety of methods. "Derivation" in this context is defined as a modification of raw data from a single thematic source, using accepted methods of calibration and/or correction. Although such work will defer, where possible, to other established sources (that contribute such derivations into the public domain), some specific derivations may be produced at NGDC. Priorities for derivation are jointly established by the supporting programs, according to needs and resources.

(d) Data Synthesis:[1]

Simulation models and other applications may require spatial variables that are missing from the observational database or not directly derivable from it. In such cases, it may be possible to substitute for the missing variable through a complex derivation using established relationships and transformations between multiple data-sets. This is a form of characterization that can be performed at NGDC and elsewhere using GIS methods, given suitable guidance on transformation equations and derivation procedures. As above, such activities will defer to other established sources whenever possible, but otherwise may be jointly designed by NGDC and ERL-C according to needs and resources.

TASK 3: DATABASE STRUCTURE AND FUNCTIONALITY

A common database structure is possible using currently evolving GIS conventions. This has many benefits in supporting quality control, peer-review, and the sharing of research applications of the database. It is important to note the co-evolutionary nature of the database structure and GIS functionality, and the need for further research and experimentation on both.

Table 1 lists some major descriptive analysis functions that are required for adaptive characterization, along with their purpose (Kineman, 1993). These functions correlate well with GIS concepts, which imply a number of conventions that are already well established, such as geographical (vector and raster) object definitions, common geographical referencing, labeling and legend conventions, and others. Future research and development in GIS, if properly informed by the global change community, can provide systems that are increasingly well suited for environmental characterization.

Linkage with GIS is important in forming the required structural and functional aspects of this integrated characterization database, and in performing many of the processing and quality control operations. Beyond this, there is also a strong need for a common environment to support peer-review and communication with a wide global change community of data users and developers.

[1] Not to be confused with information synthesis, which involves the compilation and presentation of results from multi-disciplinary studies.

Table I

DESCRIPTIVE ANALYSIS	
FUNCTION	SCIENTIFIC PURPOSE
Data Integration	Multi-thematic representation of variables, within a common analytical structure, with meta-data for verification and error analysis.
Exploratory Data Analysis	Visualization; exploration; hypothesis formation
Inter-comparison and statistical analysis	Quality and error assessment; empirical testing (hypotheses, models, and characterizations); statistical modelling
Re-sampling and error analysis	Experimental Design / re-design for various applications
Environmental Characterization & Data Synthesis	Multi-disciplinary representation of phenomena; monitoring and assessment; description of patterns and trends; model inputs and tests

Advances in data structures and meta-data are strongly influenced by developments in GIS functionality. The reverse has also been true, with system enhancements resulting from data integration requirements and the effort to track more kinds of data and meta-data to meet the demands of global change research. As clearly as this relationship exists, it is still not easy to maintain such links for generic purposes or to disseminate general system requirements to GIS developers. By maintaining strong links with GIS development, these co-dependencies can become better illuminated.

As data are processed into a common structure that is intended to be compatible with emerging GIS conventions, we must also deal with changing conventions and improvements in GIS structures. This is accomplished by linking the database with more-or-less "generic" GIS's that are currently available and offer the possibility for collaborative improvement. The use of IDRISI (from Clark University) began during the pilot and prototype phases and has continued because of the value of having a "common denominator" to facilitate integration, review, data exchange, and technical communication. This has remained true even in serving the EPA's needs, which are primarily reliant on using the database within the GRASS GIS, on Unix systems. UNIX, though common among global change researchers, does not provide the widest commonality among reviewers, data developers, and internationally distributed scientists and data sources; whereas the IBM-DOS compatible environment has made significant inroads in most technical facilities. Furthermore, maintaining operability at the lowest operable system level has introduced little if any sacrifice of quality or operability in more capable systems (other than the problem of format conversion).

The collaboration with Clark University (IDRISI) has been exemplary in showing how structure and function can co-evolve. Based on this interaction, IDRISI has been significantly modified to support global change studies, and in turn, the database structure has evolved through use within the GIS environment. Using this example, similar relationships may be developed with other GIS developers.

This project goes beyond the issue of providing a common format for data exchange between systems and users, which is already the focus of several major efforts within Government and other institutions (formats for the integrated database are relatively generic and easily converted to other standards, which generally require less detail about the data). An integrated database is a more in-depth approach, that correspondingly sacrifices breadth in its coverage of overall data availability. Instead, the focus is on the intercomparability, inter-operability, and verifiability of data-sets, using a selected portion of the overall data pool that must be prioritized by specific research.

The structure and function of a characterization database is strongly related to Geographic Information Systems (GIS) methodology, to such an extent that GIS integration and improvements in GIS methods are of primary concern, along with the provision of useful data. For this reason, agreements may be formed with GIS researchers, and experimentation by GIS developers is encouraged to promote the inter-operability of the database within the established GIS community and among global change researchers using GIS methods.

This will involve experimentation with GIS structures using at least one existing GIS as a development platform, and working closely with the software designers to incorporate improvements in:

(a) data structures and formats (raster, vector, tabular, data types, compression, etc.)
(b) raster/vector integration,
(c) integration of tabular data structures,
(d) export to other systems and formats,
(e) data exploration and statistical capabilities,
(f) scale integration and interpolation functions (including re-projection).
(g) lineage tracking
(h) techniques for error representation
(i) error estimation and tracking

TECHNICAL DEVELOPMENT

TASK 4: QUALITY ASSESSMENT AND DOCUMENTATION

The greatest problem in the distribution of environmental and ecological data-sets is their usefulness outside the institutions or programs that created them, and within a rapidly expanding global change community. While review and publication standards provide effective quality assurance for research in general (including the production of data-sets), they apply less directly to the distribution of data-sets for subsequent uses, which is often a second or lower priority in research programs and funding. The result is that when data are removed from their original research context, we have inadequate mechanisms for evaluating their design or verifying their accuracy for given purposes. This issue of verifiability becomes critical in the context of large multi-disciplinary system studies such as global change, which must rely on a common data pool.

Documentation therefore involves more than providing information about formats and data structures. Verifiability means the ability to assess the entire production method, statistical nature, and accuracy of the data-set, given its original purpose and present use. Ideally, the overall effort will provide the means for users to "re-design" or synthesize data for specific purposes with known confidence limits. This requires considerable knowledge about the data, i.e., "meta-data" and documentation. In this project, such information is obtained from existing documentation and published articles, returns from the external review process, and results of internal quality assessment efforts. Documentation is produced and added to the database *Documentation Manual*. This manual contains all available information needed to understand the design and nature of the data, from a technical/statistical (and mathematical, in the case of models) perspective.

Internal Quality Control procedures (which extend to all tasks) are implemented to ensure error-free processing and a full and accurate reproduction of data-sets in their complete form, as close to their originally design as possible, even though structures and formats may be changed to achieve full integration with the database. The need for retrospective quality assessment of many data-sets themselves (which typically have inadequate documentation or meta-data), has led to an emphasis on "Exploratory Data Analysis" methods, some of which have been experimentally incorporated into GIS software. In some cases, this has also spawned research projects related to specific data-sets, for example on the quality and use of AVHRR and Vegetation Index data, and other examples in topography, vegetation classifications, climate data, etc. Some of this research also results in new methods for data integration, such as gridding or re-gridding techniques, error representation, etc. Methods in quality assessment are also being improved cooperatively with independent software developers.

TASK 5: DISTRIBUTION AND PEER REVIEW

The three levels of distribution and review are described in the main body of the User's Guide, along with a diagram of the annual cycle (page 12 and 13).

Pre-release reviewers (approximately 3-5), selected for each new data-set on an ad hoc basis, serve the function of ensuring adequate quality and appropriate representation of new data-sets and the integration process. A team of Primary Reviewers (approximately 100 in 1991, of which 38 responded), maintained for the duration of the project, serves the function of reviewing the integrated database and project methodology. A number of issues are addressed in the review, including:

1. Technical quality and completeness of data and documentation
2. Database design
3. GIS functionality
4. Scientific content

5. Potential applications
6. Recommended improvements
7. Overall methodology

The review concept, however, goes farther. As "reviewers" use the database (and GIS methods) in their research, they are also participating in an experiment in scientific networking which may have broader implications for future collaboration. Each scientific discipline must achieve a certain "critical mass" in establishing useful methodological conventions that allow scientific exchange to proceed smoothly and quickly between colleagues. Global Change, as a field that crosses traditional disciplinary and institutional boundaries, requires an unprecedented level of cooperation and exchange to address the key scientific questions. Efforts such as these may provide viable technological and methodological approaches to this greatest of problems, as well as the increasing dependence on verifiable multiple disciplinary data.

The importance of review is especially obvious in contrasting "ecosystems" data to better defined physical data historically handled by the World Data Centers. Remotely sensed data are less well defined, and thus more subject to interpretation. At still another stage removed, however, are derived satellite products such as the Global Vegetation Index, produced from NOAA polar-orbiting satellite AVHRR data. For such derived data, precise definition is nearly impossible except on an application-specific, empirical basis. Many of the ancillary data-sets included in this project go even one step farther, being subjective classifications based on individual scientists' work. Nevertheless, such derived, and even synthetic data are essential in providing the variables needed to characterize systems, even though their value can only be determined in the context of experimentation. Thus, unlike many other fields where an individual expert can effectively represent the scientific community, global change is too diverse and broadly defined to be served so easily. A reasonable approach, adopted by many multi-disciplinary projects, is to bring a sufficient number of experts together with a common framework that can facilitate consensus. The current effort attempts to do this through a common methodological framework and review process.

TASK 6: DEVELOPMENT AND IMPROVEMENT

Part of the review effort also involves contact with the original investigators, which can often result in improved documentation and/or new data. In addition, various in-house projects may be conducted to produce new or improved data-sets from available sources. Some examples are the production of monthly GVI data from weekly and bi-weekly data, integrated topography data, and new versions of boundary data. In 1991, four external contracts were issued under the NGDC Global Change Database Program for data development and improvement.

International efforts to enhance the database were initiated during the IGBP Global Change Database Pilot Project for Africa (known then as the "Diskette Project"), which was a joint effort by NGDC and the ICSU Panel on World Data Centres (WDC). Plans are being developed to expand that effort through collaboration between the ICSU Panel on World Data Centres and the IGBP Data and Information System office in Paris (IGBP-DIS): NGDC participates in this effort through its affiliated World Data Center-A.

Where there are suitable facilities and interest, the World Data Centres help implement the database and GIS capabilities along with data exchange agreements. This is currently taking place with institutions in China, South America, and Africa.

The data integration effort will become bi-directional in 1992. First, there is a need to respond to the on-going specific data requirements of the ERL-C characterization and modeling efforts to provide needed input data for research (Tasks 1 and 2). Second, the FY92 work will begin integrating some of the outputs of the ERL-C research in the form of derived or predicted numerical sets. Since these numerical sets are intimately connected with the models that produced them, documentation will need to be extended to include adequate representation of the models themselves (Task 4). This may be done in several ways. First, traditional written descriptions will be included in the documentation as provided by ERL-C. Second, NGDC will develop ways of including the operational models themselves in the distribution products, integrated, if possible, with the basic access and analysis software provided with the CD-ROMs, with links to third-party software as needed.

SCIENTIFIC SUPPORT AND METHODOLOGY

TASK 7: CHARACTERIZATION METHODOLOGY

The key to understanding the role of an adaptive, integrated database in environmental and ecological characterization (as defined in the Executive Summary) is to distinguish between description, which GIS methods are presently well suited for, and theory, which is largely the realm of dynamic modeling. This does not overlook the fact that theory is required for description, and vice-versa.

Characterization applies existing theory, models, and data (including flows and rates) for the purpose of empirical analysis and description. It may include the development of indices and predictions, as well as descriptive models of critical processes that determine the time-dependent nature of system function and behavior. (Watson, 1978)

The advantage of GIS is that it seems optimal for descriptive analysis (including quality assessment, exploratory analysis, statistical comparisons, error analysis, etc.) and static derivations or data synthesis (e.g. overlay operations, distance analysis, interpolation, and even complex derivations of predictive indices). It may also be well suited for statistical modeling, i.e. the search for underlying patterns and trends in the database, which may be used as a basis for prediction. Because it optimizes for descriptive analysis, however, the GIS approach may not be optimal for dynamic simulation and theoretical prediction itself, which typically employs a different philosophical approach, emphasizing mathematical and statistical formulation of theory rather than the analysis of data objects.

The GIS-database approach can thus be designed to support characterization work. This approach does not attempt to represent the mathematical form of processes themselves, but rather the observed, derived, or predicted results of such models, i.e. model runs based on known states (in numerical and geographical form). Naturally, this approach allows for linkage to other mathematical or statistical modeling sub-systems, or to narrative descriptions of processes; but nevertheless remains a distinct activity of its own. Unlike computer systems that are designed to implement formula objects (i.e. simulation models), GIS typically deals with formula objects in a transitory and piece-wise manner, using known relations to calculate results in simplified steps. This approach is ideal for exploratory analysis and data development, since each step can be independently confirmed before proceeding. In this approach, the results of applied models and data derivations may be added to the database, as derived (but confirmed) "data."

This philosophy of GIS methods also has equipment implications. Being more data-intensive than computation-intensive, the approach lends itself well to single-processor computers, even micro-computers, with sufficient disk storage (improved performance requires faster or multiple I/O channels to the database, rather than faster or multiple processing of limited data inputs). This implies that not only are the two methodologies distinct (in purpose and character), but that the best overall implementation may be to link separately optimized systems.

Nevertheless, the characterization database must be driven by conceptual models of the system, whether or not it represents them in mathematical form. These models, derived from research, indicate critical processes and important phenomena, and thus determine what variables should be represented, and perhaps in what form (i.e. scale, time-step, tessellation, precision and accuracy, etc.).

TASK 8: SCALE INTEGRATION AND LINKAGE

It is well known that environmental and ecological processes are a function of scale (Rosswall, Woodmansee, and Riser, 1988); and accordingly, scale is an important defining factor for this project. The scope of the project is largely defined by scale. It was shown earlier in the User's Guide (Data Acquisition and Development Priorities, Figure 1, Page 7), that the database defined by this project fills a gap in scale between site-specific studies and global simulation modeling. As such, it can serve as a link between these other two well-established activities, scaling up from site studies and scaling down from coarse global models. Furthermore, the range of scales defined by the project encompasses interesting phenomena in its own right, and may serve well as the scale for communication between studies and with various global change outreach efforts.

Assuming that there are appropriate ranges of scales for representing given phenomena, the issue of scale integration becomes critical to global characterization. When is it appropriate to develop correspondences between data of different scale and when is it not? Since any study can define its "natural" scale boundaries, as in Figure 1 (Page 7), this question reduces to one of determining the boundaries of natural scale groups, for a given field of study. It is useful and necessary to attempt scale conversions and correspondences between data within the scale boundaries of a given study (e.g. using cover class data at 1/2 degree to help analyze satellite data on 10-minute grids). Between such scale groups, however, it may not be so reasonable (e.g., there is often no direct correspondence between land cover classes at widely differing scales because they may be based on different properties).

Instead, linkage of information between such scale groups must rely on linking the analytical results of complete studies, not data. Such linkage is accomplished by predicting phenomena at different scales, from analysis within a given scale range. For example, it may be difficult or impossible to predict distributions of non-dominant species based on GCM predictions on 2.5 to 5 degree grids (because they may respond more to the micro environment and biological interactions than to general conditions). However, it would be a reasonable approach to predict climatic changes on this scale, then modify the prediction with finer scale data, for example elevation and soils data, and then apply the prediction to species models at site scales. Similarly, it is a reasonable approach to extrapolate biogeochemical emissions from site scales to medium scales, using appropriate data on the geographic extent of similar sites. These may then be aggregated to produce the required coarse inputs to GCM models.

For both scaling up and scaling down, the "medium" scale of this database becomes important. On the one hand, we need aggregation methods to characterize (or "parameterize") variables at scales required by models, from finer scale states and processes. On the other hand, we need to understand variability of coarse scale phenomena due to more local conditions, to apply the predictions of global models. The same issue exists in the time dimension. Establishing methods for characterization across scales is a prerequisite for linking modeling with observational data.

TASK 9: APPLICATIONS TESTING AND MODEL LINKAGE

A major goal of the characterization database (besides supporting monitoring and various forms of synthesis) is to support research and modeling in global change. As mentioned above, the concept of ecological characterization, differs from process and simulation (i.e. dynamic) modeling in that it is focused entirely on synthesis of existing data and process information. Dynamic modeling, on the other hand, involves a theoretical representation of processes or system functions themselves, usually in mathematical form, and usually designed for research purposes. Some other common uses of the term "modeling" include statistical modeling, which is the search for underlying patterns that accurately describe data, and GIS or "data" modeling, which is really data synthesis -- the production of a pseudo data-set from the static combination and analysis of others. These latter uses of the term also differ from dynamic modeling because they are descriptive and do not use dynamic relationships. Perhaps the clearest distinction is that dynamic modeling (which may include probability models) claims to be valid in space and time beyond the range of available data, whereas other forms of modeling are only tenuously extrapolated beyond or between the available data, with increasing statistical error. This comparison is shown in Figure 1 below (Kineman, 1993).

The goal of adaptive environmental (and ecological) characterization using GIS technology, is to apply accepted integration and synthesis methods to data-sets to provide empirically valid representations of the natural system, given a variety of applications. Though guided by modeling efforts, it must remain entirely complementary to them (i.e. as independent as possible to provide new information and valid tests), sharing much with the field of *comparative ecosystems analysis* (Cole, Lovett and Findlay, 1991).

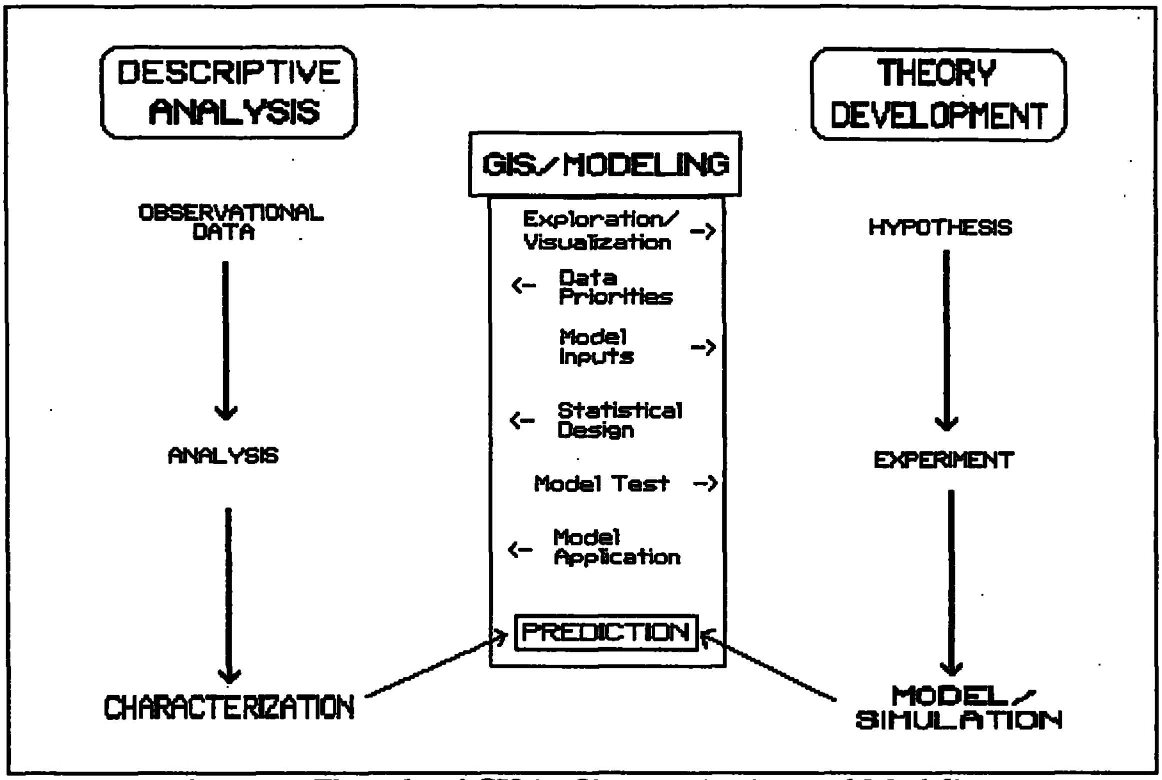


Figure 1: The role of GIS in Characterization and Modeling

As a goal-directed activity, this approach can ideally support monitoring, assessment, prediction, model inputs and tests, exploration, education, resource studies, environmental management and planning, and other information applications.

EXTENSION TO EDUCATION, OUTREACH, AND POLICY SUPPORT

While the Global Ecosystems Project itself is focused on establishing and improving the scientific database, other activities of the NGDC/WDC Global Change Database Program are concerned with public dissemination. This includes cooperative projects between the World Data Center (housed at NGDC) and the IGBP, as begun with the Africa Pilot Project, database support and consulting to the International Space Year's Global Change Encyclopedia, data planning committees such as CODATA, and support to various educational outreach projects, including United Nations training programs and developers of educational materials.

