COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 99-2	FOR NSF USE ONLY
NSF 99-29 05/17/99	NSF PROPOSAL NUMBER
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.) KDI/KN	

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)	SHOW PREVIOUS AWARD NO. IF THIS IS	IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY?
956006145	☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	YES ☐ NO ☒ IF YES, LIST ACRONYMS(S)

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE
University of California-Santa Barbara	c/o Contracts and Grants 5220 Cheadle Hall Santa Barbara, CA. 931069530
AWARDEE ORGANIZATION CODE (IF KNOWN) 0013201000	
NAME OF PERFORMING ORGANIZATION, IF DIFFERENT FROM ABOVE	ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE
PERFORMING ORGANIZATION CODE (IF KNOWN)	

IS AWARDEE ORGANIZATION (Check All That Apply)
(See GPG II.D.1 For Definitions) ☐ FOR-PROFIT ORGANIZATION ☐ SMALL BUSINESS ☐ MINORITY BUSINESS ☐ WOMAN-OWNED BUSINESS

TITLE OF PROPOSED PROJECT **A Knowledge Network for Biocomplexity: Building and Evaluating a Metadata-based Framework for Integrating Heterogeneous Scientific Data**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PREPROPOSAL NO., IF APPLICABLE
$ 2,979,726	36 months	09/01/99	9975978

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW

☐ BEGINNING INVESTIGATOR (GPG 1.A.3)
☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.D.1)
☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG II.D.10)
☐ NATIONAL ENVIRONMENTAL POLICY ACT (GPG II.D.10)
☐ HISTORIC PLACES (GPG II.D.10)
☐ SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.D.12)
☐ GROUP PROPOSAL (GPG II.D.12)
☐ VERTEBRATE ANIMALS (GPG II.D.12) IACUC App. Date ____
☐ HUMAN SUBJECTS (GPG II.D.12)
Exemption Subsection ____ or IRB App. Date ____
☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES
____
☐ FACILITATION FOR SCIENTISTS/ENGINEERS WITH DISABILITIES (GPG V.G.)
☐ RESEARCH OPPORTUNITY AWARD (GPG V.H)

PI/PD DEPARTMENT	PI/PD POSTAL ADDRESS
Marine Science Institute	Marine Science Institute University of California Santa Barbara, CA 93106 United States
PI/PD FAX NUMBER 805-892-2510	

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address
PI/PD NAME Omer J Reichman	Ph.D.	1974	805-892-2500	reichman@nceas.ucsb.edu
CO-PI/PD James W Brunt	M.S.	1988	505-277-9342	jbrunt@sevilleta.unm.edu
CO-PI/PD John J Helly	Ph.D.	1984	619-534-5060	hellyj@sdsc.edu
CO-PI/PD Matthew B Jones	M.S.	1994	805-892-2508	jones@nceas.ucsb.edu
CO-PI/PD Michael R Willig	Ph.D.	1982	806-742-2590	cmmrw@ttacs.ttu.edu

CERTIFICATION PAGE

Certification for Principal Investigators and Co-Principal Investigators:

I certify to the best of my knowledge that:

(1) the statements herein (excluding scientific hypotheses and scientific opinions) are true and complete, and
(2) the text and graphics herein as well as any accompanying publications or other documents, unless otherwise indicated, are the original work of the signatories or individuals working under their supervision. I agree to accept responsibility for the scientific conduct of the project and to provide the required progress reports if an award is made as a result of this application.

I understand that the willful provision of false information or concealing a material fact in this proposal or any other communication submitted to NSF is a criminal offense (U.S.Code, Title 18, Section 1001).

Name (Typed)	Signature	Social Security No.*	Date
PI/PD **Omer J Reichman**		SSNs are confidential and are not displayed *ON FASTLANE SUBMISSIONS*	
Co-PI/PD **James W Brunt**			
Co-PI/PD **John J Helly**			
Co-PI/PD **Matthew B Jones**			
Co-PI/PD **Michael R Willig**			

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the individual applicant or the authorized official of the applicant institution is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding Federal debt status, debarment and suspension, drug-free workplace, and lobbying activities (see below), as set forth in Grant Proposal Guide (GPG), NSF 99-2. Willful provision of false information in this application and its supporting documents or in reports required under an ensuring award is a criminal offense (U. S. Code, Title 18, Section 1001).

In addition, if the applicant institution employs more than fifty persons, the authorized official of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of Grant Policy Manual Section 510; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflict which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Debt and Debarment Certifications

(If answer "yes" to either, please provide explanation.)

Is the organization delinquent on any Federal debt?	Yes ☐	No ☒
Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency?	Yes ☐	No ☒

Certification Regarding Lobbying

This certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding $100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding $150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, ''Disclosure Form to Report Lobbying,'' in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME/TITLE (TYPED) **Dorothy C. Hall**		

TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER
805-893-4034	**hall@omni.ucsb.edu**	**805-893-2611**

*SUBMISSION OF SOCIAL SECURITY NUMBERS IS VOLUNTARY AND WILL NOT AFFECT THE ORGANIZATION'S ELIGIBILITY FOR AN AWARD. HOWEVER, THEY ARE AN INTEGRAL PART OF THE INFORMATION SYSTEM AND ASSIST IN PROCESSING THE PROPOSAL. SSN SOLICITED UNDER NSF ACT OF 1950, AS AMENDED.