The current database is experimental and not well suited to teach global change phenomena without considerable analysis and interpretation. Nevertheless, education and outreach need not wait for full scientific development, if the education/outreach program is appropriately tuned to the existing level of development and knowledge.

For example, it may be appropriate to teach and extend expertise in GIS and data management for global change using this database, if it is clear that the effort is targeted to appropriate experts and applied scientists. Similarly, at the level of technical development, the issues of statistical designs, quality assessment, comparative techniques, etc., are relevant for curricula. At the level of scientific support, it may be appropriate to explore concepts of characterization, scale, and modeling, extending results to other scientific groups, as well as developing corresponding education programs for global change phenomena, if the effort is appropriately combined with expert knowledge. It would not be appropriate, however, to use the data contained here as definitive information, without considerable study and interpretation.

Although not a specific goal of the project, eventual application to policy represents a fourth level of development that relies on prior development of all three preceding levels (Database Integration and Design, Technical Development, and Scientific Support and Methodology). As such, it is clear that by establishing a credible scientific effort in global characterization, we are also establishing a methodical way to support policy and avoid the inappropriate transfer of raw data to information levels.

REFERENCES

Cole, J., G. Lovett, and S. Findlay (eds.). 1991. *Comparative Analysis of Ecosystems: Patterns, Mechanisms, and Theories*. New York: Springer-Verlag.

Kineman, J.J. 1993. "What is a scientific database? Design considerations for Global Characterization in the NOAA-EPA Global Ecosystems Database Project." In: *GIS and Modeling: Proceedings of the First International Workshop on Integrating GIS with Environmental Modeling, Boulder, CO., September, 1991*. London: Oxford University Press. [In press]

Rosswall, T. R.G. Woodmansee, and P.G. Riser (eds.). 1988. *Scales and Global Change: Spatial and Temporal Variability in Biospheric and Geospheric Processes*. SCOPE 35. New York: J. Wiley. 355p.

Watson, J.F. 1978. "Ecological characterization of the coastal ecosystems of the United States and its territories." *Proceedings: Energy/Environment '78*. Los Angeles: Society of Petroleum Industry Biologists. pp. 47-53.
(Also subsequent publications of the Coastal Ecosystems Project, Office of Biological Services, Fish and Wildlife Service, U.S. Dept. of the Interior, Washington, D.C. 20240.)

APPENDIX B -- REVIEWER'S COMMENTS

John J. Kineman and Jeffrey D. Colby (eds.)
Global Ecosystems Database Project
National Geophysical Data Center
Boulder, CO

EDITOR'S NOTE: The following collation of review responses has been minimally edited to preserve the anonymity of the reviewers. Personal communications and specific editorial corrections for the documentation were also deleted, however they were taken into account in the re-design and production of the current version of the database. The reviews are organized according to questions on the response form that was distributed with Prototype 1 of the CD-ROM, plus any additional comments provided by the reviewer. The numbers refer to the reviewer, so that a given reviewer's comments can be matched between questions (sometimes a comment will make reference to answers in the previous questions).

I. WHAT ARE YOUR COMMENTS ON THE OVERALL USEFULNESS OF THE DATABASE EFFORTS (E.G., IS THE DATABASE HELPING YOU IN YOUR RESEARCH ACTIVITIES, IS THE PUBLISHING OF AN INTEGRATED GEOGRAPHIC DATABASE USEFUL, ETC.)?

1) N/A
2) N/A
3) First, I want to congratulate you and your team with the compilation of such impressive data collection. My first feeling however, was a slight disappointment. Most databases included are databases that I already had checked, reviewed, and/or used in one stage or another of my research. Then I realized how much time it had taken me to locate data-sets, read and reformat tapes, incorporate them into my own systems, present the data on maps and finally evaluate the data. This compilation is an enormous step forward for global (change) research. The current IDRISI format allows any user to look at the data in a glimpse, without much knowledge of programming, data structure etc. I think that this is the most important achievement of this data compilation. Although it is far from ideal, it will surely ameliorate the use of global data-sets, lead to an improved data-use in global (change) research and eventually facilitate and promote the creation and development of better and more comprehensive global data-sets.

   I would like to see an exercise session included in the manual. Probably most users will not have much experience in GIS-technology and database techniques. A step-wise approach to guide the user to mapping, displaying and overlaying routines will promote a better understanding of the data-sets and lead to more comprehensive applications. I have seen such a manual together with the IGBP-DIS [IGBP Data and Information System] pilot project "Global Change Database" and found it very useful. The preparation of such a manual should not be such large effort, because you were involved in the creation of the IGBP-DIS data-set.
4) Well organized, good data structure.
5) Someone once said "people ask for criticism when they really want compliments"; the

authors of this database thoroughly deserve compliments!! It's amazing! Of course, I'd like several other factors to be included in the database, but you'd have to be brain dead not to be impressed by the potential of this concept. You'd also have to be completely lacking in appreciation of beauty not to be fascinated by the global NDVI images. The images are intriguing enough in themselves, but they certainly are also very useful.

With the African pilot project database I was able to map out a modified version of the Koppen classification. I could see how it compared with the standard Koppen and see that some locations were unclassified by the modified system. I could also look at the NDVI images to see if areas which were supposedly drought-free really were at particular times of the year. I'll be using my programs and the global database to do similar analyses for other areas of the world.

I purchased a CD-ROM reader just to see this database and I'm very pleased with the result.

6) I did not receive the database until a few months ago and have only now been able to begin using the data. From this limited experience it is apparent that this database will be very useful in our efforts to model trace gas emissions from soils and vegetation. Although many of these data-sets are already available to us ..., it is more convenient to have the data on this CD-ROM.

7) .....I don't think that the full usefulness of this data will be known for some time. In the near term, I know of several people who are looking for this type of data, even though it is not for ecological modelling efforts. For my current interests, the elevation data is the most interesting and useful.

8) Highly useful as an integrated package. We put together a "home-grown" version of this type of global coverage, and it is useful to compare documentation and interpretations.

9) I am part of a larger ... Global Change program. The database is not helping me personally, as I am working on global aspects of carbon cycling and carbon cycling modeling. However, an integrated geographic database will be useful to our overall program which is working on modeling sectors of the U.S. including Alaska, Hawaii, and Puerto Rico.

10) This CD-ROM is just the type of database we need for teaching in our undergraduate and MSc level courses, and for some of the GIS research projects we would like to be involved with on a regional/global scale. The completeness and common basis of the raster data simplifies our tasks immensely.

11) This is an outstanding compilation of essential information for climate studies. I plan to make extensive use of it.

12) The database has been very valuable to me, in allowing me access to the Holdridge Life Zone approach description of the climate-vegetation relationship. The Holdridge approach seems to be the most readily available means of projecting vegetation patterns under global warming. Examining the included Holdridge layer helped me understand the advantage and drawbacks of using the Holdridge Life Zone approach for our study of biodiversity under global warming in the Pacific Northwest of the United States. A widely distributed, integrated database makes a lot of sense. It should allow workers from diverse discipline, institutions, and continents to compare results. I do have some misgivings about the global nature of the database though. The inevitable coarse resolution (at least of input data) will limit the usefulness for some regional and smaller scale studies pertaining to global change. I think

consideration should be given to parallel efforts to produce national or bioregional databases modelled on GED, but at higher resolution.

13) As addressed in detail later, the present usefulness is minimal because of unacceptable access problems. The database is potentially very useful for scenario development and GCM evaluation research of [our program]. It also has potential applications for [our environmental monitoring, forest, and assessment programs].
14) The GED database is a useful product. All data are stored on the same medium (CD-ROM) with standard file and record formats. As the quantity of geophysical data continues to grow, it is important to store the data in standard formats and on easily accessed media such as optical disks. Since we did not use GIS software to access or process the data, I cannot comment on linkages between the GED database and GIS. We did not, however, find any of the IDRISI software that was provided useful.
15) The publishing of an integrated geographic database is timely with the development of systems to research and model the global impact to our environment that are occurring in this day and age. Locally, we plan to develop our own historic climatic database, and the effort by [our agency] to develop [a] database for management and modeling purposes should coincide well with the development of the Global Ecosystems database.
16) The database is quite useful, especially with the IDRISI interface. I have no complaints with the database itself. The organization of the documentation is good and the directory system and file naming conventions allow easy access to the data. I have found the Meteorological data-sets quite useful in my research.
17) N/A
18) This is an enormously valuable effort. Already this database has substantially helped my research. I foresee that this database will be almost universally used and cited by global change researchers.
19) The database is <u>exceedingly</u> useful. I've been trying to obtain the NDVI data for years without success. These are all databases which are <u>critical</u> for global-scale research. Thus, having a documented common source of data is <u>absolutely</u> necessary for exchange of ideas. Also, it is wonderful for teaching classes and may attract MS.-level students into global research for their Ph.D studies.
20) The database is helping us in our research activities.
21) I think the database efforts are very useful. A major difficulty in developing spatial models is obtaining and controlling the quality of the necessary input data to run the models, especially if one must obtain these data-sets from a variety of sources. Our simple ecosystem model... can use many of the data-sets provided on the CD. With the CD-ROM data, we can compare the effects of using different input databases (Leemans and Cramer vs Legates and Willmott) on model results. The public access of this integrated geographic database should encourage additional modeling efforts by minimizing the time, money, and effort needed to obtain these initial data-sets.

    Another benefit of the database effort is the possible development of a "standardized" input data-set. It is often difficult to compare spatial models because the models have used different input data-sets. By providing a "standardized" input data-set that is easily accessible to various modelers, models can then be compared based on differences of model interpretations of reality.
22) To the kind of work I am personally involved with, this database will be almost essential, especially if updates can be provided later.
23) It would be nice if every field were on the same grid. It would also be nice if the

data were contemporaneous, i.e., January 1988, etc., for all fields.

24) 1. Extremely useful collection of information.
    2. Each subsequent release should strive to greatly increase the amount of information included.

25) N/A

26) Very useful.

27) Yes, the efforts are useful, 3 years ago I was looking for data of this format to do a project with a quick turn around time and it was impossible, your database should reduce the waste of time associated with locating & working with large-scale environmental data.

    [From cover letter:] The NOAA/NGDC Global Ecosystems Database is a useful and welcome package. The data-sets included are certainly of value to research in global ecosystems. Some of the data-sets are ones we planned to obtain from other sources (e.g., Zobler soils, World Data Bank II, Elevation and Bathymetry). Documentation of the data-sets is good, both in content and format. In fact, we could make real use of the documentation even if we had to get the data from other sources. The inclusion of published documentation is quite useful.

    Distribution on CD-ROM is a nice touch and useful. Organization of the CD is good. We operate in a UNIX workstation environment (DECs and SUNs), and we had no real trouble accessing the data files. I would have much preferred the actual data files be in the "lowest common denominator" of ASCII format rather than binary. However, with some work we should be able to access the binary data through GRASS (the GRASS appendix is helpful, but could give a bit more detail) or some other UNIX tool. The documentation of the binary format is of help here. Still, I would have been happier, and the data more useful faster, if they had been presented in ASCII.

    The database will be useful in our research on global terrestrial vegetation. We would like to use it, and we look forward to the first real release.

28) The publication of this type of data is essential for many applications within the EPA. It provides a standard set of information on which further research can ba based. My major complaint concerns the accessibility to the GIS community. Although the PC based display programs worked... most scientific research using GIS is not done on PC platforms. When trying to determine how to convert the data to other formats for use in UNIX based GIS systems the documentation was inadequate. The database will only be useful if the data [are] provided in a useful format.

29) The Global Ecosystems Database is undoubtedly useful in our research activities. Its further development and increase in the number of data-sets (parameters) will make it possible to widely use it in the field of hydrometeorology and ecology.

30) I believe that the publishing on CD-ROM in this way is very useful as a means of providing a baseline set of surface data. It helps to make the comparison of Earth observation data products between instruments and missions a realistic possibility.

31) Overall a very useful data-set; in particular inclusion of multi-temporal satellite data provides excellent opportunity to study global scale interaction between climate and vegetation dynamics.

32) N/A

33) N/A

34) 1. Selection of contents - The selection of databases - topography, temperature, precipitation, vegetation index, vegetation types, soil type -- seem to be good.

[comment moved to 5]

2. Specification of grid type - Gridded data-sets can be classified in two types .... One type may be called "grid point type" and the other, "grid box type". The former conceptually consists of point estimates at the cross points of x- and y-grid lines. The latter consists of some statistics, such as mode, median or mean, within boxes surrounded by grid lines. In addition, both types can be placed in two ways according to the position of grid lines. Usually grid lines are placed at integer multiples of the grid interval (or even number multiples of half the grid interval). Sometimes, however, grid lines can be placed at odd number multiples of half the grid interval, in order to match central point of "grid point type" data with that of "grid box data"....In general, the actual resolution of information is coarser than half grid interval, and the difference of grid types maybe irrelevant for processing of individual data-sets. When making comparative studies, however, matching of grids is more serious. When making a database like this, there can be a choice. One can interpolate grid point type data to "odd" positions so that they coincide to center position of grid boxes. [Editor's note: The reader is referred to the explanation of registration and window coordinates in the User's Guide, for the conventions and definitions adopted in Version 1 of the database].
3. 10-min. data-sets - I think it is good to include the data-sets "expanded" to have the same spatial resolution as others. However, following points should be taken into account.

   The procedure for making 10-min. grids from coarser grids should be documented. In the prototype manual, it is only documented for (3) only. For other data-sets, it is just noted that they are derived by IDRISI program. At least, whether the value is just copied from the nearest grid point or interpolated from several grid points should be explained.

   The original version should be included. For Legates and Willmott climate data (3), all the information can be incorporated in 0.5 degree version with smaller data amount than 10 min. version. [Editor's note: this would require either interpolation to "even" grid-box registration, or original data on "odd" convention and thus incompatible with the rest of the database. Expanded grids are no longer provided for software convenience, but only as needed to allow data intercomparability within the "nested-grid" convention.]

[Editor's note: comments about packed format omitted - packing is no longer used on the disc]

35) N/A
36) N/A
37) N/A

IA. WHAT SOFTWARE SYSTEMS DO YOU FIND MOST USEFUL WITH THE DATABASE?

1) N/A
2) GRASS running on a Sun workstation. Could you link it to GMT either in xyz format or grd format (netCDF) files?
3) I think IDRISI is a good choice to present and use the data-sets. PCs are cheap and quite common right now and IDRISI allows for immediate use of the data by

colleges, universities and research institutes all over the world, including third world countries. However, there is a disadvantage to limit the data availability only to this specific computer platform. The research community especially wants to link global data to models and other large applications. PCs are less suited for such tasks due to memory and speed limitations. Although IDRISI is capable to export data to other formats, one still needs a PC to do such conversion -- [and a] program to read the data immediately and correctly from the CD-ROM (the compatibility with UNIX was no problem as stated in the introduction on page 3). To improve data exchange there surely would be a need of some conversion programs to read IDRISI data files and transform them to ASCII or binaries. The source-codes of these programs (Fortran, C, Pascal) should be provided so that they could be compiled and used on any computer system. This will immediately allow for the use of the global data on other computer platforms and therefore enhance compatibility and give the CD-ROM a larger audience.

4) A good layout. Looks to be very compatible or accessible to whatever the end user has or wishes to use.
5) I mainly use IDRISI and my own programs. I've been very impressed with IDRISI, which I hadn't seen before the African pilot project. I like the modularity of the system and the simple structure of the data files.
6) IDRISI
7) Clearly, the IDRISI GIS is most useful for an initial look at the data. In my case, I will be using the data with the programs I write.
8) GRASS 4.0, NCSA Datascope, XImage
9) We used the IDRISI program that was furnished to access the database.
10) We use a variety of proprietary raster handling GIS, mostly SPANS, but also IDRISI (very old version), and some vector types such as LaserScan's Horizon and Arc-Info. In addition we write our own on an experimental basis - usually raster - to provide functionality not readily obtainable from the proprietary ones. We have used GED with success with IDRISI and SPANS, and have imported it into some of our own half-baked (in the UNIX sense!) products.
11) We are developing our own to do hydrologic modeling. We also plan to make some use of a general-purpose GIS but haven't chosen one yet.
12) IDRISI. I had some difficulty attempting to transfer information from IDRISI to ARC/INFO for plotting.
13) Since the on-site reviewers were unable to successfully port any of the data in a form accessible for any of our software packages, I cannot answer this question. If access problems can be overcome, we plan to use the database in an ARC/INFO environment.
14) N/A
15) We were not able to access or view the data or structure due to lack of availability at this time of a CD-ROM drive. We plan to purchase one soon.
16) The IDRISI system was useful in the preparation of the data for further analysis. Unfortunately, it was a bit limited when it came to obtaining hardcopies of the displays.
17) In this review of the database I used IDRISI to examine the data-sets. I also attempted without much success, to use the PC package MIPS to analyze the data. In the future I will be using IDRISI and GRASS in my analysis of the data.
18) I've imported all external databases into the BIOGIS system of Leemans and Cramer,

which was tailor made for our work.
19) N/A
20) MS-DOS/IDRISI
21) Most of our modeling efforts and analyses have been PC-based. Therefore, IDRISI has been useful for displaying the data-sets. To use the data with our models, however, will require transforming the data to a different data structure in which attributes are explicitly identified with a particular longitude and latitude. Once the data is transformed into this structure, we have used dBASE-IV to manage the spatial data; we have used SPSSPC+ and PC SAS to analyze statistics of the data; and have used GRAPHER and SURFER to display the spatial data.
22) Not really applicable as I understand. This disk provides data, not software. If the question concerns the usefulness of IDRISI in this context, then my comment is positive, because I am familiar with IDRISI. I am not sure, however, how big the problems with IDRISI are for users not familiar with the system.
23) IDRISI, and we can only display files, not access them. It would be better to be able to easily transfer ascii files to our computer system of Unix and Fortran.
24) 1. A text editor
    2. IDRISI
25) ARC/INFO, because that's how we access geographic data at our site.
26) In-house software.
27) N/A
28) I don't believe the data-set should provide elaborate or even moderate GIS tools. Most researchers have tools for viewing, manipulating and analyzing spatial data and are more comfortable using their own tools. What they need is easy access to quality data and a convenient method of moving this data into their current analytical environment. Therefore I believe an extensive set of data conversion software would be an essential addition to this database.
29) IDRISI is acceptable to us.
30) We now use Suns almost exclusively: CD-ROM access is straightforward.
31) General image processing tools - we have been using IPW: Image Processing Workbench.
32) N/A
33) N/A
34) N/A
35) N/A
36) N/A
37) At present IDRISI, but we will add a UNIX system in the Fall.

IB. HOW IMPORTANT IS LINKAGE BETWEEN THE DATABASE AND GIS (AND/OR OTHER SOFTWARE)?

1) N/A
2) Very important - allows early data manipulation.
3) For global change research this link is of uttermost importance. It permits displaying, linking and overlaying the data in many ways. Further, for users from other categories, it will give display and overlay capabilities, which makes this CD-ROM already more useful then for example the first part with environmental data of *Times* atlas. Without a GIS the data would be just a dull collection of uncomprehensive

digits.

4) Important to the novice but you can't get too coupled if you wish to stay flexible to other approaches.
5) I think this is very important, particularly as your target audience includes schools and colleges as well as researchers. Many researchers would be pleased just to obtain these data in digital form, but few schools would use the data if the link between the database and the GIS wasn't good.

   I'd read that CD-ROM software was often unsatisfactory, so I've been pleasantly surprised by how smoothly IDRISI works with the database. That isn't to say a few things couldn't be improved. For example, several of the IDRISI commands were obviously designed to work well with floppy or hard disk systems and not the vast storage capacity of CD-ROMS. It would be good to be able to abort commands like LIST with cntl-C or perform selective searches, rather than to have to scroll through the entire contents of all active directories. Similarly, it might be useful to be able to abort long color jobs if they don't look interesting.
6) It is important.
7) For many researchers, this linkage can be quite significant as the raw data will be of little use unless they can display and manipulate it without having to spend a lot of time and effort writing software. However, it has been my experience that no one, or even a group, of programs can be expected to meet everyone's needs. Therefore, the availability of the raw data and reasonable documentation as to its content, format, and access is equally important.
8) Very important for our global nutrient cycling model studies.
9) We will be basing many ...information outputs on GIS. It is therefore important to us to link the database with GIS.
10) The IDRISI link has proved the easiest method to extract and display the GED. Input to SPANS enables quite complex modelling to be done. For teaching purposes the provision of a simple front end direct to IDRISI was very welcome. The only real requirement is to be able to import the data required onto disk in the correct form for the appropriate GIS. Working direct from the CD-ROM is unlikely to be a profitable route for repeat enquiries owing to the typical lack of access speed found on CD-ROM readers.
11) Not too critical, but nice to have.
12) Absolutely essential.
13) This linkage is very important for our intended applications. The existing linkage was extremely frustrating and extensive work, even using the GRASS interface, never provided successful access to the data. The data could be loaded into GRASS, but not be successfully displayed.
14) N/A
15) The linkage to GIS will have potential for our work in the future to spatially represent the data we collect/model, and also to integrate with other systems that are developing in that format. Currently, [our] GIS procurement is still in the pipe and not available to implement at this time.
16) I found the linkage between software and database essential, especially for intercomparing data-sets.
17) I think the linkage between the database and other GIS packages is very important. Global (or continental) georeferenced data is very hard to come by (or too expensive).

18) For us not at all, i.e., the important thing is that we can easily read the files.
19) Not terribly; most people have a preferred database into which the software is translated (e.g., dBASE IV; Quatro-Pro, etc.). As long as the information in the separate data-sets is transportable, the use is open.
20) The linkage is important for our research and initial stage of environmental modeling.
21) The linkage between the database and other software is very important. Without such linkages, the data-set becomes a set of pretty pictures to look at and perhaps gain some qualitative insights. With the linkages, the information in the database becomes available for many different types of analyses: data can be related more quantitatively to information in other spatial data-sets or even in other non-spatial data-sets; or the information can be used as input into models which may make the database even more valuable.
22) This is partly what I have answered above. Still, standardization of GIS has not reached the level where one standard would be satisfactory everybody. I think IDRISI provides reasonable links to the systems we use here. Essential is, that the system allows for the retrieval of raw data without deep knowledge of any one GIS system.
23) It provides a convenient way of displaying and manipulating the data. However, again we need to be able to take data-sets easily to our own fortran programs.
24) Extremely important. In fact, more of the IDRISI Modules should be added, and a text editor should be added.
25) Its most important that it be stored in a common format. Most GIS's read the most common formats.
26) Not important.
27) N/A
28) Linkage is ESSENTIAL. Refer to previous answer.
29) Linkage between the database and GIS is essential, which should not however preclude from using other software.
30) Linkage is vital, especially to GIS on which the data have been developed. This enables those who wish to, to view data as seen by the originators.
31) This is important but not critical; more important is a simple, accessible data format so that data can be imported into local systems independent of specific hardware/software.
32) N/A
33) N/A
34) N/A
35) N/A
36) N/A
37) Very - makes initial screening simple

IC. HOW IMPORTANT IS IT TO ENSURE COMPATIBILITY WITH A "LOWEST-COMMON-DENOMINATOR" GIS PACKAGE (FOR EXAMPLE, IDRISI)?

1) IDRISI provided with the package was not a complete set. I should like [to] have [a] "shared" copy if it is not too late to request. (I received this prototype copy only in Feb. 1992)
2) Very important but requires compatibility with a non-PC environment.

3) Considering the costs in obtaining, training personnel etc. of most commercial GIS-packages, limiting the global CD-ROM to those packages, would be a severe limitation. Only large research and governmental institutes could afford these packages and use the global data-sets. A broad distribution of the data is one of the aims of the release of this global data and therefore I think that IDRISI for the PC (and, if possible, GRASS for the UNIX environment) is the best choice in presenting compatibility of the data-sets. Besides IDRISI is capable to import/export data from, respectively to other GIS-formats.
4) I think it would be very important.
5) Very important! By using a system like IDRISI, which is available on PC's, you make the database available to probably two orders of magnitude more users than if you used just GRASS/UNIX or similar. I think you'll be amazed by the analyses done by school students with this database.

   Even more importantly by adopting a PC minimum configuration you make the information potentially available to students and users in the developing world, virtually none of whom have access to UNIX workstations. I say "potentially" because they also probably don't have access to CD-ROMs. It might be worth preparing some policy about circulating subsets of the information on floppies to people in developing countries. Could I extract data the final database and send it to a colleague in China or Africa? Would you want to purchase a copy of the CD even if they don't have a reader? This would be okay if the cost was reasonable.

   You might also consider preparing a one disk "sampler" to load onto bulletin boards; it would be a good advert for IDRISI as well as the database. I'd also recommend sending a copy of the CD-ROM to [various magazines], when you release the final version.
6) It is very useful for users with limited GIS experience.
7) As with 1b above, that will depend heavily on the nature of the research and the fit between the GIS and the research needs. Although compatibility with the GIS is highly desirable, I would not eliminate data just on the basis that it was not compatible with a GIS. In the long run it is the data that is important, not the GIS.
8) Important, but also for Apple Macintosh GIS software.
9) I have as yet had no other experience with GIS. When we got a feel for IDRISI it worked O.K. However, the database will need to have Readme and Help files to make it readily usable with any compatible program package that it is set up for.
10) Very important indeed! Wide dissemination of the database will be predicated on easy and cheap use. Cheap but effective GIS on low budget platforms must be a large target market.
11) I think it is a good idea.
12) This depends on how widely distributed various types of GIS packages are. If "everybody" can get their hands on IDRISI, this makes sense. If, on the other hand, all serious consumers of the database have access to ARC/INFO (for example), distributing information on the "lowest common denominator" standard probably just slows down analysis.
13) This linkage is very important for our intended applications. The existing linkage was extremely frustrating and extensive work, even using the GRASS interface, never provided successful access to the data. The data could be loaded into GRASS, but could not be successfully displayed.
14) N/A

15) We feel the lowest common denominator is the best choice to ensure compatibility across systems that are in use.
16) The compatibility issue is directly related to the usefulness of the database. IDRISI should be either packed with the database or easily accessible to the user.
17) The database either needs to be in a low common denominator (like IDRISI) or it needs to have accurate conversion routines built into it to allow the export of the data to GIS and other packages. The current incarnation is perfect for this except for a few minor missing elements (discussed below).
18) I think this is important for many users, and I believe IDRISI is the right choice.
19) This is more important; most folks who could use the data-sets in teaching, for example, are unlikely to have expensive GIS's, as are those in less-developed countries - it seems a small price to pay to maintain the IDRISI connection.
20) It's very important.
21) As a PC user and an IDRISI user, I think compatibility is very important.
22) see above
23) Zero.
24) Very important. Since that will almost guarantee compatibility with more sophisticated packages.
25) Not very.
26) Not important.
27) N/A
28) I don't support the concept of "lowest common denominator" GIS. There are now, and will be for quite some time, numerous GIS solutions available to analyze spatial data. What is needed for compatibility is a common data transfer data format for spacial data such as the Spatial Data Standards and Testing (NIST). Since this standard is still in review other commonly used spatial data transfer formats including DLG, TIFF, DXF, BIL, BIP... would be more useful to the GIS community than the IDRISI format supplied with the database.
29) N/A
30) Not sure - most GIS's seem to have foreign import options for "raw" data inputs, therefore, if the database format and structure are both simple then compatibility need not be a problem.
31) Not important; too much diversity in systems currently being used; rather, simple, accessible, well documented data storage format is critical.
32) N/A
33) N/A
34) N/A
35) N/A
36) N/A
37) Very - for initial screening
- showing data to other users, quickly

But needs file format transfer programmes -- some are in IDRISI already, others should be added.

II. ARE THE DATA ACCESSIBLE IN TERMS OF DATA STRUCTURES, FORMATS, MEDIA, PROJECTION, ETC.?

1) N/A

2) Yes - but again would like easy access on the SUN/UNIX environment.
3) I had no problem in assessing the data files, both under the DOS and UNIX environment. Most data is in a gridded format and this can be transformed to any projection. I think that the data is very accessible on this CD-ROM.
4) Yes - good job.
5) Yes
6) Yes
7) The structures and formats used for the data on this CD-ROM are about as simple and easy to use as any I have seen. Where practical, simplicity is clearly a virtue. However, good data should not be excluded simply because it has a complex structure. Complex data just needs better, or more extensive documentation. I was most pleased to see that your data was unprojected (used standard Lat/Lon units). Most people with whom I come in contact want unprojected data-sets. The quarter-inch backup tapes are currently in wider use and would also make good second choice. Given the availability of these low cost options, the traditional magnetic tape reels should be ignored in the future.
8) Yes, and I have included a procedure to access the data with GRASS 4.0. Feel free to share this with others, if it proves useful.
9) We found vector and window formats to be useful. We could have made useful printouts and used screen projection. I thought some of the data, especially on large-scale maps was questionable.
10) It is very instructive to compare access to this GED CD-ROM with the prototype issued for the world vector base (to be) provided by DMA called DCW.

    The latter uses a highly structured file system intended for direct input to ARC/INFO, and is difficult and obscure to handle and slow to run owing to the repeated accesses to multiple directories and files on the CD-ROM.

    The GED is simplistic but easy to use, fast of access, and simple to read and convert into alien formats. This is partly because it is a simple raster, but it could have been organized in some funny quad structured fashion to make life really difficult! Simplicity is always worthwhile when large volume data is being handled.

    We did not test any projection handling, plate carreé is fine form for most of our needs.
11) Yes - very important to have data in lat/lon coordinates.
12) I found extraction of the data and use by IDRISI fairly straightforward, I did have a problem maintaining geographical reference points after extracting a regional subset of a raster file. Also, the lack of geographic reference the subset coordinators complicated transfer to ARC/INFO. Also, a means of subsetting vector files (e.g., from micro World Data Bank II) would speed processing enormously.
13) No. As mentioned above, rather than a "lowest-common-denominator" approach, compatibility should be maintained through the use of standard data structures and formats. For example, there are many industry standard image file formats, such as TIFF, BIL, BIP, and run-length compressed. However, should the database remain compatible with IDRISI only, then it is strongly recommended that software be provided to the reviewers to enable conversion of the data-sets to GRASS format as well as to one/many of the industry standard image formats. Reviewers would not be expected to obtain this software on their own. It should be provided with the review package.

    Another issue is time of access. This is really a hardware issue. To make the database accessible to the largest number of people, a PC platform is desirable. Unfortunately, this provides very slow access and transfer rates. for [our] general ... applications, the PC platform is probably unacceptable.

14) N/A
15) see 1(a)
16) One frustrating aspect in regards to the usability of these data is the lack of a suitable driver for the HP Laserjet Printer. Hardcopy of the results of analysis using IDRISI are necessary. Perhaps this problem will be solved shortly. The data format allows for easy access to the data.
17) There are a number of things to be done to improve the accessibility of the data:
    A) The origin of the projection of the data needs to be given. I imported data from GOLSON.IMG into MIPS and converted it to an Alber's equal area projected image. While the image looked fine on the screen an overlay of the Great Plains states ended up in the Atlantic Ocean. It was difficult to establish proper georeferencing based solely upon the window information for the latitude and longitude.
    B) A conversion routine needs to be added that would take the images in the database and convert them to one of the standard projections (i.e., Universal Transverse Mercator or Alber's Equal Area). While there are software packages available to do this, they require the data to first be in a vector format. I know that other raster GIS packages contain routines for projecting raster data. Being able to project the data would improve the display of it on the screen.
    C) Much of my work deals with the abundance of a given species of bird at a given area. While I have a proper georeference point for my data, I could find no convenient way of identifying say, the GVI, for that data point. A VERY important (vital?) routine to add is the ability to import a list of points (in the proper projection) and then to export those points with the proper class identifier from the data. This would also greatly facilitate statistical analysis of the data.
    D) CD-ROM, even though it's slow, is probably the perfect media for this project. As CD-ROM reader prices continue to fall, the data will become more available to more people. Other methods of mass storage simply aren't standard enough to be useful.
18) Yes, as far as we have discovered.
19) Just fine - no comments of value here.
20) Yes, the data accessible.
21) Data structure - As mentioned above, to use the data with our model will require a transformation of data structures. However, this change can easily be accomplished with a quick utility program.

    Format - Personally, I like the format used to distribute the data because it emphasizes displaying the data. This format allows investigators to quickly evaluate the usefulness of the data-sets for their research needs. (Especially when used with some of the palettes provided by the CD). In addition, this format allows modelers to quickly determine sources of problems encountered during model development.

    Media - Although we did not have a CD-ROM player, I was able to borrow one to analyze the data and to transfer the data to a media (read/write optical) that we happen to use more regularly. Overall, I believe the CD-ROM is an excellent media for data distribution.

Projection - The lon/lat projection matched the projection we use for our modelling activities so the data were very accessible.

22) I expect CD-ROMS to be widely spread soon, therefore it is the most useful low-cost distribution medium. For us, any other medium would have been more costly in either investment of money or time.
23) Yes, however, user needs to be familiar with IDRISI to be able to access the files and display the images. Again, we need plain vanilla ascii data files.
24) Somewhat. I reviewed "(6) Matthews Vegetation, Land Use, and seasonal Albedo". I would have liked to output in ASCII file of the original data in the form:
    LAT LON VALUE
    . . .
25) Yes, we would prefer the data to be formatted for ARC/INFO, but then those without ARC would be out of luck, so this is a good compromise.
26) Yes - very good.
27) N/A
28) The data structure and formats used are not common spatial data formats (see 1c) and therefore the data is not easily accessible to most users in the GIS community. The CD-ROM media is a growing area and should cause no problem for researchers. Most GIS packages have projection algorithms to convert data from LAT/LONG to a variety of projections. Therefore, if the data is distributed in LAT/LONG, it should be no problem for the researcher to project the data to a desired projection.
29) Yes, they are.
30) Those examined were fine.
31) Mostly, however, in certain cases packing and coding format are not clear (see below); for the most part accessibility is good!
32) N/A
33) N/A
34) N/A
35) N/A
36) N/A
37) Yes

III. ARE THE DATA ADEQUATE GIVEN YOUR RESEARCH NEEDS?

1) N/A
2) My needs are global modeling of the fluxes of trace gases - in this case these data are very useful - especially linked together on one CD.
3) Especially the climate and topography data-sets are very useful for my research needs. These obtain continuous values and can be interpreted immediately. The vegetation and soil data-sets are of lesser use, because these obtain classified data, which is more difficult to interpret. For example: what is the difference between potential vegetation, life zones, actual vegetation and land use. For unexperienced users it will be difficult to evaluate all these classes and therefore use the data-sets in a competent way. Probably some guidance should be given in the documentation.
4) N/A
5) The data are useful, but researchers always want more and better data.
6) Yes
7) The coarse grain (5 and 10 minute) elevation data is good for a lot of general area

work. However, it lacks the detail necessary for more precise work in limited areas. The one minute, three minute and 30 second topography data described on page 12 of the National Geophysical Data Center Solid Earth Geophysics catalog (January 1987) is highly sought after by the people with whom I am in contact. An additional CD-ROM with this data would be of considerable value and highly appreciated by a wide audience.

8) Not entirely.
9) We need to look at the data more closely to see if they can be used as a reference or base to build on.
10) Our research needs tend to be regional and continental rather than global. The GED is adequate but we are greedy.
11) I now have access to important data-set I expected would be hard to find. For now this is fine.
12) Inasmuch as our study concerns plants and animals with narrow distribution within the Pacific Northwest, in general the spatial resolution of the data is too coarse for our purposes. The Holdridge life zone data are probably adequate, given that the current state of climate change predictions does not allow any finer resolution of potential future climates.
13) I am unable to respond in detail to this question because of an inability to access the data. On paper, data-sets such as the Global Vegetation Index, Monthly Temperature and Monthly Precipitation would be valuable as historical data for use by [our team]. Also, the thematic data-sets would be valuable to [us] as reference data.
14) Not entirely. For our map of the Mid-West (see Fig. 1, attached), the level of detail of the geo-political boundary (vector) data were found to be adequate. For the Northeast (Fig. 2, attached), however, the level of detail was found to be inadequate. Important geographic features such as Martha's Vineyard and Nantucket islands are not well resolved. Other features, such as islands off the coast of Maine and the Connecticut River are missing entirely. Thus, the boundary data cannot be used to produce detailed maps for sub-continental sized regions.
15) There is only one data-set that we could find in the global climate database that would be useful, that is, the daily data compiled by Mueller (1982) based on the UK Meteorological Office series (1966-1983) and the World Survey Climatology (Landsberg, 1969-1981) with some other sources inspected as well. We could not find the locations of the 1057 stations that might tell us if we could use this particular data, but this is the type and frequency of data that we would use in our modeling work, and to build our database, compare data, or participate in the global database effort.
16) It would be nice if the raw data, i.e., individual data points used in the construction of these data-sets, were accessible. It is understandable that these data could not be widely distributed due to space requirements. However, they should be available on a special order basis.

    I would like to see satellite Geostationary (IR and visual) and polar orbiting data available on Optical Disk.
17) My research deals with scales of approximately .5 degrees on a side. As such, most of the data in this database is very useful. I am concerned however about the mixtures of scales within a given data-set. I found that the documentation was not very clear when this had occurred. As an example, The Olson World Ecosystem is given as having a 10 minute scale. In reality most of the data (with the exception of

Africa) has a resolution of only 30 minutes and then it was converted to the 10 minute scale. I might be better to include the original data and a second set of the higher resolution data where available. (In this particular case having 10 minute data for North America would be fantastic!). Regardless, it should be made very clear in the documentation that there are mixed scales in the image.

One of the benefits (and one of the detriments as well) of the data is the scope of its coverage. It would be very useful to myself and other researchers to create a disk containing primarily North American data. As much as I would like to work globally, the data I need are just not available. While I can subset the data down to North America I think that a great many useful data-sets may have been overlooked (both North American and European) because they were not global in origin. I realize that the scope of this project was to create a global database and you have done a very good job there. Perhaps you could either a) create a North America (or western hemisphere) specific data-set, or b) include other useful data-sets that may not be global in coverage (perhaps in a separate subdirectory). Those data-sets that were not global in origin could still be overlaid with images from the other data-sets as long as they had been properly georeferenced and the proper window had been specified.

A priority for improvement would be to ensure that all of the maps were comparatively georeferenced at all scales. Another priority would be to try and get those images which are mixed in scale (i.e. GOLSON.IMG) finished at the same scale of resolution. It would be very worthwhile to have finer scale data available. Finally, I think that there needs to be some improvement in the acquisition software (as specified in the paragraphs above).

18) They are very useful, yet there are many desirable improvements.
19) The data available are excellent.
20) Yes
21) Our model uses data at various resolutions (although all input data are gridded to 0.5 degrees lon/lat) and produces predictions at the .05 degree lon/lat resolution.
22) Global modelling never will have the "adequate data". I believe, however, that several data-sets on this disk are extremely useful for this purpose and, in fact, hardly available anywhere else.
23) No, operational analyses would also be useful, OLR would also be useful.
24) No
25) Yes
26) Some. (see enclosed notes)
27) N/A
28) N/A
29) They are, partially.
30) N/A
31) Yes, we have found this data-set to be quite interesting. In particular, we have found the database to be highly useful for explanatory analysis of relationships and interactions between vegetation and climate; we are also pursuing research related to map-guided detection of deforestation.
32) N/A
33) N/A
34) N/A
35) N/A

36) N/A
37) Provides a first-order analysis to get started

IIIA. WHAT ARE THE LIMITATIONS OF THE DATA?

1) N/A
2) Available on PC only.
3) There is too little guidance in the documentation how to interpret and use the data-sets. Climate, topography and coastlines are relatively obvious in what they display, but class data, like vegetation and soil characteristics, are more difficult to interpret. More guidance for users on how to use the different data-sets would be welcome.
4) N/A
5) The coverage of some areas of the climatic interpolations could be improved. For example, Leemans and Cramer (1990) note that their interpolations have few data for Australia, south-east Asia and some other regions.

   The International Soil Science Society is aware of limitations of the FAO/UNESCO soil data, especially for use with simulation models, so their SOTER (soils and digital terrain database) project is aimed at improving global and regional/continental coverage (more details from: V.W.P van Engelen, International Soil Reference and Information Centre, PO Box 353, 6700 AJ Wageningen, The Netherlands). However, the FAO/UNESCO data will remain the only global coverage for a long time.

   I believe UNEP have digitized data for the more detailed (5000?) FAO/UNESCO map units (they kindly supplied me with these data for Africa), rather than the 133 soil units given in the database (e.g. Lc62-2bc would appear as just Lc in the database). The former data provide detailed description of soil conditions.

   Box, E.O., Holben, B.N. and Kalb, V. (1989) Accuracy of the AVHRR Vegetation Index as a predictor of biomass, primary productivity and net CO^2 flux, *Vegetatio*, 80, 71-89. Provides a very good introduction to the pros and cons of using the NDVI for ecological analysis.
6) N/A
7) As mentioned in the previous paragraph, the elevation data does not have sufficient detail for many applications. In my regular job I use Defense Mapping Agency CD-ROM based DTED on a PC. Your more detailed data nicely fills the gap between what is on this CD-ROM and that which is provided by DMA. For many applications the DMA data is too detailed as well as not being generally available to the public.
8) 1. For the purpose of coarse-scale global change studies, all data-sets should be provided at the 60 minute grid cell resolution.

   2. Soil carbon and nitrogen data-sets (Post et al., 1985, Nature 317:613-6) would be useful additional layers.

   3. The gmveg data-sets (especially cultivation intensity) is probably badly outdated. A new assessment based on global coverage (Landsat) imagery would be much preferred.
9) Probably, the most serious limitations are the broadness of the scales.
10) Our wish list would definitely like to see all data collected on 2 minute (as with the FAO), or 5 minute at worst. But we will happily take what we can get. We are worried about the problems of nesting because of the scale implications, but would rather have non-nesting stuff as to more detailed level than pure data at 1:10m.