A Knowledge Network for Biocomplexity: Building and Evaluating a Metadata-based Framework for Integrating Heterogeneous Scientific Data

Project Summary—Complexity is an inherent property of living systems that arises from direct and indirect interactions among the earth's physical, chemical, and biological components. Biocomplexity includes the structural and functional attributes of dynamic systems that arise at all levels of biological organization, including individuals, populations, and communities. Importantly, ecological components of biocomplexity (e.g., biodiversity, ecosystem services) are in crisis, and are undergoing potentially irreversible changes in the face of rapid human population growth and economic development. Wise stewardship, based on all available scientific knowledge concerning these natural systems, is essential.

Data—Catalyzed by these societal concerns, and facilitated by technology advances, scientists focused on complex ecological systems have generated an explosion of ecological and environmental data. When integrated with data from other disciplines (e.g., meteorology), these data have the potential to greatly enhance understanding of biocomplexity. However, broad-scale and synthetic research is stymied because these data are largely inaccessible due to their spatial **dispersion**, extreme structural and semantic **heterogeneity,** and **complexity**.

Knowledge Networking—We propose to integrate the distributed and heterogeneous information sources required for the development and testing of theory in ecology and its sister fields into a standards-based, open architecture, **knowledge network**. The network will extend recent advances in metadata representation to provide conceptually sophisticated access to integrated data products drawn from distributed, autonomous data repositories. In addition, the knowledge network will include advanced tools for exploring complex data sets from which multiple formulations of hypotheses can be tested.

The existence of such a network will lead to broadened understanding of biocomplexity and ecological systems, and allow the application of that understanding to societal issues. In developing this network, we will create a new community of environmental scientists who will be able to focus on complex, multi-scale problems that, to date, have proven to be intractable. We will perform foundational research in computer science and informatics to create new tools for discovering, retrieving, interpreting, integrating, and analyzing data from these diverse sources. Our prototype network will be useful across a variety of disciplines and will provide a basis for the growth of multidisciplinary research groups focused on biocomplexity.

Collaboration—To accomplish these goals, we have created an intellectual consortium that comprises the National Center for Ecological Analysis and Synthesis (NCEAS), the Long-Term Ecological Research Network (LTER) and the San Diego Supercomputer Center (SDSC). Our partnership has a successful history and includes (1) advanced expertise in ecology, informatics, and computer science, (2) a comprehensive understanding of the critical obstacles that data heterogeneity and dispersion create for advancing synthesis and understanding, and (3) strong commitments to addressing those obstacles that deter broad-scale and synthetic analyses.

Impact—The results of the proposed research will have broad implications for our ability to understand and manage sustainably the complex ecological systems and biological resources on which all humans depend. Information on biocomplexity is voluminous and complex, but currently is inaccessible to research scientists and policy makers. The intellectual advances in information science that we propose will, for the first time, provide an accessible infrastructure for identifying, integrating, managing, and, ultimately, synthesizing the nation's ecological and biodiversity information resources.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.C.

*Proposers may select any numbering mechanism for the proposal, however, the entire proposal must be paginated. Complete both columns only if the proposal is numbered consecutively.

Project Description

‘Biological information about biodiversity and ecosystems is among the most complex scientific data to manage electronically, yet it is vitally important that we do so. There are intellectual challenges in the area of biodiversity information analysis, synthesis, presentation, validation and long term storage that require considerable information science and computer science research and infrastructure “—**Panel on Biodiversity and Ecosystems, President’s Committee of Advisors on Science and Technology [45]**

I. RATIONALE

Complexity is an inherent property of living systems that arises from direct and indirect interactions among the earth's physical, chemical, and biological components, of which humans are an integral constituent. Biocomplexity includes the structural and functional attributes of dynamic systems that arise at all levels of biological organization, including individuals, populations, and communities. Importantly, ecological components of biocomplexity (e.g., biodiversity, ecosystem services) are in crisis, and are undergoing potentially irreversible changes in the face of rapid human population growth and economic development. Wise stewardship, based on all available information concerning these natural systems, is essential and requires a multidisciplinary approach based on understanding biocomplexity in all of its ecological manifestations.

The last decade has witnessed a tremendous explosion of ecological and environmental data, catalyzed by societal concerns and facilitated by advancing technologies. These data have the potential to greatly enhance understanding of biocomplexity. However, broad-scale or synthetic research is stymied because data are largely unorganized and inaccessible as a consequence of their tremendous **heterogeneity, complexity,** and **spatial dispersion** in many separate repositories. In addition, information required to answer ecological questions derives from data in multiple disciplines, ranging from geochemistry to climatology, and including all branches of life sciences. Thus, a comprehensive effort to promote knowledge networking through the quantitative integration and synthesis of biocomplexity data will incorporate many of the issues surrounding complex adaptive systems (see special features “Complex Systems” and “Complex Adaptive Systems in Ecosystem Science”) [14-15] as well as new developments in computer science research.

Among the most applicable of these areas of computer science is research on integrating distributed, heterogeneous, and autonomous databases. Although substantial progress has been made in schema integration for heterogeneous databases [55][4], new challenges have surfaced for information integration with the advent of data warehousing and the Web. Indeed, recent surveys dealing with information mediation for data warehouse creation have pointed out the difficult issues that still remain for integrating heterogeneous sets of distributed, autonomous data [56]. These issues apply directly to the types of multidisciplinary biocomplexity data that we have described above. Fortunately, extreme information heterogeneity is an active research area in the web and digital library communities. Recent advances in schema integration tools such as the Resource Description Framework Schema [8] and XML Schema specification [5-6] create exciting new opportunities for data integration and synthesis.