11) I need a finer resolution database. Orographic precipitation modelling requires 1km - 2km (1min or less resolution).
12) Primarily spatial resolution (see above). The means of combining life zones for the aggregated Holdridge life zones data-set was not specified. This limited usefulness of that data-set.
13) N/A
14) N/A
15) N/A
16) The analyzed products are only as good as the raw data used to produce them. I was happy to see that the error analyses produced by Legates and Willmott was included.
17) N/A
18) Ground-based climate data as they stand are rather problematic - they have uneven data coverage and quality, and don't include e.g., radiation and extreme event. The NDVI period covered is fairly short (compared with e.g., the climate data). The vegetation data-sets have numerous limitations e.g., doubt about many of the sources, doubt about the method used by Matthews to infer natural vegetation, and chaotic documentation by Olson.
19) None
20) N/A
21) We have experienced tremendous problems using Matthews vegetation data-set (1° lon/lat resolution) in conjunction with Legates and Willmott (gridded to 0.5° lon/lat). We believe the source of the problem is in the difference of resolutions of the two data-sets. When we use vegetation data-sets digitized to 0.5° lon/lat we have not had nearly as many problems.

    Soil texture is another input into our model. We have been using texture from the FAO map gridded to 0.5° lon/lat. Your Global Database only has Zobler gridded to 1° lon/lat. This resolution difference might also be a problem for the reasons described in the previous paragraph.
22) This would be major topic of study. Some short comments:

    the climate data fields (2) and (3) are all nice, "smooth" data fields, which obviously reflect the true conditions very differently in different regions (due to shortages of input data). The user has no chances to estimate the spatial pattern in the reliability of the data. More important, the resampling procedure from the Legates and Willmott grid to the standard grids in this database, seem to be a dubious and "dangerous" procedure to me. Ideally, the gridding from original raw weather station data to whatever desired output grid should be performed in one single step (see pt.4 for further comments on this).

    none of these data-sets can ever be expected to be "final", but some of the limitations in the Olson data-set (4) can be derived indirectly from the upgrade documentation in this volume.
23) Only a few meteorological parameters are on disk.
24) I would like at least three times the number of categories for the albedo data.
25) N/A
26) Resolution - often too coarse. Patchy data (large blocks w/same value)
27) N/A
28) N/A
29) The data application is limited by the scale of time averaging, the inadequate amount

of geophysical parameters, included in the database.

30) N/A

31) Obviously, a larger time series of satellite data would be desirable. Also, precipitation and temperature data contemporaneous with satellite data would be very useful. Consistent spatial resolution would be useful (e.g., use UTM projection)

32) N/A

33) N/A

34) N/A

35) N/A

36) N/A

37) #2 Soils -- need a water-holding capacity and other quality interpolations - such as the GISS GCM routine

#1 Climate -- monthly time series and a consistent averaging period required -- IIASA data is length of record

IIIB. WHAT ARE THE PRIORITIES FOR IMPROVEMENTS?

1) N/A
2) Access on other platforms in other formats.
3) The documentation could be improved considerably by adding summaries, scope and limitations of each data-set to the "data-set description" pages. I believe that this should have the highest priority.
4) N/A
5) Improvements in coverages of temperature and rainfall interpolations. Inclusion of full FAO/UNESCO soil unit data.
6) additional data-sets
7) For my areas of interest, your more detailed topographical data is top priority. Since the data already exists, it would appear that this addition to the CD-ROM based database might be made with minimal expense and effort.

   With respect to the current data, it appears that some of the data-sets have voids in them. This impression is the result of a limited look at the data with the IDRISI GIS. If such voids exist, their elimination would appear to be a priority area for attention.
8) 1. Updated land use coverage.
   2. Soil attribute data layers.
9) Research is needed to build reliable sector data banks over a period of years.
10) Firstly a more detailed spatial resolution, and secondly a wider range of topics.
11) Higher resolution terrain data.

12) These are perhaps software concerns, but I would suggest the following as priorities for increasing usefulness.
    1. build in a means for maintaining geographic reference of data-sets through various transformations including reclassification, subsetting, and raster-to-vector conversions;
    2. provide a means for subsetting vector files
    3. enhance the range of hardware (i.e., printers, plotters) compatibility for IDRISI's "plot" module so as to facilitate producing hard copy.
13) N/A

14) N/A
15) N/A
16) 1. Error analyses for the other products included.
2. Accessibility of raw data.
17) A priority for improvement would be to ensure that all of the maps were comparatively georeferenced at all scales. Another priority would be to try and get those images which are mixed in scale (i.e. GOLSON.IMG) finished at the same scale of resolution. It would be very worthwhile to have finer scale data available. Finally, I think that there needs to be some improvement in the acquisition software (as specified in the paragraphs above).

Olson World Ecosystems - (A few data-set notes - to me this is the most useful data-set on the disk, it would be helpful if Dr. Olson could be persuaded to complete the 10 minute updates for the rest of the globe, or at least the Western Hemisphere). Why are so many categories not used? I think it would be better to eliminate the not used categories to reduce the amount of searching necessary to find a category corresponding to an area of interest. If color choice is a problem then perhaps a special palette could be provided. What documentation? Except for a brief, useful, review of how the data-set was updated the documentation is nonexistent. What are the classifications based on? How have the classifications been used (references?)? The reference from Milleman and Boden, 1985 should be included as part of the supplemental literature (especially as it is not readily available in most libraries). [Editor's note: This is an incorrect reference that was mistakenly used in the Prototype release -- the correct reference is Olson, Watts, and Allison, 1985, ...]

18) - work with GCM/weather prediction modelers to provide long-term "data" on all key variables from model output generated shortly after updating.
- include a longer period of NDVI data, and also monthly and annual averages over the whole period.
- work with IGBP-DIS toward a satellite-based global vegetation data-set.

19) 1. other data-sets as these scales
2. plus remote sensing at finer scales
3. inclusion of hardcopy of each mapped variable for comparison to electronic ones on screens to confirm data are correctly displayed.

20) N/A
21) For our research purposes, the availability of improved NDVI data-sets is highest priority. Next, the availability of time series maps of land use and climatic data is highly desirable.
22) - improved gridding techniques for continuous variables, such as the climate variables
- improved world ecosystem database
- improved (corrected) topography and bathymetry database
- new soil texture database

23) Plain ascii files for transfer to a wide variety of systems.
24) 1. Ease of coordinate/data extraction
2. A note that the "Menu.exe" utility uses the default directory c:\IDRISI\
3. Acquisition software on this CD-ROM

25) Agricultural practices and other historic databases.
26) same as 3a.
"resolution - often too coarse

patchy data (large blocks w/same value)"

27) N/A

28) N/A

29) Database expansion on the basis of comments in 3a.

30) N/A

31) - Longer time series of GVI
  - What about including data from Earth Radiation Budget Experiment (ERBE?)
  - More current climate data
  - Sea Surface Temperature?
  - How about including point data from which climate data/images were interpolated?

32) N/A

33) N/A

34) N/A

35) N/A

36) N/A

37) (Probably outside your scope of work:)
  - socio-economic data -- population, land use, agricultural zones, cities, etc.
  - analytical routines -- aridity index (as in UNEP Atlas of Desertification), agroclimatic zones, vegetation biomes -- could all be set up in IDRISI (?).
  - GCM scenarios of climate change

IV. DO YOU HAVE DATA CREATED OR OBTAINED THROUGH YOUR OWN RESEARCH THAT YOU WOULD LIKE TO CONTRIBUTE FOR THE NEXT RELEASE? PLEASE DESCRIBE.

EDITOR'S NOTE: Because of the sometimes sensitive nature of data availability issues, answers to this question are provided only in summary form, in the main body of the User's Guide. The project will pursue and inventory relevant data sources, as availability is confirmed.

V. PLEASE PROVIDE SUGGESTIONS FOR OTHER DATA THAT SHOULD BE OBTAINED OR FOR CHANGES IN THE CURRENT Data-setS. PLEASE BE AS SPECIFIC AS POSSIBLE, GIVING SOURCES OF DATA IF KNOWN.

1) It is possible to add information such as;
Nature: Volcanoes, large faults, historic landslides
Human structure: Dams, mines, large cities, observatories, and their names; and population density
2) Complete model runs with all fields from global climate simulations would be very useful eg., NCAR, $1XCO_2$, $2XCO_2$ runs these could be updated regularly.
3) The US-EPA and NCAR have compiled a tape with several GCM climate change scenarios. It could be very useful if the main parameters, <u>temperature and precipitation change</u>, are included in the global data-sets. This will make the CD-ROM more useful for global change research.... The reference could be (I am not completely sure): Jenne, R.L. 1989. Data from climate models: the CO_2 warming. National Center for Atmospheric Research, Boulder, CO.
The "<u>Trends 1990</u>" - data is somewhat outside the scope of this CD-ROM, but

could be useful for global change research. If there is space left the data-sets could be included, but most of them are not geo-referenced and could not be formatted for IDRISI. The correct reference is: Boden, T.A., Kanciruk, P. & Farrell, M.P. 1990. Trends '90: A compendium of data on Global Change. ORNL/CDIAC-36, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

The Zobler and UNESCO/FAO soil database are the only digital ones currently available, but their quality is limited. I have had correspondence with the FAO and learned that they are working on a new better digital implementation of the proper FAO-soil map of the world. The CD-ROM should include this database instead of the other two if possible. The FAO-database should be available by autumn....

I find the soil databases useful, but for modeling purposes I need databases with more relevant parameters then soil, texture and slope class. For some purposes I have created some derived data-sets with different soil characteristics, like water holding capacity, fertility and soil depth. Because of the simple correlation schemes that I have used, I do not dare to present these derived data-sets to others, but I would promote to find or create global data-sets on these properties, created by experts in this field.

Only climatic averages are included in the data-set. It should be nice to include also some climatic trends and series based on series of monthly values. The best data-set around is the one compiled by Bradley et al. (1985). I know that recently there has been an update of this database.....

Reference:

Bradley, R.S., Kelly, P.M., Jones, P.D., Diaz, H.F. and Goodess, C., 1985. A climatic data bank for the northern hemisphere land areas. 1851-1980. DOE Tech. Report No. TR017, US Dept. of Energy, Carbon Dioxide Research Division, Washington D.C.

4) N/A

5) A GCM-developed climate change scenario for temperature and precipitation in about 50 years time would be interesting e.g. as used by,
Emanuel et al. (1985) Climatic change and the broadscale distribution of terrestrial ecosystem complexes, *Climatic Change*, 7, 29-43.

I guess one of the GCM groups would be able to supply a more reliable prediction than the one Emanuel et al. used years ago.

Interpolated data for monthly evaporation (Penman-Monteith estimate) for the whole world would be very useful. The estimates of evaporation based on just temperature and precipitation, as used by Holdridge, are not very reliable. Better estimates of the water balance would improve predictions of ecosystem distribution. Monthly interpolations for solar radiation could be done at the same time (see Hutchinson, M.F., Booth, T.H., McMahon, J.P. and Nix, H.A., 1984, Estimating monthly mean values of daily solar radiation and the more commonly available sunshine hours measurements). These factors along with monthly maximum and minimum temperature (and the existing precipitation relationships) provide the basic data necessary to run simulation models see for example,
Booth, T.H. 1991, A climatic/edaphic database and plant growth index system for Africa, Ecological Modeling (in press).

In the longer term it would be good to interpolate gamma parameters (see Samani, Z.A. and Hargreaves, G.H., 1989, Applications of a climatic database for Africa, *Computers and Electronics in Agriculture*, 3, 317-325), so users could generate synthetic sequences of rainfall and link the data for specific locations to agricultural models, such as those being developed by International Benchmark Sites Network for Agrotechnology Transfer (Dept of Agronomy, Univ. of Hawaii, see their DSSAT model system). In the shorter term, people could use the actual time series data in the World Weather Disc CD-ROM (Weather Disc Associates, 4584 NE 89th, Seattle) to generate time series data for any location on the sample grids (i.e., by comparing means at locations with actual data with those sample locations and calculating a correction factor).

6) It would be very useful to have more detailed land use data and natural vegetation types. For example, regional vegetation in the U.S. in about 100 categories. (e.g., Oak-Hickory forest, bluestem prairie).
7) [comment moved to 3b]
8) same as 3b
   1. Updated land use coverage.
   2. Soil attribute data layers.
9) Thought should be given to inclusion of maps that track generation of greenhouse gases. The database could be improved through addition of parameters more directly related to sources and sinks of greenhouse gases and quantitative indices of radiative forcing.

   Information on CO_2 generation from depletion of forests in tropical countries and consumption of fossil fuels in industrial countries would be useful. Data on tropical forest depletion is available from FAO. (Food and Agricultural Organization of the United Nations in Rome).
10) Presumably you have had talks with DMA about DCW when it hits the public domain. This might provide some of the extra infill information for relief? How about including hazard and other data (possibly already in the World Data Center) - not necessarily on a raster basis eg: global earthquake/seismicity files.

    Is there any general geological or geochemical data available on a raster basis? This could be useful? Air pressure? Air quality? What about economic minerals maps?

    I realize all these above would require effort to get from paper map form into classified raster, but they would be valuable. The present GED appears to have as its major constituents input from classified satellite imagery, plus terrain, submarine, and soils information from other sources. Perhaps there is a case to be made for serious nested raster import of paper mapping products?
11) Drainage connectivity array consistent with the terrain data.
12) You might consider including some global change projection data-sets, e.g., GCM output under doubled CO_2, or Holdridge life zone maps based on the GCM output. The advantage of this would be a common basis for a variety of "what-if" exercises by a variety of workers. The drawback would be the implicit 'stamp of approval" on the global change scenarios included. Whether the benefits would justify the controversy that might ensue would require some careful thought and discussion to determine.
13) N/A
14) N/A
15) a) As mentioned in response (3) Daily means, max, min, etc. for the Southeast Region, specifically, the closest official NOAA weather station.
    b) Monthly global vegetation index. Monthly resolution is not adequate for our

modeling purposes. Our interest would be in identifying when the phenological season begins by latitude and longitude. Weekly would be satisfactory. Sources include Gutman, G. in Appendix A [of the GVI User's Guide] - *Bibliography of GVI Research Reports*.

c) The monthly data-sets of temperature, precipitation, and cloudiness have some utility to our research program. Derived indices of moisture demand; PET, AET, etc., are useful. Weekly or daily values are necessary for serious modeling of vegetation functioning.

16) Satellite infrared and visual data should be available on a grid which is determined from space requirements. The data should have as high a spatial resolution as possible. The international Satellite Cloud Climatology Project [NASA] makes these data available...

17) Some of the data that I would like to see available in the next release include:
a) The various GCM models
b) Changes in the Holdridge Life Zones under the various GCM scenarios (possibly available from NCAR)
c) Monthly evapotranspirations.
d) Monthly moisture indices (Thornthwaite) along with a map of Thornthwaite's classification of moisture requirements).
e) For North America - Kuchler's potential vegetation, USGS digital line graph (DLG) vector data and possibly some digital elevation model (DEM) data.
f) Finer scale (1-4 km) AVHRR data.
g) When available, more time series data.

18) Why not include paleo data-sets (e.g. for 6000 and 18,000 yr. B.P.) for model validation purposes. For example, DOE has a data-set of global lake - level changes. Other such globally standardized data-sets may come into being soon. Suggest liaison with NOAA [Paleoclimatology Program].

See above my comments on previous page. For the concept of using weather prediction models to generate grid-based, physically consistent data contact. John Kutzbach at the Center for Climatic Research, Madison. This is one way to provide data important for ecological and hydrological modelling e.g., winter minimum temperature, net radiation, storm frequency.

19) Global Land-Use Patterns: through my IIASA project, David L. Skole of University of New Hampshire, Center for Earth, Oceans and Atmosphere, obtained and now has digitized, Soviet maps of global land use patterns. Because climate along with landforms constrains land use, these and similar data are very important sets for inclusion in your database.

Climate Change Scenarios - the current set of EPA-distributed climate scenarios from GISS, GFDL, CCC, etc. could be added profitably, with suitable cautions about their limitations. Especially useful would be intermediate data not usually available on daily temperature or precipitation extremes as model output.

Atmospheric chemistry data - I am unaware of any data-set which could provide global distributions of atmospheric NO_x, SO_2, etc. Perhaps enquiry among the users of your CD would bring out this useful database.

20) I would be delighted to include the collection of the data about paleo-environmental change for the last 130,000 BP. At least 125,000 18-20,000, 10,500, 8000, 6-5500, little ice age, little climatic optimum of holocene. The data about anthropogenic climatic change also are very important.

21) We would be interested in another global data-set of cloud cover (0.5° lon/lat resolution) that is based on satellite data.

In our quest to improve our vegetation data, we have been "hand digitizing" the following vegetation maps for our model. You may be interested in obtaining the following data-sets:

Belov, A.V., I.I. Bruks, E.K. Fedozova, S.A. Gribova', T.I. Isachenko, N.M. Calibernova, Z. Y. Caramisheva, G.S. Catenina, G.M. Ladygina, V.V. Lipatova, N.R. Lifvinova, N.I. Nicolskaya, E.I. Racheovskaya, I.N. Safronova, V.N. Haamtsov, T.C. Yurcovskaya, I.S. Ilina, Y.S. Prozorov. 1990. *Vegetation of the USSR*. Moscow State University, Geography Department, Moscow USSR.

Hou, H.Y., S.C. Wang. 1979. *Vegetation Map of China*. The Laboratory of Plant Ecology and Geobotany, Institute of Botany, Chinese Academy of Sciences, Map Publisher of the People's Republic of China. Bejing, China.

Joint Federal-State Land Use Planning Commission for Alaska. 1973. *Major Ecosystems of Alaska*. U.S. Geological Survey, Fairbanks, Alaska.

Kuchler, A.W. 1964. *Potential Natural Vegetation of the Conterminous United States*. American Geographical Society, New York, New York.

Rowe, J.S. 1972. (Reprinted 1977) *Forest Regions of Canada*. Publication No. 1300, Department of Fisheries and the Environment, Canadian Forest Service, Ottawa, Canada.

UNESCO. 1981. *Vegetation Map of South America*. UNESCO, Paris, France.

White, F. 1981. *Vegetation of Africa*. UNESCO. Paris, France.

22) I have heard of US activities to create an improvement for the terrain data-set (I expect you have better and more specific information about that).

A most important type of data I would like to see on this disk is anomaly fields derived from GCM's, similar to the standardized files available from NCAR. These files should be at both, GCM resolution and gridded to some of your grids.

23) Numerical weather prediction analyses data.

Operational NMC analyses that give 3-dimensional picture of atmosphere structure. Contact ... NCAR and ... NMC/CAC; Outgoing Longwave Radiation, ... NCDC.

24) The GCM output mentioned in the manual is a good idea.

A collection of the satellite imagery of earth from the Spot and Landsat would be extremely useful (Available from EOSAT ....).

25) N/A

26) Substitute full CIA World Data Bank II for micro version. I think this has greater resolution and more attributes. Contact NTIS ....

You might want some data from the DMA's Digital Chart of the World Project. Contact ... DMA....

27) N/A

28) N/A

29) ...Hydrometeorological and environmental data-sets which have been created or are being created in different organizations in Russia and outside it. The sources can be specified...

30) Seasurface temperature
AVHRR
GHOSTA (UK Met. Office Database)
Atmosphere data - mean values

Total water vapor (SSM/I) - already done by NSSIDC

31) Continental-scale higher resolution NDVI data from NCAR or EROS data center.
Time series of seasurface temperature (NOAA)
Sea Ice? NOAA?
Chlorophyll concentration from CZCS - NASA
32) N/A
33) N/A
34) I think inclusion of variables related to surface energy and/or water balance, e.g. net radiation, evaporation, runoff etc. can be useful. However, since there are perhaps no data-sets ready to be included, they would be put off to later editions. I wonder why Wilson and Henderson-Sellers' vegetation data-set is not included [Editor's note: we were awaiting copyright release]. It is different in philosophy from Matthews digitized potential natural vegetation, while Wilson did existing vegetation. Both approaches have their merits and demerits. Wilson's soil data-set is derived from FAO/UNESCO soil map, so it may not be necessary now when Zobler's data-set is included.
35) N/A
36) N/A
37) Our goal (see the attached project description) is to compile the FAO agroecological zone data (improved to worldwide coverage) and model on a CD-ROM. This would integrate the natural resources, agricultural suitability and socioeconomic interpretation into one database/GIS that users around the world could test -- either globally or regionally by filling in finer-scale data and analyses.

VI. IS THE YEARLY RELEASE SCHEDULE REASONABLE, GIVEN THE PACE OF GLOBAL CHANGE RESEARCH AND/OR YOUR RESEARCH?

1) N/A
2) The latest release should be available over the network with annual updates of the CD-ROM.
3) I do not think that a yearly release is practical. Updating of most data-sets will take a longer periods, especially because the creation of data-sets is not the main aim for global research, although often on the agenda. This can easily be determined from the documentation. Many papers describe the data-sets only in summarizing terms and emphasize on the application. If each release of the database should be an improvement over the previous, a biannual release schedule should be more feasible.
4) We're not that familiar with your users but a yearly publication of a CD is in concert with some of the things we're doing.
5) Yes
6) Yes
7) For the data on the disk a yearly update would appear adequate. For the additional topography data mentioned above, the sooner the better.
8) Obviously, the sooner the better.
9) I would consider it reasonable.
10) Yes
11) Yes
12) Yes
13) Yes, except in the case of annual reports associated with [our] program. An annual

report is required 9 months after the close of sampling (July). We already have some problems obtaining current climate data in time and I imagine the lag would be even greater with this system. For more traditional research and development activities associated with [our programs], the annual time frame should be adequate.

14) N/A
15) Yearly release of updates to the database are reasonable based on the need for data of high quality, refinement, and release on a ROM disk.
16) The yearly release schedule is adequate as long as some flexibility is included to account for variability in the production of new and important data-sets.
17) A yearly release schedule is reasonable only if the data significantly changes. Rather than include the same data each year it would be more useful to only include corrections and new data-sets in each yearly installment. Along the same lines it is very important to keep researchers appraised of errors found in the database. Perhaps a dial in bulletin board would be more cost efficient than periodic mailings.
18) Yes - we don't need to update too often; there is something to be said for the use of standard databases that are "valid" for a reasonable time, e.g., 1yr.
19) Seems very reasonable although if you are funded well enough, you could send out a quarterly/monthly newsletter which describes B-test data-sets recently available, for those who cannot wait!
20) It's reasonable, at this stage.
21) I'd be concerned about the quality of the data if I was to develop data-sets for annual release. However, with several groups working on global data-sets, I think a yearly release schedule is reasonable if different groups contributed to a specific release. Indeed, at a later date, it may be desirable to have more frequent releases.
22) Yes
23) Yes, although it would be nice if it were more often. Also, these files really should be temporally consistent.
24) Yes
25) Yes, to ensure the high quality of the data and to let it be reviewed.
26) Yes
27) N/A
28) Yes
29) The yearly release schedule is reasonable.
30) I believe that yearly release is probably the maximum rate at which these data should be released. The processing of checking data formats etc. and validating actual values is necessarily time consuming.
31) Yes - it is more appropriate to provide data of dependable quality with good documentation, than to release data prematurely!
32) N/A
33) N/A
34) N/A
35) N/A
36) N/A
37) Yes -- see [question] #12.

VII. IS THE DOCUMENTATION HELPFUL, COMPLETE, ETC.? (PLEASE ATTACH COPIES OF THE DOCUMENTATION SHEETS FROM THE MANUAL WITH YOUR EDITS, IF POSSIBLE).

1) N/A
2) Very satisfactory.
3) The documentation gives currently the most important information. I would like, however, the addition of a short summary and some notes on the applicability of the database. The summary (c. 100 words) would describe the data and would allow for a faster evaluation of the content then by looking at the proper documentation or requesting the additional literature references. The extra notes would summarize the scope of the data, its limitations and applicability.

Each contributor of global data-sets should be urged to fill in a standardized information sheet with room for all important information. If the contributor has not filed such information sheet, the data-set should not be included.....

- The correct reference to the IIASA climate database is now:
Leemans, R. & Cramer, W.P., 1991. The IIASA database for mean monthly values of temperature, precipitation, and cloudiness on a global terrestrial. Research Report RR-91-18 November 1991, International Institute of Applied Systems Analyses, Laxenburg, 61 pp.

The reference Solomon and Leemans (1990) should be moved to additional references. It presents an application and does not need to be included as primary reference, but as additional reference. Another additional reference could be:
Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon, A.M. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeography* (in press)

......I find the documentation of the Olson data-set insufficient. In my global change research, I found this the best vegetation database currently available, better than, for example, Matthews, because very clear descriptions of each vegetation class are given in the accompanying report (Olson, J., Watts, J.A. & Allison, L.J. 1985. *Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: A Database,* NDP-017, Oak Ridge National Laboratory, Oak Ridge, Tennessee.) This reference should be included in the documentation as primary reference and not as additional, and if possible, also a reprint of this report.

The soil databases (Zobler and UNESCO/FAO) are all based on the FAO soil map of the world. I find it essential to include the original reference of the map the documentation. This reference is:
FAO/UNESCO 1974. *Soil Map of the World,* 1:5,000,000. Food and Agricultural Organization, Paris, France.
4) Yes, quite thorough for a beta test.
5) Yes. I've only had a couple of minor problems so far.

When I first tried to use ACCESS I received an "Out of environment space" message. I had to trim my existing long path statement. I believe there's a DOS command to increase the environment space, but I've not had time to track it down. SHELL C:command.com /E:400 /p increases environment space to 400 bytes from default of 1606.

I thought the description of the use of SUBSET could be improved by giving an example of its use. If I want to just look at Australia do I have to extract all the relevant data using SUBSET or is it possible to change the default window that COLOR uses, so that it would show Australia first rather than North America? It would be a good feature to add if it's not presently possible. Incidentally, the new windowing feature of COLOR is excellent.

6) N/A
7) I found the documentation to be generally complete and appropriate. Some of the copied material was difficult to read and some of the graphics lost much of their usefulness due to blurring of the copies. Where possible, you might ask the contributors to resubmit cleaner originals. The IDRISI and World Data Bank documentation was probably the least appropriate and complete. My full comments on IDIRISI and World Data Bank are provided below. To reduce your overall production cost you might consider putting as much of the documentation as possible on the CD-ROM. If maximizing the potential use of this excellent data is a goal, then keeping the total distribution costs to a minimum would appear to be a necessity.

   I hope that the IDRISI documentation extract provided in Appendix A is not representative of the full documentation, especially if the intent is to use IDRISI as some form of standard GIS for working with the data at the high school level. Although the various modules (programs) provide considerable capabilities, it is unclear from the extract how the capabilities are integrated. Although the menu program provides an interface to each of the programs, it does nothing to integrate the capabilities. Based on what is provided, it is my impression that it will take a fairly computer literate and experienced teacher to figure out how to use the software/data combination. The teacher will then have to figure out how to package the programs so that the students can make use of the capabilities. The general syntax and contents of the documentation is typical of a computer technical reference manual - not something which will be effectively and efficiently used high school teachers and students....

   In all fairness, I must admit that even this partial IDRISI GIS has been of considerable help in looking at the data and gaining an appreciation for the considerable value of your efforts.

   If there will be additional reviewers who will be working on this project in the future it might be useful to do some rework on the extracted documentation and build some step-by-step tutorials to get them started and save some time.... The value of the data you are assembling should not be diminished by the IDRISI documentation.
8) In general, documentation is adequate.
9) The documentation is helpful and complete, However, there needs to be better explanation, with examples, to make the data from the CD-ROM more readily accessible and user friendly.
10) Most impressed with the documentation. The only major problem is that for non-IDRISI users there may well be some difficulty in getting going. The IDRISI section does not indicate how a sensible path through the various modules might be chosen.
11) Yes, but I'm just starting to use it.
12) As mentioned above under 3a), description of the method of combining life zones for the aggregated Holdridge life zone data-sets is needed. Also, some discussion of the similarities and differences between the two monthly weather data-sets would be helpful.
13) No! This, along with the "lowest-common-denominator" issue were the strongest comments from my technical reviewers. For example,
"Specific documentation on the structure of these image files does not exist, e.g. data types are defined as 'byte' and file types are defined as 'binary'. I did find one

reference that some grid files (only the 8-bit raster files), are identical to GRASS cell files. I did attempt to manipulate a ten-minute GVI file into an appropriate GRASS cell subdirectory, but I was not successful. In Appendix B, the documentation reports that 'import of the data into GRASS is relatively easy...' however, the special software is not provided and more detailed instructions are necessary."

14) Some of the documentation is confusing. Descriptions of the data-sets found on the DATA FILE STRUCTURE pages are terse. At first we did not realize that most of the data record formats described by documentation in sections (1) to (11) are relevant not to the GED files, but to earlier versions of the same data. Almost all of the information that one must know to access the GED versions is found in the GED Introduction. Since, when first reading the documentation, one might easily pass over these few pages of information, a note or statement on each Data-set DESCRIPTION or DATA FILE STRUCTURE page that directs the user back to the GED Introduction might be helpful. (See notes on attached copy of <u>data file structure</u> page.)

    Some of the additional information provided in sections (1) to (11) discussed the sources and history of the data-sets. This information is important and should remain part of the GED documentation.

15) See pages 8-10 for example. It is difficult to tell which data-sets are of the same format/file if they are separately collected and formatted. Are these data-sets in some kind of relational database or separate files? We may have been able to ascertain this a little better if we had been able to view the CD-ROM disk.

16) The documentation is essential. It was quite complete and useful. The photocopied journal articles were very light and hard to read in some places. They should be reproduced using a text editor.

17) General notes concerning the documentation. The documentation should continue to be released as a hardcopy document and should also be broken up into two volumes. The first volume should contain all of the introductory material, data-set descriptions, and an overview of the data (probably synopsized from the original sources). The second volume should contain the supplementary material. In particular, a copy of the actual paper that describes that data-set (not just the reference) would be very helpful. Additionally, one or more papers that demonstrate the use of the data-set should also be included. I realize that in some cases this represents a significant effort to secure copyright approval. However, if the manual was well done it could probably be published as a stand alone book as well as supplemental information for the database (I'd certainly read it).

18) Mostly yes, but I am dissatisfied with the Olson and Matthews documentations. Olson's "legend captions" doesn't provide enough information to document the changes he has made over the original version, and contains much irrelevant and even incomprehensible data. Matthews does not give enough information about data sources, and particularly the method for assigning seasonal albedo values.

19) Documentation written by original authors to fit a prescribed format, then edited by you, might be more useful. The authors of the test database documentation did not seem very knowledgeable of the development of the sets, nor of who their true authors were - citing Ray Milleman as author of Olson's Database.

20) Now it is.

21) In general, I found the documentation very helpful and a delightful collection of references. However, there were certain aspects of documentation I found confusing.

    Although I liked the Global Vegetation Index User's Guide as a reference, it's inclusion with the NDVI data-set caused a bit of confusion. At first, I wasn't sure whether or not calibrated NDVI was included as part of the data-set. Then, I wondered why the calibrated data-set was not included with the data-set. If there were controversies in determining the calibrated NDVI, they were not apparent from the documentation.

Why was the Leemans and Cramer data considered "Monthly" data when it was averaged monthly? To separate it from Legates and Willmott? Why not label the sections Leemans and Cramer data and Legates and Willmott data?

22) Unfortunately, I do not have time to edit the documentation for this review. I think the current format is reasonable if one prefers a relatively rapid production of this database. Ideally, of course, more standardized and less individual information would be most useful.
23) Need better installation procedures. The documentation that was provided did not work on our PC. (Append command doesn't work.) Need more information about files so we don't have to use other software on PC's.
24) 1. I truly did not enjoy putting the documentation together.
    2. The title of the database should be changed. It should read something like "Global Change Databases" or "Earth Science Databases"
25) Doc files look well organized and uniform.
26) Yes - very good.
27) N/A
28) The documentation for data transfer to GIS packages was inadequate with the exception of the IDRISI package. If data conversion software to convert the data to common spatial data formats (see 1c) had been provided extensive documentation on the details of the data format would have been un-necessary.
29) The documentation is rather complete.
30) Documentation on data - itself appears good, more help with implementation would be useful.
31) Yes, the documentation is quite good except for the cases below where the packing scheme was unclear.
32) Were the instructions clear and accurate for reading and displaying the contents of the database?

The Introduction section is clearly written, but Appendix A (IDRISI Implementation) leaves much to be desired if (as I attempted to do) the subset of IDRISI included with the database is to be the primary means of browsing through the data. Much of the problem appears to relate to the interface between IDRISI and DOS (and how that relationship is described in Appendix A). In particular (pages referenced being to Appendix A):

(a) On page 1, there appears to be a typographical error ("Append /E" should actually read "APPEND \E" - that is, a backslash should be used). *[Editor's note: This is actually incorrect. The proper syntax is forward slash, i.e. APPEND /e -- please refer to the User's Guide or your DOS manual]*
(b) Near the top of page 3, "ALL." should be "ALL" - that is, the period should be deleted.
(c) The use of ACCESS and APPEND should be more clearly described. In my installation, even though the APPEND path appeared proper, based on the response to entering APPEND at the DOS prompt, the environment shown in response to entering ACCESS did not include the APPEND path. Nevertheless,

IDRISI appeared to be able to access the image files (but, for some reason, not the Micro World Data Bank II vector files) when "color a" was selected, although not when "list" was selected. In the latter case, IDRISI responded that the file was not found in the default directory.

(d) For users depending on the provided subset of IDRISI, it would be useful to provide some minimal level of documentation of IDRISI. It took some amount of trial-and-error to determine which commands were needed to display a map of each data-set.

(e) I used IDRISI to browse through each of the 11 chapters (looking at the first data-set in each chapter, if it contained more than one - e.g., GMVEG in Chapter 6). There was a problem, possibly specific to the personal computer system used, with FAOSOIL (but not GFAOSOIL) - the file appears to be have been read, but only the title and legend, and not the mapped data, appeared on the screen - perhaps the file was too big. If this is a widespread problem, a warning about system requirements would be useful. And, as mentioned in (c), the Micro World Data Bank II vector files were not read, perhaps a result of the way that IDRISI was used; as mentioned in (d), more instruction in IDRISI might have helped.

33) N/A
34) N/A
35) N/A
36) N/A
37) Needs some improvements, especially on origins of the data, quality checking, etc. The user beware caveat is too easily lost in the rush to have results.

VIII. DO YOU THINK THAT WE HAVE ADEQUATELY ADDRESSED THE ISSUES OF DATA QUALITY? NOTE THAT DATA QUALITY ALSO REFERS TO DATA DOCUMENTATION, ALLOWING THE USER TO JUDGE THE FITNESS OF THE DATA FOR A PARTICULAR USE.

1) N/A
2) A broad statement spelling out the nature of the data and the responsibility of the user would help at the <u>very</u> beginning.
3) No, I don't think that you have adequately addressed the issues of data-quality. You have compiled a large series of very useful data-sets, reformatted them and presented them in a unified format. I find this already a major achievement. Currently, the CD-ROM contains only the data-sets and some of their documenting papers. No quality assessment is done at all. I would like that the contributors not only write a short summary describing the data (see also point 7), but also a similar note on the scope and limitations of their data. I foresee that many quality reviews of the individual data-sets will be published in the scientific literature when the data becomes generally more available. If no notes are included on the limitations of the data-sets, negative reviews of individual data-sets could start rumors that the global data-sets are trivial. The addition of a critical note could remove such threats.
4) Yes, unless your user community is very limited and well defined its difficult to fairly assess data quality. Its more important to provide data availability. Let the end user decide for himself what's best.
5) It's really the responsibility of users to read the original papers and judge the data

quality for themselves.

6) The uncertainties and assumptions associated with each data-set are not clear.
7) Based on my initial review, I think you have. As noted below, the documentation in IDRISI and World Data Bank leave something to be desired.
8) In most cases, yes. In the case of NDVI, however, there could be more documentation on the potential uses, calibration to plant production, and limitations.
9) I think you have done as well as you could. However, interpretation from the maps, I believe, could lead to significant errors.
10) Data quality is always difficult to assess: particularly where data-sets have been patched together to provide the global coverage. The ETOPO5 terrain data does not register perfectly with other mapping (in places). Unfortunately I did not have time to try a superimposition on ETOPO5 and the DCW prototype vectors. I imagine there might well be some problems.
11) Yes
12) Except as noted in 7), yes.
13) Again, a detailed reply is not possible since all our time was spent trying to access the data. I am somewhat skeptical, however, of the quality of the data contents if the same people prepared it as prepared the operational documentation.
14) It is difficult for those who compile different kinds of data to adequately address data quality issues. On the other hand, it might be possible to establish certain minimum criteria for the data-set overall and to indicate whether or not each author has verified the quality of his data to the extent of those criteria. Of course, each user of the data must also realize that the data are provided for use as is and that quality assurance is to some extent the responsibility of the user.
15) Data quality looks to be excellent based on the amount of time and effort to refine and document the process that was used to compile it.
16) The error analysis should be included for all data-sets if possible. The raw data should be available if the user would like to conduct his/her own error analysis.
17) I actually see little evidence that the issue of data quality has been addressed at all. In the integrated bathymetry/elevation data-set you refer to a mis-registration without following up on the problem. I do think you did a good job in explaining some of the quality control that went into converting that data into a format usable in the database (but there are specific concerns - see comments above). I actually think that, in general, there is very poor overall documentation as to whether a data is fit for a given use (see comments above). Given the EPA's concern over quality assurance issues (and my own agency's as well), I think that more time needs to be spent covering the quality and usefulness of the data.
18) For now you can only use data-sets that exist; but I would like to see some incentive for the providers of vegetation data-sets to be more explicit about their classification procedures.
19) I had no problem but then, I only used the databases for which I had personal knowledge.
20) Yes, of course.
21) Other than noted above, yes.
22) cf my answer to question 3a). I think, more could be done, and specifically, I think the problem of resampling should be addressed more. One solution would be to always include the original data in some practical format.
23) Yes, data-set description in the .doc files and in the manual provide enough

information for determining whether or not data is adequate for our use. It isn't yet. We are really only going to use vegetation because we need it for January, 1988.

24) Yes
25) Don't know.
26) Yes
27) N/A
28) Did not review data content (see 3).
29) It is difficult to answer the questions because we do not have enough time to analyze the data quality and completeness to a required extent. However, we promise to send to your address all our comments which may arise in the course of our work with the data.
30) For the FNOC data there appeared to be little documentation beyond basic description of data format. I couldn't find any discussion of data quality or much on historical development.
31) Again quite good; but more complete documentation/citations should be provided for <u>all</u> data-sets.
32) N/A
33) N/A
34) N/A
35) N/A
36) N/A
37) " [see question #7]

IX. PLEASE LIST THE DIFFERENT TYPES OF APPLICATIONS THAT YOU OR YOUR FACILITY ARE INVOLVED WITH.

1) N/A
2) GIS, hydrology, vegetation modeling, atmospheric density, atmosphere/biosphere interaction, global environmental management.
3) My department is involved in research on different aspects of the global change. There are two main fields. The first is the carbon cycle modelling with emphasis on global and regional emissions, global land use patterns, carbon fluxes, and population growth. The second field is modelling the impacts of global climate change on eco- and agro-systems. We use global climate data-sets, climate vegetation relations and vegetation data-sets for different applications.
4) N/A
5) I'm interested in predicting where and how well different tree species and provenances will grow under present and future environmental conditions. (see papers enclosed).
6) Measurement and modeling trace gas emissions including:
   - methane emissions from cattle, rice, termites, natural gas distribution systems
   - non-methane hydrocarbon emissions from vegetation
   - NO and N_2O emissions from soils
   - reduced sulfur gas emissions from soils and vegetation
7) Developing and distributing basic data to support spatial research and generation of maps and map based products.
8) Modeling of tropical C and N cycles and trace gas fluxes from soils. Global respiration modeling.

9) We are involved in monitoring mitigation and adaptation of gas exchange between the atmosphere and biosphere, ecological dynamics, disturbance phenomena such as fire and insects, and human interactions.
10) Our work has been mostly local scale, but we are hoping to move into environmental global models and modelling in a general way with our final year undergraduate and Master's course students. I have started assembling pieces that can be made into a course that would serve as an introduction to the possibilities of using regional and global scale data. Your CD-ROM, and its hoped for successors are invaluable in this respect. In many ways more useful than the DCW data as it is presently easier and faster to handle.
11) World Climate Research Program's Global Energy & Water Cycle Experiment (GEWEX); Hydrologic Forecast Models.
12) In my present position I am assembling a database of narrow endemic vascular plants and vertebrates of the Pacific Northwest, their distribution and habitats, and collecting information necessary to assess the potential impacts of global warming on these organisms.
13) Global Climate
    - scenario development
    - GCM scenario analysis
    - integrated regional modeling
    - climate impact assessment
    - EMAP (Forests and Integration and Assessment)
    - ecosystem monitoring
        - ozone
        - wet deposition
        - dry deposition
    - air chemistry monitoring (wet and dry deposition)
    - background studies
    - baseline definition
    - within ecological resource assessment
    - trends identification and description
    - associative studies
    - across ecological resource assessment
    - interpretive studies
14) .....basic and applied research in the atmospheric and planetary sciences. [A] staff of about 60 scientists and support personnel work in the fields of atmospheric chemistry, meteorology, climatology, physical oceanography, remote sensing, numerical weather prediction and planetary physics.
15) Our applications at this time are working and validating tree growth and stand level process models using data we will collect experimentally and that available from other sources such as Global Climate database.
16) ....analysis of large and small scale climate research. We have been involved lately in the determination of rainfall from satellite data-sets and the verification of existing satellite-based rainfall algorithms.
17) Researchers at our facility are involved with a great number of topics. My own work deals with the influence on climate on the distribution and abundance of North American birds. I am modelling how birds respond to climatic influences and will then try and project how a given species' range may change under various global

warming scenarios. Other work done here includes studies on population changes in neotropical migrants, impacts of global warming on wetlands in the Prairie Pothole region, impacts of grazing and land use practices on distributions and abundances of birds in the northern Great Plains region, and many issues concerning waterfowl population trends and biology. Many of the data-sets in this database would be very useful to work going on here at the center. While a North American database would be even more useful, this data can be sub-setted and provides a very good start. I am particularly looking forward to using this data-set in GRASS once we get our new GIS lab set up.