By building on these opportunities in informatics research , we propose to develop a prototype knowledge network for understanding biocomplexity that will enable integration of data drawn from diverse disciplines such as ecology, meteorology, and geochemistry, and their subsequent synthesis into useful knowledge about biocomplexity.

II. MISSION

The *mission* of this project is to enable scientists from a wide range of disciplines to address biocomplexity questions by providing for discovery, retrieval, interpretation, integration, and analysis of heterogeneous and distributed information about biodiversity and the earth's ecosystems, and directly addresses the concerns detailed by PCAST [45]. We propose to accomplish this mission by achieving the goals of:

A. **Building the Knowledge Network for Biocomplexity: Research and Prototype Development.** Our approach includes development and research in three domains: data, information, and knowledge. The technological research within the **data domain** provides an integrated network infrastructure with access to distributed data sources. Research within the **information domain** involves using cutting-edge technologies for structuring metadata to enable powerful semantic querying, data integration, and analysis. The **knowledge domain** comprises development of a probability-based analytical tool with a natural language interface that formally models alternative formulations of hypotheses based on available data (see Section IIIA).
B. **Using and Evaluating the Knowledge Network: Biocomplexity Research.** To demonstrate the ability of the network to substantially advance ecological understanding, we will explore an overarching question in ecology, the relationship between diversity and ecosystem function. Closely associating the development of the knowledge network with this ecological investigation will provide critical feedback into the nature and design of the biocomplexity network (see Section IIIB).
C. **Education, Outreach, and Training.** We take advantage of existing partnerships with repositories of ecologically-relevant information (e.g., Organization of Biological Field Stations, California Nature Reserve System, Long-Term Ecological Research (LTER) Network) to expand the domain of data searches and the scale of core ecological research. In addition, we will disseminate the computer science and informatics products of our research to potential users of our prototype network (scientists, graduate students, policy makers) in ecology and other disciplines (see Section IIIC).

III. INTEGRATING RESEARCH, DEVELOPMENT, AND EDUCATION

The knowledge network that we propose will result from a collaboration among informatics, computer, and ecological scientists. In addition to generating the first Knowledge Network to promote synthesis regarding ecological dimensions of biocomplexity (see Section III A), our efforts will catalyze the development of a new generation of multidisciplinary scientists through investigation of a profoundly important ecological question concerning the relationship between biodiversity and ecosystem function (see Section III B). Moreover, we will support and engage a large cohort of graduate students in the process of scientific discovery through a series of graduate seminars that utilize site-specific data to address scale-dependent questions concerning the relationship between biodiversity and ecosystem function, and then extend those analyses across multiple biomes through use of the Knowledge Network (see Section III C).

Network development, research in computer science, ecological research concerning biocomplexity, and educational activities are purposefully linked in the proposal. The Knowledge Network will provide a testbed for integrating multidisciplinary, multi-scale data for addressing critical environmental questions. The efficient discovery of new ecological insights from this system will provide validation of the Network. Similarly, advances in computer science research involving probabilistic testing of hypotheses will guide ecological research and accelerate progress in understanding complex phenomena in general. Finally, research products from a series of graduate student seminars will supplement the data within the Knowledge Network, while simultaneously training students in its use.

A. Building the Knowledge Network for Biocomplexity: Research and Prototype Development

We will improve the ability to discover, retrieve, interpret, integrate, and analyze biocomplexity information in the Nation's repositories of data by designing and implementing a distributed knowledge network that facilitates the transformation of data to usable information, and ultimately to scientific knowledge. To succeed, we must address several important technical challenges:

- widely **distributed ecological data** sources;
- extreme **structural and semantic heterogeneity** of ecologically relevant data;
- lack of methods for **automating scientific hypothesis testing**.

We will accomplish this research and development at three sites with Internet 2 connections, including the LTER, NCEAS, and SDSC. In addition, an **Informatics Working Group** will be recruited to provide expert contributions to the techniques and approaches proposed here. This working group will consist of 15 persons drawn from the fields of informatics, ecology, computer science, and earth science. In the remainder of this section, we discuss the technical strategies for addressing each of these three challenges. Section A.1 discusses the development of a data-oriented network that facilitates access to distributed data and data discovery via a standardized approach to metadata representation. Section A.2 proposes mechanisms that enable the interpretation and integration of heterogeneous scientific data sets through machine-mediated schema integration. Section A.3 discusses the extraction of knowledge from this information store through the semi-automated modeling of scientific hypotheses (figure 1).