18) - Development of global biome models
- Application of global biome models to paleoclimatology and paleoecology, including data model comparisons
- Coupling biome models to atmospheric general circulation models and analysis of implications for ecosystems
- Development of vegetation dynamics models
- Analysis of vegetation sensitivity to global warming scenarios and CO_2 change.

19) 1. Research in developing global biosphere models ....
2) Research in modeling tree migration in upper Great Lakes and Canada.
3) Graduate education in global change (use in seminars, GIS courses, graduate research).

20) N/A

21) I'm not sure I understand the question, but here's an attempt to answer it:
.....basic field research programs studying ecosystem processes (including trace gas fluxes) in:
a) lakes
b) streams
c) salt marshes
d) oceans and inlets
e) tundra systems
f) temperate forests
g) tropical forests

In support of these activities, the Center has been developing several models to examine ecosystem processes and the effects of various factors on those processes. In addition, the center is developing global- to continental-scale ecosystem models to describe the effects of climate and other factors on ecosystem processes.

[Cover Letter] Our group is interested in trying to link the results of field studies to the effects of global change in ecosystem processes (e.g., NPP, net N mineralization, trace gas fluxes) at continental to global scales. We are developing spatial models to provide this link.....

22) - global change impact on terrestrial ecosystem scenarios, using ecosystem models
- education of geography students on global ecology, database management and GIS

23) 1. Climatic research.
2. Intraseasonal climate variations research.

24) 1) Carbon cycle modeling
2) Agricultures contribution to Global Change
3) Potential C sequestration in agriculture
4) Earth albedo

25) Global Climate Change

US/North American agriculture carbon cycle interactions.

26) CD-ROM general reference/education atlases

27) N/A

28) 1. Atmospheric Research
    2. Visualization for scientific research
    3. Exposure Modeling
    4. Risk assessment

29) The database has been used in the course of the global climate change analysis and the investigation of their connection with regional anomalies.

30) Earth Remote Sensing (RS)
    RS product processing
    RS product calibration and validation

31) 1) First ISLSCP Field Experiment:FIFE
    - Association between satellite data and terrain variables
    - Estimation of surface biophysical variables from satellite data

    2) California Biodiversity Project
    - vegetation mapping and ecological modelling of habitat types for entire state of California
    - GAP analysis to identify ecological preserves based upon biodiversity criteria

32) N/A

33) N/A

34) N/A

35) N/A

36) N/A

37) Crop-yield modeling – daily + monthly water balance + agroclimatic index; point + spatial scale
    Resource/population/land use models
    World food trade
    Landscape vegetation zones
    Energy demand
    Greenhouse gas sinks
    (mostly all corrected to climate impact assessment)

X. WHAT APPLICATIONS DO YOU SEE THE GLOBAL DATABASE BEING USED FOR?

1) N/A
2) see (9) and educational purposes in training workshops.
3) Education: With these global databases it becomes more easy to present many kinds of global scale patterns, to link and evaluate them. The CD-ROM could in time become more useful for education of our physical global environment then an atlas.
   Research: The database could be used to describe species distributions in terms of climatological parameters. Some research has been done already by no there becomes data available for continental and global applications. If the CD-ROM is available throughout the world many new and unforeseen application will surely be developed.
4) N/A

5) Ecological analyses
   Agricultural/Forestry analyses
   Under both present and future conditions.
6) Global biogenic emissions, which are a function of soils, vegetation type, temperature, light intensity, biomass density, and other factors.
7) Clearly, environment, weather, and geography research and education comes to mind. The topography data will probably be used in a wide variety of research, education, and production work. You might consider asking your customers to let you know what they are doing with the data.
8) Spatial ecological modeling and biogeochemistry changes in response to climate change.
9) I think it should be useful for reference and base-line data from which to build more definitive sectoral data.
10) Global/regional modelling of atmosphere, environmental change. Hazard assessment, monitoring, regional assessment of reinsurance risks (given geological overlays etc).
11) Global Hydrologic Modeling and Improved Parameterization of GCM's.
12) Regional and large scale studies of interaction between physical environment and natural systems, and/or physical environment and human-dominated systems seen to use the most appropriate uses of the global database. These studies would include generalized descriptions of individual regions. Regions with similar environments in different locations or regions with contrasting environments could be usefully compared.
13) For [our] activities, the database's primary use would be in the development of background products and the definition of baseline conditions against which ecosystem status and trends assessments could be compared. For climate studies, the database would be used as a baseline against which GCM baseline output would be compared. It could also support ongoing research into techniques for the derivation of more detailed regional climate change scenarios.
14) In addition to using geo-political boundary and elevation data to plot maps for use in regional data studies, [our program] scientists have expressed interest in using a global database such as the GED; (1) to verify surface vegetation types for satellite-based cloud studies and (2) to obtain surface boundary condition data for use with numerical weather prediction and general circulation models.
15) We envision some kind of spatial representation of the Forest Ecosystem with separate layers to include different plant and tree species, growth through time based on interactive input, and linkage to expert systems as they develop for different types of modeling environments.
16) They are directly applicable to construction and verification of climate models and the assessment of climate and climate changes. The packing of several data-set on one optical disk will allow easy intercomparison of the data-sets.
17) At present I see the global database being used primarily for biogeographic and terrestrial ecology questions. This is not necessarily all that bad. Before we can understand what impacts global climate change may have on plants and animals we need to have an idea of how they respond to "normal" conditions. Most ecological and biogeographic studies have had a very limited geographic coverage. This has largely been due to, I think, the lack of available data and lack of understanding of the scales involved. The global database goes a long way toward resolving the issue of data availability. Scale issues are important and not so easily solved but that is not

the job of the database. I reiterate the point that I think that better data extraction routines need to be written that will allow users with no availability to sophisticated GIS to be able to access the data. Without the data extraction routines I think you might risk being perceived as collection of pretty pictures.

18) All kinds of work involved with global ecosystems modelling, global hydrological modeling and biosphere - atmosphere modelling.
19) see 9 above
20) The decision of the forecast problems. The different model creation (e.g., models for CO_2 increase and others).
21) The global database will primarily be used as input data-sets for the modelling efforts. However, data may also be used to help put together summaries for publications related to the field research program.
22) The ones listed under 9) (we are a geography department with all standard activities usually found at such a department; the ones I mentioned are the ones of relevance to this product).
23) 1. Comparisons with GCM's
    2. Comparison with analyses like winds and moisture fields.
24) At least everything I have listed in question 9.
25) Climate Modelling. Once equivalent versions of the same data categories have been gathered for a long enough period of time, historical trend analysis - prediction will be possible.
26) I can't speak to any but our own, but I am sure you will be surprised by unexpected uses.
27) N/A
28) N/A
29) The database can be applied to research in hydrometeorology, ecology, for environmental monitoring, etc.
30) As a basis for RS (remote sensing) data to be developed and understood. As understanding of a particular data-set develops it may itself become part of the global database - once proven adequate to agreed standards. In this way, the global database may become an international standard.
31) - Exploratory analysis of global scale vegetation - climate interactions
    - Continental phenology studies
    - Change detection of large scale practices of deforestation
    - Analysis of time-space properties of satellite data.
32) N/A
33) N/A
34) N/A
35) Morphometric analysis of 5 and 10 minute DEM data using classification system of Hammond.
36) Evaluation of Climate GIS Databases for Identification of Crop Environments: case study for Northern Latin America.
37) Global and regional Agro-Economic Zone project -- [as described elsewhere]

XI. WHAT PROBLEMS DO YOU FORESEE IN MAKING THIS DATABASE AVAILABLE FOR EDUCATIONAL PURPOSES AT VARIOUS LEVELS? THIS MAY INCLUDE HIGH SCHOOL THROUGH COLLEGE, PUBLIC EDUCATION, TRAINING PROGRAMS FOR SCIENTISTS AND MANAGERS.

1) N/A
2) (1) Training the educators.
   (2) Obtaining appropriate hardware.
3) One of the problems that I foresee is the scale of the dataset. Most of the data sets are on a 0.5 degree or even coarser grid. For educational purposes users will not be fully aware of the limitations of such medium scale resolution and use the data for only small regions, like their own backyard, and find discrepancies with the observed patterns there. I think that a thorough discussion on different scales, scope of applications, etc., should be included as a guidance for the potential user. Unfortunately, I do not know of a paper that addresses this issue.
4) Maintaining a level of documentation that would serve both communities. The educational community would prefer a canned tutor session or sample session to get by some of the technical expertise needed otherwise.
5) None foreseen.
6) N/A
7) As with any new product, simply getting the word out to the intended audience is the first problem. The second will be with the schools getting, or allocating, the money to purchase the database, IDRISI, and the CD-ROM drive if they don't have one. Getting the product to the schools will only be the first part of the problem. If the rest of the IDRISI documentation is like what is in the manual I received for this review the bigger problem will be in using the data - especially at the high school level. For the school environment I would strongly advise that the IDRISI documentation be considerably expanded and include a tutorial. The value of the data will be greater if it is supported by software which is geared to that level. Based on the flyer on IDRISI I received from Clark University it appears to me that many of the IDRISI capabilities are quite advanced for the high school level. For this level some other less technical program might be advisable. In this respect, the IDRISI documentation is the low point in the package I received.
8) Integration with dynamic models.
   Lack of data at various spatial scales (i.e. 60 minute)
9) The main danger is making errors in interpreting the data, and extrapolating the information beyond the applicable scale.
10) No problems at all! It would be an idea to develop a good interactive teaching pack to help the uninitiated (academics as well as students) operate prepackaged courses, and to include a genuine modelling (in the Spans sense) capability, possibly based on IDRISI, and certainly based on nothing more elegant than a PC386 with Super VGA. There is nothing more irritating in the educational environment than to discover that your 20 PCs don't have the vital piece of specialist hardware required. Packages (and people) get thrown out of the window when that happens.
11) N/A
12) As the database becomes more widely distributed, I would foresee problems both of accessibility and (over-) interpretation. I do not think CD-ROM readers are available uniformly; at lower educational levels the problems probably become worse. GIS

expertise is probably equally limited in distribution. As currently distributed, problems in producing hardcopies of plots would also limit the accessibility of the information, I found the "plot" module of IDRISI necessary for combining raster information of interest with vector files from Micro World Data Bank II that provided geographic reference. I wasn't able to produce a hard copy of the plot though, and I imagine most users would lack the hardware to do so. At the same time, the farther the information gets from the compilers, the less appreciation there will be of inherent errors due to resolution of input data, resolution of distributed files, methods of interpolation, etc. Pointing out the usefulness and the limitations should be a major component of documentation and promotional material.

13) The biggest drawbacks at this point are the poor access documentation and the access speed. In its present form, the data transfer time is excessive (this refers strictly to the hardware aspects, not those associated with the documentation).

14) N/A

15) No problems in making the database available for educational purposes, only possible misuse of the data in not analyzed within context - extrapolations, statistics, unscientific procedures, conclusions.

16) There should be no problems. However, they may be of limited use due to the lack of sufficient hardware and software and the expertise to use them. Training programs will be necessary.

17) 11 and 12) I have lumped these two categories because many of the comments are applicable to both. In some respects it seems that you need two different copies of the data acquisition and view software for the two different groups. I think that the database itself is useful for many different levels of use, from high school through research. I would suggest the following however:

1) Put the data acquisition software on floppy diskette rather that on CD-ROM. The benefit to putting the software on floppy is that you can have two different versions of the software to access the same data (see below).

2) Create two different versions of the data acquisition software. It would seem to me that high school, college classes, public education, etc. would mostly be interested in actually viewing the data. I would suggest that you provide software that would allow for easy viewing/zoom viewing of the data, possibly the ability to string several images together into a slide show, and the ability to export the color images to another format (i.e. a screen capture utility that would export the image as a .PCX (or other) image). The export of the color images would allow for the data to be used in many different novel ways. The second version of the software would be similar to that distributed now with some of the recommendations that I and I'm sure others have made.

The problems I see in distributing this database are largely financial. Most CD-ROMs are overpriced in comparison to the cost of manufacturing them. Even the \$500-\$600 charged by the NGDC for other data-sets is probably excessive when compared to the average charges levied by EROS for CDs of its data (average cost \$35). The cost issue is especially important if you wish to attract users in the education and public sectors. Another problem might be the acquisition of rights to publish copyrighted material (which I assume much of the supplemental material that you included to be). I think the inclusion of this previously released material is very important and should be pursued.

18) The database as a whole is not really suitable for this purpose because many of the

data-sets are too experimental, the documentation differs strongly among data-sets, etc.

19) It probably should include the GIS (IDRISI), at least as an option, when others obtain it. Still, its not a terribly user-friendly system for public schools - a good PASCAL programmer could build a windows interface which is particularly useful in such complex applications which are only occasionally used by operators.

20) N/A

21) Technically, I see no problem in making the database available for educational purposes. I think most institutions can afford the necessary equipment, if they don't already own it.

The problems I foresee in making this database available to educational institutions is the extra time and effort it will take for people to learn a GIS in order to access the data. In addition, I think that once people have developed GIS skills to access the data, they will be frustrated by the lack in the availability of spatial data necessary to carry out their specific analyses. But, you've got a start somewhere and I think making this CD-ROM available is a good start.

22) Frankly, I don't foresee great problems. IDRISI is in the process of getting more and more established as a general low-cost GIS, and any IDRISI-user with a PC and a CD-ROM drive can access this database.

23) For educational use we should have all fields at all resolutions.

24) Audiences other than "Highly Technical" will likely have no use for this data nor will they appreciate its value. For those audiences, the data-sets must be "Put into the appropriate contexts and perspectives", i.e., a sort of interactive slide show of global change information.

25) Simply need to ship appropriate access software (or scripts/macros) for large, in-place applications, such as for ARC/INFO). Then let the users integrate it into their GISs.

26) Data may be too raw for non computer-literate to use.

27) N/A

28) For educational purposes the IDRISI software may be appropriate depending on the audience. I do not believe the function of the databases on CD should be educational. Maybe two sets of CD's (one educational and one for research) would be more appropriate. I believe in the effort to make the data useful to "all" possible users the authors of this CD have made the data inaccessible for most GIS users without special programming skills.

29) N/A

30) I see no insurmountable technical problems. However, the data are open to misinterpretation by those remote from the point of construction.

31) 1) Technology: many public education facilities cannot afford the hardware to extract and display this database.
    2) Expertise: appropriate use of these data requires high level of training regarding data limitations, sources, and appropriate uses.

32) N/A

33) N/A

34) N/A

35) N/A

36) N/A

37) Equipment requirements are relatively high (for UK Academics and particularly for

schools). Will need "packaging" -- original slide show diskette was better for a quick look -- some sort of hyper-text menuing/interpretation system but these could be added or tailored by individual users. What about developing a menuing prototype that we can alter? We give 100s of demonstrations a year, including IDRISI, project results, PCX graphs, screen dumps, etc.

XII. WHAT PROBLEMS DO YOU FORESEE IN MAKING THIS DATABASE AVAILABLE TO VARIOUS SCIENTIFIC GROUPS? THIS MAY INCLUDE GLOBAL CHANGE RESEARCHERS, RESOURCE MANAGERS, AND A VARIETY OF SCIENTIFIC GROUPS IN VARIOUS FIELDS.

1) N/A
2) - Advertise its availability in EOS regularly.
   - Make it available over the network.
   - Provide workshops around using the data.
3) The quality of the data-set becomes more important for research issues. If no statements on the quality and limitations are included, the research community will make their own judgement, often based on unrealistic expectations and an unfortunate misunderstanding of the evolution of many of the data-sets.

   Hopefully many of the contributors of current data-sets will present new, better versions in the near future. In the creation of these new versions, an important task for your group will be to guide the creation of the proper, comprehensive and compatible documentation describing all aspects of the data-sets. Only if such documentation is available the data-set should be included in the global compilation.
4) Getting to a common level in terminology and to avoid compromise that hurt the integrity of the data or make it suspect to the data purists.
5) People will like it so much, they'll want even more data.
6) The main problem I would foresee from the wide dissemination of these data is that it could be used without a good understanding of the uncertainties and assumptions associated with each data-set.
7) I suspect that the scientific and technical community will tend to pick up on the availability of this database faster than the academic community. Cost should also be much less of a problem for these groups. In both cases it will simply take time to get the word out.
8) Lack of recognition of the limitations of data sources.
9) I foresee the same limitations as with students, but with less probability for misinterpretation of data.
10) No mention has been made about cost... I would hope that the final product will be as cheap as possible! For teaching purposes in my estimation a price over $1000 would kill the project dead, $100 would mean bonanza sales. Research and business would not have the same financial constraints. Obviously only 1 copy is needed per establishment, but there may be problems with licenses on the basic GIS (very necessarily) distributed with the GED.
11) Only the cost which can't be very high.
12) Probably the largest problems here are hardware software, and personnel concerns pertaining to data accessibility. CD-ROM readers may not be easily accessible to many scientists and managers. Using major libraries (which presumably do have CD-ROM readers) [for] data distribution might help. The large number of GIS systems in

use, and the unavailability of any GIS to some individuals also may constrain accessibility. Finally, in agencies or institutions with centralized GIS functions, my experience is that personnel are already quite busy. Adding mediating use of this database to their responsibilities may not be feasible.

13) I do not think that IDRISI is the way to go. Also, unless access, transfer and display speed on the PC can be improved, I think this group ... would really rather opt for workstation (this system could not be made to work on a DEC workstation either), mini (VAX) or mainframe versions of the system. Finally, the IDRISI menus do not contain an export function. This is necessary not only for GIS file conversion, but for data use in non-GIS settings. At the very least, a standard ASCII file format should be offered as an output option.
14) Storage of files in IBM or DEC readable binary format might discourage individuals or groups that do not use IBM or DEC machines from using the data. We, for example, do not have the proper equipment to use the special GIS CD-ROM access software that was provided on floppy diskette for this review.
15) Similar to above - using data out of context or disregarding extrapolations or limitations in collection and/or compilation. Misunderstanding of these same. Example: Spotted Owl habitat data in the Pacific Northwest released to public and used out of context to influence legislation.
16) I do not foresee any major problems. A wide distribution will allow more interdisciplinary research which is vital for the climate change problem.
17) [see Question #11]
18) Even in this case, some attempt at a more standardized documentation format would help.
19) Probably, the same generic problem and solution as in 11, above.
20) N/A
21) I think global change researchers and resource managers need real time data to carry out their respective analyses appropriately so that the data on the CD may be of limited use to these groups. In designing our model..., we recognized the lack of spatial data needed to develop a complex model so we designed [the model] to be a rather simplistic ecosystem model. Hopefully, as more data-sets become available, more complex models can be developed.

    As for other scientific groups, I think the same problems listed above for educational purposes applies here. Time is a precious commodity so the benefits from the available data-set need to overcome the costs of learning the GIS needed to access it.
22) Users have to figure out how to get the data out from IDRISI and into their application.

    Also, they have to be made aware of the unavoidable limitations (and their spatial pattern) of such data. The really big danger is that users assume this kind of data to be THE TRUTH. I can, however, not give much advice on how this problem can be overcome.
23) Hardly any scientist uses a PC. There needs to be a more generic database that a wide variety of platforms can access. Unix and Fortran on workstations are the most widely used systems of scientists.
24) The acquisition costs must be "low to free". Many scientists will be unwilling to pay for data that is available for free from other sources.
25) Ditto question XI, #25).

26) None
27) N/A
28) Since intense scientific use will undoubtedly uncover QA/QC problems with the data, managing the updates to future revisions will be a challenge. How the updates are selected and applied to improve future releases of the CD will be a major task. In addition, data ownership issues (political, commercial,...) will also have to be overcome if more data is to be made available for distribution.
29) N/A
30) None - except as above!
31) The major issue here is using simple, accessible data format with good documentation. As long as these issues are given continual high priority I foresee few problems.
32) N/A
33) N/A
34) N/A
35) N/A
36) N/A
37) Data quality issues and desire to go back to original sources to avoid (a) other's errors, (b) issues of copyright and publication credit.

    What about an e-mail network on global databases and applications? A quarterly electronic newsletter would help keep in touch!!

APPENDIX C -- IDRISI IMPLEMENTATION

John Kineman and Mark Ohrenschall
National Geophysical Data Center
Boulder, CO

INTRODUCTION

IDRISI is a Geographic Information System for the DOS environment developed for research and education by Dr. J. Ron Eastman of the Clark University Graduate School of Geography. It is maintained and distributed internationally on a non-profit basis, and has an installed base of over 6,000 at the time of this printing. The following instructions are provided as a convenience to IDRISI users, but do not necessarily constitute a recommendation of this system for a given application.

Further information can be obtained from:

The Clark Labs, IDRISI Project
Graduate School of Geography
Clark University
950 Main St.
Worchester, MA 01610-1477
(508) 793-7526
EMAIL: idrisi@ollie.clarku.edu

The Global Ecosystems Database as structured on CD-ROM is intended to be as generic as possible for use in a variety of systems. User software in various operating systems can access the data files directly. Nevertheless, it was considered essential that the database be fully operational in at least one existing GIS, preferably one that is readily and inexpensively available on the widest installed base of computers, the IBM compatible PCs. IDRISI compatibility was chosen for the project to satisfy this requirement and to support distribution for review and data development purposes. IDRISI has thus shared a co-evolutionary relationship with the global database. Through various cooperative efforts between the database developers at NGDC and the software developers at Clark University, considerable changes have been made in IDRISI to make it more compatible with the needs for global database processing and global change studies. Likewise, the database has been revised considerably to reflect advances in GIS technology in general, and as implemented in IDRISI. This interaction has been for research and development purposes only, and does not constitute an endorsement or recommendation for the use of any particular product in conjunction with the database. Every effort has been made to make the database as system-independent as possible.