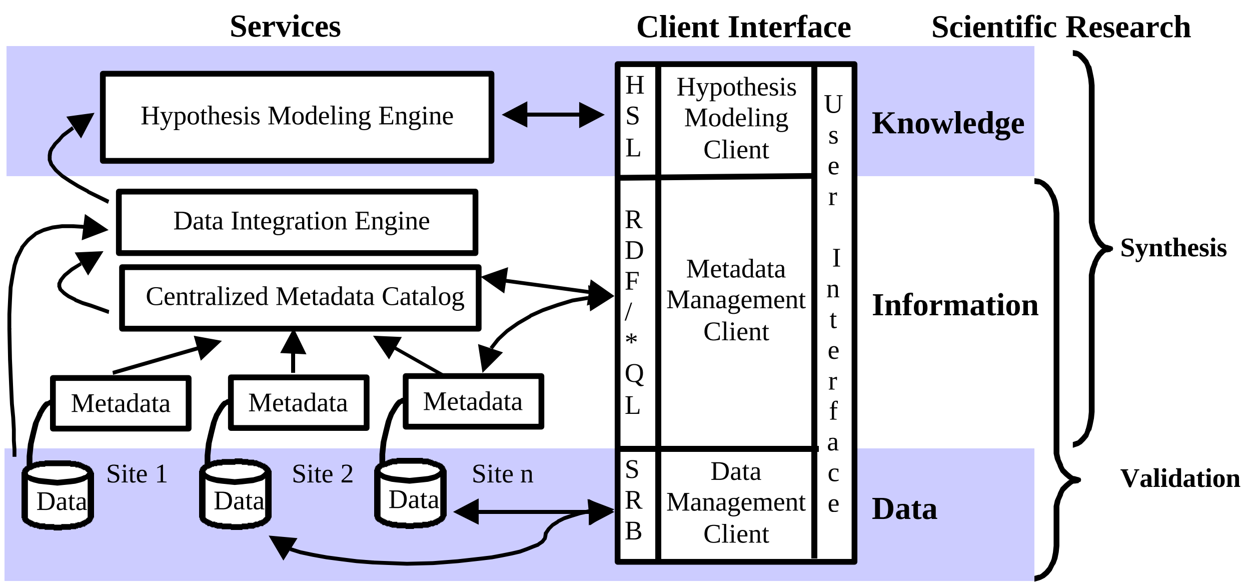

Figure 1: Principal components of a knowledge network prototype: a set of distributed data repositories and their associated metadata catalogs; a centralized metadata search service that allows unified searching across multiple metadata types; a data integration engine that uses high-level metadata to integrate heterogeneous data; and a hypothesis modeling engine that allows for formal evaluation of scientific hypotheses. Scientific activities such as Validation and Synthesis (Expansion, Amplification, and Extrapolation) are explained in section IIIB.

A.1. Data: Unified access to distributed biocomplexity data

Challenge—The foundation of this prototype knowledge network will be the ability to access highly **distributed ecological data** through a unified network interface. We will accomplish this by extending current research on digital library collections to the domain of scientific data access [10-11]. The typical digital library partitions information into a set of information bearing objects (IBOs), each of which is described by metadata that provides for discovery of these resources (see [27]). These IBOs can sometimes be opaque in that the contents of the objects are not used by automated library systems—rather, only metadata about the IBOs is used—especially when complex objects such as scientific data are involved. Our proposed work will substantially extend the use of metadata for formally describing the content of scientific data objects, thereby enabling intelligent data processing in subsequent components of the prototype knowledge network. Specifically, we will develop a distributed data access system based on the Storage Resource Broker (SRB) and a unified metadata catalog for resource discovery.

Solution—Access to data in heterogeneous storage infrastructures at each of the distributed sites will be mediated through a standardized application programming interface (API) that provides uniform access to filesystem, database, and offline storage resources. This interface will be provided through the Storage Resource Broker [1-3][48][53]. SRB technology is a critical research component of SDSC's Massive Data Archiving System and is being used in the nascent California Digital Library (see [42]) and for providing access to NPACI data caches [38]. The SRB will provide high level security, access, audit, and distribution features (e.g., mirroring) that are not easily achievable with other, simpler approaches (e.g., http). The SRB extensibly handles multiple types of storage systems including different file systems, mass storage systems such as UNITREE and HPSS, and binary large objects managed by databases such as Oracle and Illustra. To improve the open interoperability of the SRB, we will develop a Common Object Request Broker Architecture (CORBA) interface to the SRB API, thereby allowing the SRB to be integrated into distributed component systems and registered with a centralized Object Request Broker (ORB). By utilizing the SRB technology for integrating distributed, heterogeneous storage facilities, we will be able to quickly evaluate the effectiveness of a highly distributed storage architecture for knowledge networking within high-speed networks such as Internet 2. Moreover, by providing access to data stores at distributed research sites such as field stations, we will be able to determine how effectively researchers can access relevant data both within and outside their focused domains of interest without high bandwidth connections.

Metadata provide the information necessary to discover and retrieve data of interest from the distributed environment. Metadata describing environmental data currently are developed at individual research sites using multiple, potentially overlapping, and inconsistent metadata conventions. Consequently, there is a simultaneous need for both content standardization and metadata interoperability, as well as metadata transformation facilities. Several research communities have recently advocated the use of structured markup languages such as eXtensible Markup Language (XML) [7] and Resource Description Framework (RDF) [29] as new and innovative mechanisms for metadata interoperability. XML provides a mechanism for syntactic interoperability that allows for standardized processing tools to be utilized on a variety of metadata types. RDF provides a generic way of representing metadata content so that different content standards can be independently developed but then commingled within a single conceptual framework.

We will achieve metadata interoperability in the proposed knowledge network by extending current research at NCEAS on Ecological Metadata Language (EML), a prototype XML application specifying the content and structure of ecological metadata based on the recommendations of the Ecological Society of America [16][36][41]. In particular, we will use RDF to make EML modular and interoperable with metadata content specifications currently under development by other disciplines. Relevant metadata content standards will come from other biocomplexity data domains (e.g., LTER Catalog of Core Datasets [35]; NBII metadata [13]) as well as non-ecological domains (e.g., geospatial metadata [12]; Dublin Core metadata [61]). RDF provides a framework for representation of metadata that reduces the redundancy in metadata content standards currently in use (e.g., most scientific metadata standards [e.g., CSDGM] duplicate information found in other standards [e.g., Dublin Core]) and facilitates the development of automated conversion tools for metadata that map from one content standard to another. Consequently, this work will simplify and automate the use, exchange, and conversion of metadata while providing for a uniform querying capability across metadata content types. By using RDF to provide this functionality we can evaluate the extent to which these cutting-edge technologies are effective for the knowledge networking of complex scientific data.