While some inconveniences still exist for IDRISI users, such as the decision in this version to abandon the concept of providing all data on a single grid size, it is possible that future versions of IDRISI may implement automatic grid expansion and contraction to take advantage of this database design. Likewise, many features of IDRISI may not yet be fully utilized in the digital database, such as the recent ability to track lineage and uncertainty (which require further development of the meta-data). Clearly, for both database and software, considerable potential lies ahead, while in the meantime a basic level of operability is maintained. The current version of the database (Version 1) is compatible with IDRISI 4.0.

An IBM/Intel 80386 compatible computer is recommended for reasonably acceptable performance in IDRISI, especially when working with the full global grids at 10-minute scale. A co-processor (80387) or Intel 80486 compatible computer will greatly improve the performance of IDRISI, especially on some of the more compute-intensive operations. Minimum acceptable performance, however, will require at least an IBM-AT or compatible (Intel 80286) with suitable disk space for processing and storage of a user database, which may be any subset or geographic window of the global files provided. As a practical minimum, display requires at least EGA graphics with 16 colors. VGA graphics will provide better screen resolution. For best results, Super-VGA IBM 8514A compatible graphics will provide 1024 x 768 resolution and 256 colors. 8514A drivers are now available for most popular Super-VGA chip sets, and some graphics boards are hardware compatible. An accelerator card will greatly enhance screen I/O. IDRISI 4.0 takes advantage of coprocessor support, which greatly increases the processing speed of many functions.

Please consult the IDRISI manuals for more information on equipment and set-up options.

GETTING STARTED

Since the software provided on floppy disk (IDRISI Explorer) is fully compatible with IDRISI 4.0, it may be added to the IDRISI directory to provide the extra CD-ROM functions. Refer to the section titled "ACCESS AND EXPLORATION SOFTWARE" in the *User's Guide* for instructions on setup, data exploration, and export/conversion operations using the CD-ROM software.

Using the CD-ROM with IDRISI 4.0 requires some additional steps to allow IDRISI to access the CD-ROM data.

BASIC SET-UP

Since it is not possible to write to a CD-ROM, it is important to establish a working directory on read/write media (e.g., hard drive), and access the CD-ROM data through a data search path. The paths specified in IDRIX have no effect on the rest of IDRISI.

A simple DOS batch program utility (ACCESS.BAT) is provided on the accompanying software floppy disk to help you set an appropriate path for use with IDRISI. This path setting will also work with IDRIX, which should then have its directories set to the working directory (see IDRIX SOFTWARE, page 26 in the User's Guide). The program merely uses the DOS APPEND command to provide access to the resource directories on CD-ROM. You can TYPE out the file as an example of how to set APPEND paths or use the program to set standard path definitions for the CD-ROM. The following procedure uses this .BAT file to get you started quickly with the main database. More detailed information about the APPEND command is given in the section that follows, and in the IDRISI User's Guide.

(1) Install IDRISI according to the installation instructions provided in the IDRISI User's Guide. Be sure to establish a working directory on your hard disk. This is done by simply creating the work directory in DOS, e.g.,

MD D:\WORK1

Next, be sure to initialize the IDRISI environment to this directory using the ENVIRON command in IDRISI. Options 1 and 2 in ENVIRON set the default drive letter and working directory (see your IDRISI manual for further instructions).

(2) Confirm that your CD-ROM reader is properly installed and that the CD-ROM drive has been initialized for access by DOS. You might test this after inserting one of the GED discs into the drive, by executing a DOS directory list (DIR). This should give the following results:

```
DIR R:

 Volume in drive R is  GED-1.0-A
 Directory of R:\

DOCUMENT      <DIR>       04-01-92  12:00p
GLGEO         <DIR>       04-01-92  12:00p
SOURCE        <DIR>       04-01-92  12:00p
        3 file(s)                0 bytes
                                 0 bytes free
```

(3) Insert the software disk into the appropriate floppy drive (e.g., 1.2MB 5 1/4"; or 1.44MB 3 1/2" drive), and change the default directory to that drive, e.g.,

CD A:

(3) Copy the ACCESS.BAT program to your IDRISI directory, e.g.,:

COPY A:\ACCESS.BAT C:\IDRISI

(4) You are now ready to begin by typing "ACCESS" followed by the drive letter of your CD-ROM and an optional path keyword (see below). For example:

ACCESS R: GLGEO

This will establish a data path to the primary database on the CD-ROM.

(5) You should now be setup properly. This can be tested by executing the LIST command at the DOS prompt in your IDRISI directory, resulting in a list of data-sets. If this does not work, execute ENVIRON and check that the drive and path are correctly set to your working directory. If you still have trouble, try the procedures described under "ADVANCED SETUP" below.

ADVANCED SETUP

Setting a Data Path Using the ACCESS Program:

The ACCESS.BAT program can be used by simply typing ACCESS followed by the drive letter for your CD-ROM, followed by an appropriate keywords for the database you wish to access on CD-ROM. For example, any one of three standard data paths can be set by entering the following command line:

ACCESS {drive} [path-option]

where {drive} is the drive letter assigned to the CD-ROM, expressed as a letter followed by a colon (e.g., "F:"), and [path-option] is one of the following:

"GLGEO"	The main GED GLOBAL GEOGRAPHIC database with nested grids -- GLGEO\RASTER, GLGEO\VECTOR, and GLGEO\VALUES directories.
"SOURCE"	Source (non-nested) data-sets for experimenting with integration techniques -- SOURCE\RASTER, SOURCE\VECTOR, and SOURCE\VALUES directories.
"ALL"	All data directories on the CD-ROM

Executing ACCESS without a path option will default to "GLGEO". An incorrect syntax will cause the program to display instructions (i.e., type ACCESS followed by the Enter key).

Setting a Data Path using the DOS APPEND command:

Microsoft or IBM DOS (Disk Operating System) has a rarely used command called "APPEND" (please refer to your DOS manual). APPEND will establish a path that is used whenever programs search for data files (non-executables), just as the PATH command establishes a search path for whenever the operating system searches for executable programs. The software can use this APPEND path to find data files on the CD-ROM, while maintaining a "work" directory on your fixed disk.

<u>APPEND must be initialized with the command APPEND /E prior to any other APPEND command</u>, for example as part of the AUTOEXEC.BAT file. (If the first use of APPEND does not invoke the /E option -- note <u>forward</u> slash -- then data paths cannot be made available to the DOS environment without rebooting the computer.)

The best way to operate within IDRISI initializing APPEND, is to set the applications environment (using the ENVIRON command, as described above) to your local work space on the fixed disk where you will keep all new and temporary files that the software might create during a work session. Then execute the DOS APPEND command (after it has been properly initialized) to set the data search path to the desired directories on the CD-ROM. This path is then stored in the DOS environment (which can be viewed with the DOS "SET" command). For example, the data path for all directories in the "GLGEO" database can be set with the following command:

```
APPEND R:\GLGEO\META;R:\GLGEO\RASTER;R:\GLGEO\VECTOR;R:\GLGEO\VALUES
```

Where APPEND is a DOS command, R: is the drive letter for the CD-ROM (this may be a different letter for different installations), and GLGEO\META, GLGEO\RASTER, GLGEO\VECTOR, and GLGEO\VALUES are the various directories on the CD-ROM, corresponding to the necessary ASCII header files, the raster database, the vector database, and attribute values files respectively (however, the order doesn't matter).

The source data files may be accessed in the same manner, separately or in combination with the above directories:

```
APPEND R:\SOURCE\META;R:\SOURCE\RASTER;R:\SOURCE\VECTOR;R:\SOURCE\VALUES
```

or

```
APPEND R:\GLGEO\META;R:\GLGEO\RASTER;R:\GLGEO\VECTOR;R:\GLGEO\VALUES;
R:\SOURCE\META;R:\SOURCE\RASTER;R:\SOURCE\VECTOR;R:\SOURCE\VALUES
```

The second path above will provide access to the entire database (including the source data files).

Setting the Applications Environment:

It is essential that the software applications environment be properly set to your working drive and directory. This can be done by executing the ENVIRON program. Type ENVIRON (or select from the menu system), and then choose the appropriate options to modify (e.g., the drive and path, line 1 and 2 respectively). The default drive and data path should be set to a specific work space on your fixed disk, not to the CD-ROM itself. This will direct all write operations (results of windowing, conversions, on-screen digitizing, graphs, and temporary files) to the work space. The APPEND data path (described above) will provide the necessary access to the archived CD-ROM data. NOTE that any files in your working directory with the same name as those being accessed on CD-ROM will be found and used first, and the appended files will be ignored.

SET-UP NOTES FOR WINDOWS 3.0 OR 3.1

If you wish to run IDRISI through Windows, do not initialize the DOS APPEND command until you are in Windows. Windows 3.0 seems to have conflicts with APPEND if it is invoked before starting Windows. Instead, enter Windows and open a DOS window. Then initialize APPEND by typing APPEND /E (note forward slash). Then execute the desired APPEND commands or use the ACCESS.BAT program. The APPEND commands should not affect other Windows applications this way, and when you close the Window, the APPEND commands will be canceled.

OPERATIONS

The following are some general comments and special notes. Please consult the IDRISI User's Guide and Technical Reference for detailed information on the use of IDRISI.

ACCESS AND EXPLORATION

All of the functions in IDRIX are available in the full IDRISI package as stand-alone modules. For example, COLOR and COLOR85 function similarly to the display options in IDRIX, and are described under the heading IDRIX SOFTWARE (Page 26 in the User's Guide).

MAP ANALYSIS

IDRISI, as with any raster GIS, realizes its greatest power in combining various digital map ('images') to characterize various user-defined results. Prior to such combinations, however, there is often considerable exploration, analysis, and processing that is required to prepare each data layer. While not attempting to list all the functions available for this step, it is useful to organize a GIS session into these three steps prior to combining data layers. The IDRISI User's Guide provides a discussion of these steps (primarily under the headings, "Core Modules", "Geographic Analysis Ring Modules," "Statistical Analysis Ring Modules," and part of the "Image Processing Ring Modules"). Using many of these functions in IDRISI, the user can not only prepare data layers correctly, but can also produce companion layers that may help in analyzing and tracking error. It is very useful to keep a detailed list of each step in analysis and processing for later reference, and it is essential to carefully think through any anticipated series of processing steps. Most GIS, including IDRISI, are extremely flexible in what they can do, but the price of this flexibility is the ease at which apparently interesting but completely meaningless results can be generated. The power of GIS functions to avoid this is entirely in the user's hands.

CAUTION!

As with any GIS, the user must be aware that coarser-grid data, when represented on a finer grid, 'nested' or not, retains its original spatial definition. For example, if we choose to work on a 10-minute grid using the GVI data, and access a .5 degree data set in an overlay operation, we are combining .5 degree spatial statistics with 10-minute spatial statistics, with results that can only be evaluated in the context of the user's requirements, and the documented statistical meaning of the data (e.g., cell average, point sample, spatial mode, class value, etc.).

If, on the other hand, one chooses to work at a coarser grid than some of the data that will be used, the problem of aggregation is introduced when accessing the finer-grid data. In this case, it is advised to first re-interpret all data-sets onto the coarser grid, using appropriate methods and documentation unique to each data set. A variety of tools should be available in IDRISI, GRASS, or other GISs to accomplish this, again, according to the user's requirements.

DATA-SET Intercomparison AND COMBINATION

Because of limitations in IDRISI's multiple-scale grid overlay capabilities at the time of this release, it is best for IDRISI 4.0 users to expand or contract the data files provided on CD-ROM to matching grid dimensions before undertaking any overlay, intercomparison operations, or other processing functions involving two or more data files. Since IDRISI's EXPAND and CONTRACT functions run in batch mode, this operation is relatively simple, given enough disk space in the work directory. As always, it is the user's responsibility to keep track of the true spatial meaning of the data values, regardless of the scale of representation (see note above). In general, this is best accomplished if data are expanded to the finest grid in the comparison. In this way, coarser data will be correctly represented in terms of geographical objects, even though the pixel density may be greater than needed.

A large number of intercomparison and combination functions are available in the full IDRISI implementation, in addition to those provided in the CD-ROM software. Some extremely useful functions include statistical comparisons such as REGRESS, which provides a regression analysis between two raster maps, or CROSSTAB, which gives a class-by-class tabulation; OVERLAY operations that provide algebraic combinations of two digital maps to produce a third; and other functions that access up to 12 input maps in various kinds of analyses, including classification techniques (MAKESIG, PIPED, MINDIST, and MAXLIKE), principal components analysis (PCA), cluster analysis (CLUSTER); functions in the "GIS , and others. Once again, the IDRISI User's Guide provides a detailed explanation of these combinatorial functions (including "Distance Operators," "Context Operators," "Image Classification," and portions of the "Statistical Analysis Ring").

USING SOURCE GRIDS

The source data files are similar to the other data files in format, but are provided as examples because differences in their spatial and/or temporal sampling presented problems in integrating them with the rest of the database. For example, the 2-minute FAOSOIL data cannot be overlain with 5-minute data, even for comparative purposes, without resampling one of the data-sets. IDRISI can do this using the RESAMPLE command, or by selecting a suitable combination of EXPAND and CONTRACT to achieve a 2:5, or 5:2 grid change. The question of standardizing to the 'nested' grid convention, or implementing other solutions should be explored.

USING OLDER VERSIONS OF IDRISI

Since there are probably a number of IDRISI 3.x users who have not upgraded to IDRISI 4.x for various reasons, a utility, called DOC4TO3.COM is provided with the software to convert IDRISI 4.0 header (.DOC) files to IDRISI 3.2 header (.DOC) files. The program is self-documenting. Also, the IDRISI 3.2 header (.DOC) files are available on floppy disk, if requested.

IDRISI version 3.x and earlier software did not allow the user to set Read-Only status for data files (although it could be set for .DOC files). This was corrected in version 4.0. On most CD-ROM reader installations, especially if accessed via network (i.e., using LAN software) we have experienced no problems with the older IDRISI modules because the CD-ROM or network software intercepts the write protection information; however, users should be aware of this potential difficulty with other configurations and CD-ROM access software.

TERMINATING YOUR SESSION

When done with IDRISI, type "APPEND ;" or reboot to cancel your path assignments. Other software may not work properly with an active APPEND path.

AUTOSCALING WITH FLAG VALUES ("COLOR[85] A" COMMAND)

The current version of color will include flag values in the autoscaling calculation if the flag value is within the data range (most if not all flag values should be outside the data range). By convention in the GED database, the maximum and minimum data values indicated in the .DOC header file refer only to the data values, and do not include flag values that occur outside the data range. Thus, IDRISI COLOR A (autoscaling) will assign the color palette to the data range appropriately, but will either include the flag value in the lowest color category or wrap the color palette to determine a color assignment.

To avoid this problem, you may wish to re-assign the flag values so that they will be given a unique color assignment by IDRISI.

HELP

If you have difficulties implementing IDRISI with the Global Ecosystems Database, you may contact NGDC support at the address indicated under TECHNICAL SUPPORT at the beginning of the *User's Guide*, or contact Clark University at the address given above.

APPENDIX D -- GRASS IMPLEMENTATION

David A. Hastings and John J. Kineman
National Geophysical Data Center
Boulder, CO

INTRODUCTION

The Geographic Resource Analysis and Support System (GRASS) is a public-domain UNIX based Geographic Information System developed for environmental analysis by the U.S. Army. Information about GRASS can be obtained from:

U.S. Army Corps of Engineers
Construction Engineering Research Laboratory
P.O. Box 4005
Champaign, IL. 61824-4005
Telephone: 1-800-USA-CERL ext 220
217-373-7220

GETTING STARTED

A special version of the Global Ecosystems Database CD-ROMs is planned for direct use by GRASS in the UNIX operating system. Meanwhile, import of the data into GRASS from the current DOS/IDRISI structured CD-ROMs is possible using one of two methods. The most convenient method for those who are networked with IBM-PC/DOS compatible machines equipped with a CD-ROM reader, is to access the CD-ROM with the DOS user software provided and convert the database or any portion of it to GRASS format on hard disk using the SUBSET option and choosing a GRASS structured output option (if this feature is implemented in your version of the software). The files will be created in the correct structure and can then be transferred to the GRASS/UNIX system using any binary file transfer method (which will preserve the binary formats created by the conversion process). You may need to edit the resultant "cellhd" files to be compatible with the version of GRASS that you use. See examples given below for ideas on how to do this. Use of the DOS software is described in the *User's Guide* under ACCESS AND EXPLORATION SOFTWARE.

The second method is to access the CD-ROM directly from UNIX. Pre-constructed GRASS "cellhd" (header) files are provided on floppy disk in UNIX-accessible format (available on request). Alternatively, the user may refer to the GRASS Programmer's Manual for the relatively straightforward procedure for developing the required "cellhd" files. While there should be no problem accessing the ISO 9660 format of the CD-ROM and copying data files to your system, problems will arise in the numerical formats of the data because of differences between DOS and UNIX various configurations of hardware and operating systems, and because of the unique data storage types used in GRASS. These problems are discussed in the *User's Guide*, in the section on DATA STORAGE FORMATS.

The 8-bit raster grid files in the database are identical to GRASS cell files. They can be ported directly into a GRASS cell subdirectory and used with the appropriate headers (either taken from the accompanying floppy disk or created using GRASS functions).

The 16-bit integer raster grid files are another matter. GRASS has a non-standard manner of handling data with values over 255 (i.e., two-byte integers) and signed integers. GRASS users will have to convert the 16-bit (two byte integer) data into a GRASS compatible form on import to GRASS.

GRASS databases allow for mixed pixel sizes, geo-referencing, and differing image coverage within a database. Thus GRASS will support the use of all GED data on either their original grids (i.e., in the 'SOURCE' directory) or the 'nested' grids of the GED (i.e., in the 'GLGEO' directory). GRASS will also automatically re-sample a data-set ("on-the-fly") to whatever window and grid size the user specifies. However, this flexible overlay and re-sampling capability of GRASS, though convenient, has a drawback: Without careful attention to detail, the GRASS user may easily misrepresent a data-set through GRASS's automatic (and transparent) interpolation to a user-defined grid. This can be avoided by standardizing the default grid and window to one of the 'nested' grids used in this database.

CAUTION!

As with any GIS, the user must be aware that coarser-grid data, when represented on a finer grid, 'nested' or not, retains its original spatial definition. For example, if we choose to work on a 10-minute grid using the GVI data, and access a .5 degree data set in an overlay operation, we are combining .5 degree spatial statistics with 10-minute spatial statistics, with results that can only be evaluated in the context of the user's requirements, and the documented statistical meaning of the data (e.g., cell average, point sample, spatial mode, class value, etc.).

If, on the other hand, one chooses to work at a coarser grid than some of the data that will be used, the problem of aggregation is introduced when accessing the finer-grid data. In this case, it is advised to first re-interpret all data-sets onto the coarser grid, using appropriate methods and documentation unique to each data set. A variety of tools should be available in IDRISI, GRASS, or other GISs to accomplish this, again, according to the user's requirements.

ACCESS PROCEDURE FROM UNIX SYSTEMS

Please refer to earlier sections of the User's Guide for documentation of the default data structures on CD-ROM. The file formats are simple, and are based partly on de-facto standards of data exchange.

RASTER FILES

With the above cautions in mind, the standard way to import a binary raster grid into GRASS is:

1. Import the grid file into the "cell" subdirectory of your database,
2. Make a cell header file in the "cellhd" subdirectory,
3. Run "r.support" to make other associated files.

The procedure has slight variations:

1. If you are using GRASS 4.0 or a previous version of GRASS,
2. If you are using the "unprojected" (lat-lon) projection (i.e., all data in GLGEO),
3. If the data are 8- or 16-bit binary grids,
4. If you want to use the data from the CD-ROM, or if you prefer to move them to your hard disk.

The discussion below describes how to handle these variations.

IMPORTING AN 8-BIT RASTER DATA FILE INTO GRASS

An 8-bit GRASS grid file is identical to 8-bit binary .img files provided here. To use GED .img files in GRASS, merely move the files to the "cell" file subdirectory for your new GRASS GLOBAL/PERMANENT (or whatever) database, and remove the .img extension. Edit sample cell header files into the "cellhd" subdirectory of your GLOBAL database. Sample "cellhd" file formats are as follows:

For GRASS 3.x, you will probably use the X,Y "unprojected" projection, although a few GRASS 3.x users may have a version that supports the latitude-longitude projection. A sample 3.x header is:

This sample header will work for ALL 10-minute unprojected 8-bit binary raster data files. It is worth making a sample header file called "basic.10min" like this, and place it into the "cellhd" directory. Every time we import a 10-minute 8-bit binary data file into GRASS from the CD-ROM, we give the file a name (such as gvioct86) for the grid file in the "cell" directory. We then go to the "cellhd" directory, execute the command;

cp basic.10min gvioct86

to make the appropriate cell header file for the new data-set.

proj:	0
zone:	-1
north:	0
south:	-1080
west:	0
east:	2160
e-w resol:	1
n-s resol:	1
format:	0
compressed:	0

Global raster data-sets with different raster cell sizes usually have identical cell header files, but for the "e-w resol:" and the "n-s resol:" entries. Instead of "1" for the 10-minute gridded data, we would place 3 for 30-minute grids, 6 for 1-degree grids, 12 for 2-degree grids, .2 for 2-minute grids, etc. All other parts of the cell header file remain unchanged, for 8-bit data. Again, it is worth making sample "basic.2min", "basic.30-min", "basic.1deg". etc., cell header files to use when importing data from the CD-ROM.