Deliverables—For our prototype implementation, we will use the SRB to link a distributed set of data repositories from field stations at the LTER network, the OBFS network, and the California Natural Reserve System, for a total of 27 data sites. In addition, we will create a centralized metadata catalog that harvests metadata from each of these registered repository sites. The exchange format will be RDF serializations of Ecological Metadata Language (EML) and other metadata (e.g., Dublin Core) that can be incorporated into a unified metadata catalog. The structure of the catalog will accommodate these multiple metadata standards and will be based on the RDF metadata model but will be implemented in a relational database (e.g., Oracle).

In addition to the metadata and data repository services described here, we will provide a user-oriented, cross-platform interface for interaction with the repository and metadata services, including: 1) metadata development, verification, query, and update, and 2) data object storage and retrieval. This metadata management client, extended from a Java metadata editor developed at NCEAS [41], will provide the capability to edit and update metadata and query the centralized metadata catalog via a standard CORBA interface. The data storage and retrieval client will communicate with the SRB servers via SRB client libraries, again mediated through a standardized CORBA interface. These combined functions will provide transparent access to the distributed network of ecological data repositories.

Evaluation— This research will provide a test of rapidly advancing technological approaches to metadata specification and exchange in a scientific domain, while obviously advancing the kinds and amount of environmental science that can be accomplished through enhanced access to large stores of well-described data. We will validate the system's effectiveness (recall and precision of query) by utilizing the knowledge network to discover and access biodiversity and net primary productivity data that are known to exist across six LTER sites based on the results of a previous collaboration between NCEAS and LTER (see "Validation" in section IIIB) [59][62]. Once validation is performed, the query will be expanded to gather a comprehensive set of biodiversity and productivity data across the full, geographically dispersed network of 27 data repository sites.

A.2. Information: Resolving structural and semantic heterogeneity via schema integration

Challenge—The scientific utility of a knowledge network is ultimately measured by the extent to which it supports the generation of novel and synthetic insights, which requires transformation of **structurally and semantically heterogeneous** data into integrated and meaningful information. The facilities described in Section IIIA.1 provide the researcher with simplified access to greater amounts of ecologically relevant data than previously have been available without extensive and potentially prohibitive amounts of human-assisted labor. Yet these data will not be comprehensible to the end-user unless relevant attributes describing the scientific content of the data are recorded within the accompanying metadata and are available for machine-processing within the Knowledge Network. Predefining a complete set of relevant metadata elements for all scientific uses is unlikely because we expect that the needs and interests of the end-users of our prototype will range widely and in unanticipated ways. Our system will thus need a mechanism for expressing the semantics of data content that is extensible and amenable to machine-based understanding [28]. In particular, we will develop the capability to express higher-order, conceptual statements within the metadata to unify or differentiate scientific concepts and associate these with specific components of data sets.

A second critical challenge arises from the need to integrate data from heterogeneous sources before scientific analyses can be undertaken. Queries on this knowledge network will potentially return a large number of heterogeneous data sets containing relevant information. The integration of these data will be difficult because many if not all data sets will be unfamiliar to any particular end-user. Moreover, these data will derive from multidisciplinary sources which structure and label data differently due to historical and methodological reasons. We describe below how we address these needs by developing extensible facilities for the **structural and conceptual integration** of biocomplexity data. We refer to this effort as **schema integration**, a major research issue within the database community [55-56]. However, our approach will derive more from recent advances within the Web and digital library communities, where structured metadata are key to querying heterogeneous and autonomous collections of digital objects [34][39][44].

Solution—We will extend our structured approach to metadata to incorporate the expression of semantic relationships within and among the data sets comprising our knowledge network. In particular, we will research the use of newly developed schema specification languages, such as the RDF Schema Specification [8] and the XML Schema Specification [5-6], for developing metadata constructs that can describe relationships existing within the data of interest in an extensible way. The use of RDF Schema or XML Schema provides us with all the advantages of XML and RDF as described above, i.e., a metadata specification that is interoperable, extensible, and machine-readable.

Both **structural** and **conceptual** integration of heterogeneous data sets will be primary areas of concern. Development of **structural descriptors** will enable automated transformation of referent data sets into formats suitable for integration with other similar data sets. For example, transformations might include simple conversions among units and measurement calibrations. These structural descriptors will include descriptions of data set format, data types (e.g., alphanumeric, floating point, integer, date), data units, and data constraints, such as permitted value ranges and cardinality. We will develop a schema for structural metadata that provides sufficient information for machine-automated parsing and validating of the referent data using only their structural metadata. Since most ecological data are tabular attribute sets, we will initially focus on ASCII structured tabular data files. For integration of more formally structured data formats, such as data in a relational database or NetCDF format, metadata conversion mediators will need to be developed.

Another critical precursor for data integration is the ability to develop **conceptual descriptors** within the metadata, to enable the expression of scientifically relevant concepts and their association with specific elements of a data set. We will focus on developing schemata that clarify conceptual relationships and dependencies among attributes within a data set, as well as allow for specifying higher order concepts that do not necessarily explicitly appear in the referent data sets. This research will require close collaboration with experts in the ecological science domain, who will inform the technology developers as to necessary or desirable semantic features to be developed, and the specific definitions of their properties. This work constitutes a major effort to develop a discipline-specific vocabulary that is highly relevant to querying and conceptualization of ecological data. For example, ecologists might generally have an interest in biodiversity or predation, but these concepts usually would not exist as a named attribute within a data set even if data relevant to the concepts were present. Rather, the relationship between a concept of interest (e.g., biodiversity) and a relevant measurement (e.g., species density) in a data set is generally not explicitly expressed in the data or metadata.