For Grass 4.0, you can also use the "unprojected" projection, with a header as shown above. If you use both GRASS 3.x and GRASS 4.0, you may prefer to retain a common georeferencing, especially if you are going to use vector files as well as raster files. However, if you use only GRASS 4.0, you may prefer to use the latitude-longitude projection. A sample 4.0 header in latitude-longitude is:

proj:	3
zone:	0
north:	90N
south:	90S
west:	180W
east:	180E
cols:	2160
rows:	1080
e-w resol:	0:10
n-s resol:	0:10
format:	0
compressed:	0

Note that data at resolutions other than 10 minutes would have to have the "cols:", "rows:", "e-w resol:", and "n-s resol:" categories modified accordingly. A one-degree data-set would have 360, 180, 1:00 and 1:00, respectively, for these parameters.

Note that cell header files have identical names as raster data files that they describe. GRASS separates different types of files into different subdirectories, but keeps the same names for all files that relate to a given data layer. See the GRASS 3.0 or the GRASS 4.0 Programmer's Manual, available from the U. S. Army Corps of Engineers, Construction Engineering Research Laboratory, P. O. Box 4005, Champaign, Illinois 61824-4005 for information on GRASS directory structures.

Alternatively: You do not necessarily need to copy the 8-bit raster data from the CD-ROM, if you prefer to use the CD-ROM as a (slow) hard disk. You can build symbolic links between the files (including full directory structure) on the CD- ROM and the "cell" subdirectory of your GRASS database. You then need only create cell header files as described above, and run SUPPORT as described just below.

Finally: After importing (or linking) the 8-bit raster data to the "cell" directory and creating cell header files, run the GRASS command that completes the ingest of a data layer (in GRASS 4.0 this is r.support; in GRASS 3.x the program is merely called support). Follow the function as prompted. Your data are now imported into GRASS.

IMPORTING A 16-BIT RASTER DATA FILE INTO GRASS

GRASS' 16-bit data format is designed for the "ultimate" in portability between UNIX workstations, not all of which handle integer data in the same way. The GRASS multibyte format is described in the GRASS Programmer's Manuals referred to above.

This format is not "industry standard," and necessitates conversion of the data. The conversion procedure is relatively straightforward.

First, check to see if the integer data must be byte swapped. As a rule of thumb, if your UNIX workstation uses INTEL-like integers (such as some IBM-compatible PCs or Digital Equipment VAX computers running UNIX), no byte swapping is necessary. If your UNIX workstation is based on a Motorola processor, on a RISC processor in a similar family of workstations, you will likely have to swap bytes. If you DO have to swap bytes, run the following command (Check the exact syntax for your computer):

dd if=infile of=outfile conv=swab

where infile is the 16-bit file that you want to convert, such as FNOCMOD on the CD-ROM. In this case, we have copied FNOCMOD to our "cell" directory of our hard disk, calling the output file infile.

Next, create cell header files using the same procedure described under importing 8-bit data files. The only change is in the "format:" line in the cell header files. For 8-bit files this should be 0. For 16-bit files, this should be a 1.

Next, run support (r.support in GRASS 4.0) to make the associated files.

Now, you have a 16-bit data-set that is properly imported into GRASS in every way, but for the conversion from 2-byte integer to GRASS' multibyte format. We now will do this in two steps.

First, we want to ensure that we are working on the entire global database, rather than a local or regional window. Run window (g.region in GRASS 4.0). Set the window/region to global. This can be done in many ways. If your "default_wind" file (describing your default working window/region) is already set to global, you can reset your working window/region to your default window. Otherwise, you can manually edit the window/region.

Second, you run r.mapcalc (in GRASS4.0, Gmapcalc in Grass3.x) with the following syntax:

r.mapcalc 'outfile = if(infile-32768,infile-65536,infile,infile)'

Where infile is the 16-bit integer raster data file from the CD-ROM (byte-swapped if necessary), and outfile is the file converted to GRASS multibyte format. Of course, you can name the output file anything that you would like. Since, in our example, we were importing FNOCMOD, we would probably want to use this name again for our final GRASS version of the data.

This procedure may not work for some UNIX workstations.

VECTOR FILES

The vector data provided here are in .vec files, with .dvc header files. The .vec files are in ASCII, and are akin to the files produced by digitizers. As such, they should be easily modifiable in a word processor, or in a database management system. Converting the data in this way, you should be able to match an ASCII vector file or digitizer file format, if your GIS handles vector or digitized data. We have converted these files to GRASS, and to OSU MAP-for-the-PC format using a full-fledged word processor, such as Word Star. GRASS takes ASCII data in a well-documented format and converts them to binary vector files. These files can be rasterized, or used in GRASS in their vector format. OSU MAP-for-the-PC takes ASCII digitizer files and rasterizes them.

A word of caution. IDRISI vector files use the bottom-left corner as the origin, whereas IDRISI raster files use the top-left corner as the origin. You therefore may need to subtract 1080 from the Y coordinate, then multiply the Y coordinate by -1 to convert the coordinates. This can be done with a word processor that permits mathematical operations on numerical data, and which has a column mode. Otherwise, a word processor can reconfigure the vector file format (from its two interfingered line formats to a uniform four-column format), and a database management system can convert the coordinates (with a reconversion to your GIS's vector file format in the DBMS or word processor).

The vector data on the CD-ROM, are currently all line segments with the following format:

245	4
6267	622
6270	623
6272	625
6304	601
248	2
6427	1021
6390	1064
0	0

This file has two line segments. The first has attribute class value 245 and is described by four X,Y (or horizontal position, vertical position) pairs. The second has attribute class value 248 and is described by two X,Y pairs. The bottom line (two zeros) marks the end of the file.

The ASCII version of a GRASS vector file resides in the **dig_ascii** subdirectory of the GLOBAL (or whatever) database. It has the format shown below;

where 'L' stands for line segment ('A' would denote a polygonal area). The number to the right of the 'L' denotes the number of Y,X (vertical position, horizontal position) pairs that describe the line segment.

L 4	
622	7267
623	6270
625	6272
601	6304
L 2	
1021	6427
1064	6390

The coordinates must be in the same units and projection as your raster database. This may entail conversion from the units used on the CD-ROM. For example, for CD-ROM vector files stored in arc-minutes of latitude and longitude, you will be OK, if your raster database also uses multiples of arc-minutes. If you need to convert units, many word processors or database management systems can import, convert, and re-export data from these ASCII files fairly easily.

Note then, to convert the data from the CD-ROM to GRASS, the attribute class value is changed to L or A as appropriate, and the columns of X,Y pairs are exchanged to make the Y,X pairs of the file in the GRASS dig_ascii subdirectory. We assign the attribute class value as follows:

GRASS stores the attribute class value not in the dig_ascii subdirectory, but in a separate **category** file (with the same name) stored in the dig_att subdirectory. This file has the following format:

L	622	6272	245
L	1062	6382	248

This is in the format of feature identifier (L or A), Y coordinate, X coordinate, and attribute. If a line segment in the vector file has no associated category in this file, it is considered unlabelled. This means that during a vector-to-raster conversion, such areas will be converted to category zero, and such lines will be ignored.

The L stands for a line segment, the Y and X positions are markers (which must be inside a polygonal area, or closer to a line segment than to any other line segment but not at a node.

Given the files in the dig_ascii and dig_att subdirectories, run v.import (in GRASS 4.0, import.to.vect in GRASS 3.x), answering the prompts, to convert the data to binary vector files that GRASS uses in its processing.

HELP

If you have difficulties implementing the database in GRASS, you may contact NGDC support at the address indicated under TECHNICAL SUPPORT at the beginning of the *User's Guide,* or contact CERL at the address given above.

APPENDIX E -- DOCUMENTATION TEMPLATE DEFINITIONS

John J. Kineman
National Geophysical Data Center
Boulder, CO

DATA-SET DESCRIPTION

Each data-set has its own chapter in the documentation. These chapters include relevant information provided by the contributor or other information sources. At the beginning of each chapter there is a structured Data-Set Description that provides a **Data-Set Name,** and summary information about the **Source, Original Design, Integrated Data-Set, Primary References,** and **Additional References;** and lists of the **Data-Set Files, Reprint Files,** and **Source File Examples** (if provided), for each data-set, exactly as the files appear on the CD-ROM. The purpose of this section is to describe everything that comprises the complete data-set as provided here, including key documentation references; and to give technical and statistical information about the data structure.

The information categories in the **DATA-SET DESCRIPTIONs** are defined as follows:

Data-Set Name: Name of the data-set as provided and documented in the GED. The same name appears in the Table of Contents and data-set chapter title, although preceeded by the Principal Investigator and Analyst (if appropriate). The format of the chapter titles is {PI} and {data-set name} in italics, preceded by the analysts' names, if the data have been significantly re-worked.

Principal Investigator(s): The principal scientist(s) responsible for the actual numerical or classed data values represented in the data-set, and/or the principal institution, if relevant.

SOURCE

This section refers to the source data-set acquired for this project, documentation references, and the full lineage prior to integration into the GED. This information clearly identifies the version of the data used and gives proper citation of the actual numerical data (regardless of format, media, etc.) and its principal investigators.

Source Data Citation: Citation of the particular data-set used as a source in the GED. This is not a literature reference, but a citation of the source version of the digital data-set. The format of this citation is: {Principal Investigator}. {Availability date}. {Data-set description or name, including geographic and temporal coverage}. {"Digital" or "Analog"} {type, e.g., "Raster," "Vector," "Map," etc.} Data on a {cell size, if standard} {projection, e.g., "Geographic (lat/long)"} {grid dimension} grid. {City, State}: {Institution /publisher}. {number of files} on {media}, {size}. For example:

> NCDC Satellite Data Services Division. 1985-1988. Weekly Plate Carreé (uncalibrated) Global Vegetation Index Product from NOAA-9 (APR 1985 - DEC 1988). Digital Raster Data on a Geographic (lat/long) 904x2500 grid. Washington DC: NOAA National Climatic Data Center. 199 files on five 9-track tapes, 425MB.

Contributor: Person or institution responsible for disposition of the data, and for releasing the data into the public domain. In most cases this will be the PI, however in some cases data have corporate or institutional ownership prior to release, or are released through an indirect route. This field also provides contact information, if available.

Distributor(s): Data or research center(s) with official responsibility for distributing earlier versions of the data-set, up to and including the source used for the GED. To work cooperatively with other distribution centers, NGDC will refer requests for source data to these official distribution points. Information for each of the institution abbreviations referenced in this field follows this section.

Vintage: The approximate date(s) of the project (digitizing work) creating the data-set represented in the current release. This will preceed the publication date and may be later than the period of the data. Also notes continuing projects.

Lineage: Chronological list of the previous versions of the data-set from the original data up to the source version used for the GED (i.e., does not include integration into the GED, which is described in the next section). Sufficient information should be given to identify the previous versions and their source.

ORIGINAL DESIGN

This section refers to the nature of the data before integration into the global database structure. This is important information when considering the reliability and potential application of the data-set to new problems, perhaps not foreseen by the original investigators. It is also useful for those who wish to track back to the original data for quality control or verification purposes, or to compare with the integrated form of the data, which may bear changes that are important for a given application.

Variables: The specific environmental/thematic measurements included in the data-set, the units used, and the numerical or class precision (class precision is a qualitative or descriptive indicator, e.g., "species", "major types", "primary/secondary classes", etc.)

Origin: Description of instrument, data sources, and/or method of original investigation or observation.

Geographic Reference: The coordinate system or projection, and projection parameters (i.e., grid orientation, origin, central meridian, zone, etc.), for the original data-set.

Geographic Coverage: The geographical limits covered by the data-set.

Geographic Sampling: The original spatial interval or sampling resolution of the data, the type of spatial object, and the numerical statistic (i.e. Vector point, line, or polygon unit with various attributes; or Grid point sample, cell average, mode, etc.).

Time Period: The time period represented in the data-set. In the case of time series, this indicates the beginning and ending of the data series. In the case of long-term averages, this field indicates the period from which data were combined.

Temporal Sampling: The original time interval or sampling resolution of the data and the type of statistic (i.e. discrete sample, peak values, running average, typical or average period, etc.).

INTEGRATED DATA-SET

This section describes the integrated version of the data-set, as provided by this project on CD-ROM. While every attempt has been made to preserve the full content and nature of the original data, some alterations may have been necessary to achieve a common structure and geographic registration. For raster data, this may involve interpolation to one of the conventional "nested" grids, or various forms of re-registration of the grid, or perhaps both. To aid in assessing the appropriateness or accuracy of interpolation methods, all interpolated data-sets have corresponding examples of the original form of the data in the SOURCE directory. The original grid representation is clearly documented in the preceding section (Original Design), and interpolation methods are indicated in this section (Data Integration). The user must understand that values represented on a new grid still retain the statistical meaning from their original grid. The information provided here is thus important for proper interpretation and use of the data-sets.

Data-Set Citation: The recommended way of referring to the integrated data-set as published in the GED. This is not a literature reference, but a unique citation for the digital data-set that distinguishes it from other versions, including the source. The format is: {Principal Investigator}. {current publication date}. *{data-set name,* including geographic and temporal coverage}. "Digital {type, e.g., "Raster," or "Vector"} Data on a {cell size, if standard} {projection, e.g., "Geographic (lat/long)"} {grid dimension} grid. In: *Global Ecosystems Database Version 1.0: Disc {disc number},* Boulder CO: NOAA National Geophysical Data Center. {number of independent spatial layers and attributes} on {media}, {size}. For example:

> NGDC. 1992. *Monthly Generalized Global Vegetation Index from NESDIS NOAA-9 weekly GVI Data (APR 1985 - DEC 1988).* Digital Raster Data on a 10-minute Geographic (lat/long) 1080x2160 grid. In: *Global Ecosystems Database Version 1.0: Disc A.* Boulder, CO: NOAA National Geophysical Data Center. 45 independent and 24 derived single-attribute spatial layers on CD-ROM. 190MB.

Analyst(s): Individual(s) responsible for the data processing and integration methods resulting in the integrated form of the data-set provided here.

Projection: The coordinate system or projection, and projection parameters (i.e., grid orientation, origin, central meridian, zone, etc.), for the integrated data-set.

Spatial Representation: The spatial interval (between grid or vector values) and other spatial characteristics of the integrated version of the data-set, including the type of spatial object, and the numerical statistic (i.e. vector point, line, or polygon unit with various attributes; or Grid point sample, cell average, mode, etc.). This may differ from the original sampling design due to registration differences or aggregation from finer resolution into a standard grid dimension. If re-sampling was performed, this field indicates the method used.

Temporal Representation: The time step and other temporal characteristics of the integrated version of the data-set. This may differ from the original sampling design due to phase differences or aggregation from finer intervals into standard time sequences. If temporal re-sampling was performed, this field indicates the method used.

Data Representation: The form of representing the data values in the integrated data-set, including any numerical type conversion or re-classification that was performed. Usually this will involve only type conversions, although occasionally actual numerical changes may have been necessary. The units and precision

	are also indicated (e.g., "Real numbers expressed to .001 inches/month," or "Byte integers representing units of 0.1 degree," or "Two-byte integers representing units of meters above sea-level, rounded to the nearest 30 meters").
Layers and Attributes:	Spatial Layers and Attributes. The number of geographically different data layers (i.e., independently distributed spatial data layers) represented, and the number of associated attributes, regardless of file structure. For example, multiple data mapped on political units may in reality have only one geographic distribution but many attributes.
Compressed Data Volume:	The volume in bytes of a compressed version of the data files. This gives a more accurate estimate of scientific information content than actual file sizes that are format dependent and may reflect inefficient storage methods for the sake of access. The volume estimate is obtained using PKZIP(-es), a commonly available data compression program.

PRIMARY REFERENCES

This section lists references that are intended to be the primary literature reference for documentation of the data-set. These may include unpublished documents and published articles. The references may be divided into sub-categories relating to the various data-set elements. One of the goals of this compilation is to provide reprints of all Primary References in addition to this manual. Due to availability and copyright restrictions, however, some documents may not have been available at the time of publication. Articles that are labeled with an asterisk (*) are reproduced as scanned image-files (in PCX format) on the CD-ROM.

[sub-categories]	The primary published or unpublished document(s) associated with the production or release of the data. If a literature publication, the article is one that was intended as documentation, and which may or may not describe related research. A copy of the primary reference is provided with this manual. Published articles are provided in scanned (bit-mapped) digital image format, to ensure accurate reproduction.

ADDITIONAL REFERENCES

This section lists key references to the nature and/or application of the database, or other information that may be especially useful to the user. The references may be divided into sub-categories relating to the various data-set elements.

[sub-categories]	References to key publications that are relevant to the nature or use of the data-set.

DATA-SET FILES

This section lists the location, names, number, and sizes of all the digital files that comprise the complete data-set, as provided on the CD-ROM. This information is divided into the following file categories:

Spatial Data:	spatially distributed data files (*.IMG and *.VEC)
Attribute Data:	tabular ("values") files (*.VAL)
Headers:	ascii header ("document") files (*.DOC, *.DVC, and *.DVL)
Palettes:	Color palette assignment files (*.PAL)
Time Series:	Time series files (*.TS)
Disk Volume:	Total size of all data-set files on disk, in bytes.

REPRINT FILES

A complete listing of all bit-mapped (scanned) image files of previously published documentation articles. These files are provided in PCX format for users who have appropriate software to display them. They are the full-resolution digital versions of the documentation articles that appear in the individual Data-Set chapters in the printed and digital versions of the *Documentation Manual*. Each file contains one page of the scanned article.

SOURCE EXAMPLE FILES

Listing of all source data files provided as examples for comparison and experimentation. Some data-sets required changes, such as re-gridding, to fully integrate them with the database. Although the integration methods are described, these example source files can be used to verify the results, or to test other methods. The location, names, number, and sizes of all files is given.

DATA FILE DESCRIPTION

The **Data File Description** section refers to the actual data files (after processing) that are included in the database, providing technical information contained in the ASCII header files stored under the META sub-directory on the CD-ROM. Each element of the database (i.e., unique spatial-temporal variable or theme) contains data file-pairs, or a series of file-pairs, each consisting of a spatial data file and its corresponding header file. For *Raster* (i.e., 'image') data, these have the file extensions ".IMG" and ".DOC" respectively. *Vector* data and header file extensions are ".VEC" and ".DVC" respectively. A data element may include *Attribute* data for re-classing or re-labeling the data file. Attribute data files and their corresponding header file extensions are ".VAL" and ".DVL" respectively. A data element may also contain color palette (".PAL") files, and time-series (".TS") files.

Version 1.0 of the database employs the Idrisi 4.0 file structure that was jointly developed with Clark University (described in section IV of the *User's Guide*, in the section titled *Database Structure*). Other formats can be produced from these files using the accompanying conversion software or user-developed programs.

The file descriptions are organized as indicated below. To avoid redundancy in the file descriptions, only the first header of a series is shown, followed by a table of "series parameters" that show only what changes through the series.

DATA ELEMENT: The variable or theme represented in this part of the data-set, and the units used.

STRUCTURE: Raster or Vector topology (e.g., nested grid-cell or grid-point, arc/node vector-line, etc.)

SERIES: Number and type of data series (e.g., "45 month time-series", etc.)

SPATIAL DATA FILES: Description of the raster or vector data files, as provided in the raster or vector header file, or first header file from a series, exactly as it appears in the META directory on the CD-ROM, including legends. If a series, the header file example will be followed by a table of Series Parameters, showing all parameters that change in the multiple header files of a data file series (e.g., titles, min/max values, etc.).

ATTRIBUTE DATA FILES: Description of attribute data files, as provided in the header file associated with an attribute (i.e., tabular 'values') file, as represented on the CD-ROM, including legends. As above, multiple files may be represented with a table showing only those parameters that are different from the example file.

NOTES: Any notes or additional information for the data element, for example notes about any color "palette" files or time-series files included with the data-set.

DATA INTEGRATION AND QUALITY

This section provides a narrative description of methods and significant procedures used in the integration process, such as re-gridding, re-projecting, registration changes, temporal compositing, etc. In general, as little change as possible is made from the original data; however, to achieve comparable data structures from source data with unstandardized formats, changes are necessary for some data-sets. The philosophy of long-term development, as conventions become established, is that major changes should be performed by the original investigators, if possible. This section also includes any information on data quality issues as a result of quality assessment work.

*U.S. GOVERNMENT PRINTING OFFICE:1995-673-018/00076