Given the ability to express relevant scientific concepts, we will also need a mechanism for clarifying how these higher order concepts themselves might interrelate. We propose to explore the suitability of metadata as a means for knowledge representation. Indeed, the RDF Schema Specification has been proposed for this purpose within the Conceptual Knowledge Markup Language (CKML [26]), and its developers borrowed extensively from both database models and the field of knowledge representation in developing this methodology [8]. RDF Schema provides for inheritance via Class/Subclass propagation, which should enable the development of "concept ontologies" for rich expression of hierarchical relationships [28]. Ecology, like many scientific disciplines, employs a number of concepts that have great heuristic value, but multiple operational definitions when actualized within a quantitative analysis. Our mechanism for expressing semantics within metadata will enable development, refinement, and preservation of standardized definitions for key ecological concepts. For example, biodiversity can be quantified in a number of ways. Two examples include:

richness $\quad b_1 = H' = -\sum p_i \ln p_i$ where p is the proportionate abundance of species i,

evenness $\quad b_2 = \left(\frac{H'}{\ln S} \right)$ where S is the total number of species ,

Both of these quantitative specifications of biodiversity are derived from a set of canonical variables that can be expressed using a controlled vocabulary. The canonical variables represent fundamental measurements within the data sets, including concepts such as ‘individual’ or mass in *System Internationale* (SI) units. When presented within the context of the hypothesis modeling system proposed below (see IIIA.3), this framework will enable the comparative evaluation of multiple formulations of hypotheses using multiple and ‘competing’ definitions of ecological concepts that are built from these canonical variables.

Deliverables— First, to formalize key structural and conceptual relationships typically relevant to ecological data sets, we will develop and deploy metadata schemata, likely using the RDF Schema specification, that allow for the expression of higher order concepts relevant to the description of ecological data and their potential use. This will involve research into the ease and efficacy of using schema specification languages for developing metadata that provide for the integration of scientific data sets. This work represents our primary mechanism for achieving data integration, and will entail significant enhancements to the functionality of the metadata and data management client described in the preceding section (Figure 1; IIIA.1). Second, our development of canonical descriptors for ecology, and formalization of a number of key ecological concepts using these canonicals, constitutes another major product. RDF Schema provides a rich framework for formalizing these concepts, and a mechanism for expressing their alternative representations in an explicit and interoperable way. Finally, we will develop a prototype data integration engine for the semi-automated integration of heterogeneous ecological data sets. The data integration engine will utilize these structural and conceptual metadata descriptors in order to identify, retrieve, and combine raw data from the distributed sites.

Evaluation— Evaluation of the metadata schema specifications that we develop will be conducted in concert with the biodiversity research project described below (IIIB). In particular, we will engage scientists to evaluate the effectiveness of the metadata schemata for expressing structural and conceptual aspects of ecological data. Scientists also will be asked to evaluate the efficacy of the data integration engine prototype in terms of its ability to transform data into formats suitable for further synthetic analysis. Finally, the expression of canonical variables will be tested through their use in the hypothesis modeling section described next (IIIA.3).

<u>A.3. Knowledge: Automated modeling of scientific hypotheses using data resources</u>

Challenge—As described above, many ecological concepts can legitimately be formulated in a number of ways, based on different functional combinations of canonical variables. Consequently, the problem remains as to how researchers can systematically and formally evaluate the multiple formulations of hypotheses that result from utilizing these alternative functional forms of concepts and all available data. Because hypothesis development and testing are fundamental features of science across all disciplines, we propose an approach facilitating the development and evaluation of candidate hypotheses that both exploits the available data resources and enables a more effective comparison of alternative formulations of hypotheses.

Achievements of the past decade in machine learning provide new approaches to probabilistic modeling of learning and induction that can be applied to hypothesis modeling. Among these are belief networks [46], adaptive probabilistic networks (APN) [50-51] and work on learning in Bayesian networks [19][21-23]. Recent work utilizing belief networks has demonstrated the ability to estimate parameters in deterministic models [57]. We will exploit these advances to compare the performance of a set of hypothesis-variants generated from a set of ecological concepts expressed in terms of canonical variables for a given set of data.

Solution—We propose to apply methods from machine learning to evaluate the relative merit of *a priori* hypotheses as a means of guiding the exploration and analysis of available data. We believe this is a novel approach to the use of machine learning. As data resources grow, it becomes increasingly difficult to evaluate the most promising direction to focus research efforts. Our research will investigate the extent to which the limitations in existing machine learning approaches can be overcome to take greater advantage of both the insight of the researcher(s) and available data. Although we focus on the ecological domain in this study, the approaches represent foundational research that is applicable to any area of science where the formulation of research questions is an important activity.

In general, a hypothesis is defined using scientific terms that can be multiply defined. Thus, any particular hypothesis can be decomposed into many alternative formulations, each of which uses a different instance of the possible definitions of the concepts expressed in the original hypothesis. For example, the research hypothesis "Biodiversity varies as a function of Net Primary Productivity (NPP)" can be interpreted in multiple ways, depending on the definition of biodiversity, the nature of the functional relationship, and the definition of NPP. To the extent that there are multiple definitions for each of these, there are alternative possible formulations of the hypothesis. Our approach will provide a method to decompose a hypothesis into its components and then evaluate the relative merit of each of the resulting alternatives. To continue our example, we can state the working null hypothesis as "H_0: biodiversity does not vary with net primary production (NPP)". We are given an initial set of data (δ_1), a definition of biodiversity (in this case species density (β_1) as number of species per unit area), and a definition of NPP (η_1) as the above-ground dry biomass. The variation of biodiversity with NPP admits a variety of possible functional relationships, (e.g., linear, quadratic) and may extend to more advanced techniques such as semi-variance analysis. We can examine the nature of the functional relationship (i.e., "varies") between β_1 and η_1 using conventional model-fitting and statistical hypothesis-testing methods such as regression. However, to choose between alternative formulations of the hypothesis prior to statistical hypothesis-testing we wish to consider the availability of data collected at various scales of measurement (δ_k) and also the implications of alternative univariate ($f(\eta_j)$) or multivariate ($f(\eta_{1\ldots}\eta_{\lambda,})$) definitions of biodiversity or NPP. This also has to be done with appropriate consideration to the independence requirements of multiple-hypothesis tests. So we add to the initial set of hypothesis modeling components the following: 1) additional data sampled at differing known scales (δ_2, δ_3, …); and, 2) additional definitions of biodiversity and net primary productivity.

Effectively we are extending (in parameter space) and amplifying (in data space) the initial hypothesis across its neighborhood in the general form:

$$\forall(i,j,k \mid H_0: \beta_i = f(\eta_j) \text{ given } \delta_k)$$

For the univariate estimators (β_{I}, η_{j}) alone the number of possible alternatives is the product of the number of different biodiversity estimators (i), and NPP estimators (j), with the combinatorial expansion of the collection of datasets (i.e., N choose p, where N = total number of datasets, and p= 1…N datasets taken p at a time) accessible through the knowledge network. These alternatives represent a top-down specification of a potentially large number of hypotheses to be evaluated against a given knowledge base. This is significantly different than the usual application and interpretation of knowledge-networks organized to solve (i.e., discover hypotheses) from the bottom-up. Consequently, we propose to interpret the hypothesis-space not as the solution-space of a belief network but rather as a set of propositions to be evaluated against a belief network. In this context we treat the belief network as the current best description of a specific research domain (e.g., biodiversity). By computing the gradient of the likelihood of the alternatives with respect to the hypothesis model and the data, we will obtain a measure of the relative merit of the alternative formulations of the hypothesis.

Deliverables—Specific functional components of our system will include: 1) a hypothesis specification language to provide a means of unambiguously defining a model (i.e., hypothesis) in terms of its canonical variables, 2) a knowledge representation using Bayesian network and graphical model approaches, and 3) methods of estimating conditional probabilities and their gradients.

Hypothesis Specification Language—The hypothesis specification language will utilize ecological concepts and canonical variables as described earlier in Section IIIA.2. We do not propose to develop a complete, new language compiler. Rather, we require the ability to translate simple declarative statements into canonical variables, which can be satisfied by something as simple as the yacc compiler available on most Unix systems. This language will be used to define and control the relationship of the canonical variables (e.g., *H', p,)* to the higher-level concepts used in the hypothesis (e.g., biodiversity, net primary productivity).

Knowledge Representation—There are numerous methods for the representation of knowledge [24][30][40], and this remains an active area of research. The method we propose to use will be a variant of the belief network representation of Pearl [46] and current graphical model approaches [19][25]. Our final selection will depend on the trade-off between convenience of presentation and computational efficiency, but we will initially utilize a matrix or list-based representation until we choose a particular implementation approach. Considerable effort will be spent in year 1 exploring the most effective means of representing prior knowledge.

One of the main difficulties in Bayesian systems is the incorporation of prior information from incomplete and multivariate data. The current approach is to make simplifying assumptions about the underlying distributions (e.g., Wishart, Gaussian) of the parameters in continuous models such as regression functions to obtain an estimate of the prior probability of a predictive model such as $X \rightarrow Y$ where X and Y may both be multivariate [22]. Experience from statistics tells us that these assumptions are usually unrealistic and/or unverifiable. This research proposal will provide an opportunity to examine the sensitivity of Bayesian networks to these assumptions and to their relaxation.

<u>Estimation of Conditional Probabilities and Gradient</u>—The ability to compute the gradient of the likelihood of the alternatives, as referred to above, implies that we can discover relevant data resources and extract from them attributes (i.e., data parameters) in a form suitable for integration with existing data. Following [20] assume that the general form for the posterior estimate of the likelihood of a particular hypothesis, m_i, given a set of empirical data, d_j, can be formulated as:

$$p(m_i | d_j) = \frac{p(m_i)\, p(d_j | m_i)}{\sum_j p(m_i)\, p(d_j | m_i)}$$

Where $\mathbf{M} = (m_1,\ldots,m_k)$ is the set of hypothesis variants and $\mathbf{D} = (d_1,\ldots,d_l)$ is the set of all data. The terms in the numerator correspond to a sub-tree of the full set of alternative combinations of hypotheses (i.e., models) and data. If, instead of computing the single value $p(m|D)$ we compute the difference between the combinations of hypothesis variants (i.e., subtrees) and data relative to a given *a priori* hypothesis (i.e., $p(m^* | D^*)$) as:

$$p(m_i | D) = p(m^* | d^*) + \sum_j \frac{\partial}{\partial dj} p(m^* | d^*) \Delta dj + \sum_k \frac{\partial}{\partial m_k} p(m^* | d^*) \Delta m_k + e$$

$$\Delta p(m | D) = p(m | D) - p(m^* | D^*)$$

we have an estimate of the relative strength of a hypothesis within a neighborhood of variants where the differences are computed relative to the nominated *a priori* hypothesis, $p(m^* | D^*)$. If the neighborhood is sufficiently small, in a manner analogous to the use of multivariate Taylor series for perturbation analysis in the continuously differentiable case, we can develop estimates of $\frac{\Delta p(m|d)}{\Delta m}$, $\frac{\Delta p(m|d)}{\Delta d}$ in the hypothesis model even if the components are suspicious with respect to the requirement of differentiability. This method should be extensible to other potential components of an extended model of hypotheses. We then will have a summary description of the relative strengths of the competing hypotheses. This provides a sensitivity analysis of the proposed *a priori* hypothesis as a function of alternative formulations and the available data. The result can be displayed as a 2-dimensional matrix of $\frac{\Delta p(m|d)}{\Delta m}$ x $\frac{\Delta p(m|d)}{\Delta d}$ representing the response surface of the hypothesis space. Maxima will represent combinations of model and data that have the greatest positive effect on the predictive value of the hypothesis while minima represent the converse. This surface will provide insight into the relevance, relative strength and weakness of alternative research approaches. This information can be used to guide the researcher to the most promising research effort given the best information at hand.

The calculation of likelihood terms is an active area of research [22]. We expect to spend considerable effort in exploring the most effective means of computing the conditional and unconditional terms. Empirical data and estimates of parameter distributions will be used to compute the entries in the conditional probabilities from which the likelihood gradient is computed. These probabilities can be estimated from the p-values derived from the estimation of the parameters of a postulated functional relationship (e.g., linear, quadratic, etc.).

Evaluation— Assessing the effectiveness of this semi-automated hypothesis modeling facility will require comparison of the results against a set of hypotheses with known characteristics. Given an initial hypothesis, we will need to verify that: 1) the decomposition of the hypothesis and representation of the multiple, alternative formulations in our hypothesis specification language are correct; 2) the data parameters that were used to calculate the likelihood for each of the hypothesis formulations are appropriate and properly integrated; and, 3) the gradient field analysis produced by the system will yield the same results that would be reached by a scientist manually analyzing the information at hand. We propose to do this evaluation by comparing results from this modeling system with those from the biodiversity and productivity analysis completed by Waide et al. [59] [see following section].

B. Using and Evaluating the Knowledge Network: Biocomplexity Research

Thus far we have provided the rationale for a Knowledge Network for Biocomplexity, and described our approach to its technical design and implementation. It is widely recognized that successful extrapolation of scientific insights on environmental issues to broader spatiotemporal scales critically depends on access to greater data resources [45]. This need is the primary catalyst for our efforts to construct this Knowledge Network for biocomplexity information. We now describe how ecological researchers, in conjunction with expertise from other disciplines, will use and evaluate the Knowledge Network by addressing a research question of vital theoretical and practical importance—the relationship between biodiversity and organic productivity. This question has stymied researchers, and is largely intractable due to the lack of sufficient data resources to develop powerful statistical tests. The environmental research described in this section will not only test and inform the design and development of the Knowledge Network but also stands on its own as a pathbreaking attempt to synthesize all available information regarding the relationship between these two key aspects defining the richness and sustainability of the natural world.

B.1. Rationale for Biodiversity and Ecosystem Case Study

Understanding the relationship between biodiversity and ecosystem function is a challenge to contemporary science [9][54] because of the complexity of the patterns and mechanistic interconnections that underlie them [60]. Indeed, ecosystems arise from the interactions among living organisms, as well as from the interactions between the biota and non-living sources of energy and nutrients [31][43], making ecological understanding a multidisciplinary endeavor anchored in the biological, earth, and atmospheric sciences. The structure of the biotic community is defined by the identity and density of its constituent species (biodiversity), whereas the functional attributes pertaining to that organization include processes such as productivity, decomposition, and nitrogen mineralization. We have purposely chosen to address this complex and scale-dependent question, not only because of its intrinsic scientific merit, but because it typifies the kinds of overarching theoretical constructs and practical constraints that distinguish research in the ecological sciences from the reductionist approaches of chemistry, physics, or molecular biology [47]. In addition, this multidisciplinary research question typifies the type of ecological research that requires broad integration of heterogeneous data from autonomous and distributed sources. Finally, this question is critical if applied sciences such as conservation, restoration, and management are to utilize science-based methods to minimize the negative consequences of environmental degradation to contemporary society [45].

B.2. Multi-scale research on biodiversity and ecosystem function

Using the knowledge network developed as part of this proposal, we will execute research on the scale-dependent relationship between biodiversity and ecosystem function in a series of stages. In addition to addressing ecological questions of inherent importance, the research will provide feedback, grounded in the realities of complex ecological data, to the development of technologies in support of a Knowledge Network for biocomplexity. Our hierarchical approach to research involves four interrelated phases: (1) **Validation**, (2) **Expansion**, (3) **Amplification**, and (4) **Extrapolation**. We anticipate that the knowledge network will provide rich sources of data for analysis of ecologically relevant questions concerning the relationship between biodiversity and ecosystem function, but at the same time, execution of the research will provide a vehicle for the improvement of technologies and serve as the "proof of concept" for the utility of the knowledge network.

Validation—The ecological research proposed here emerges from collective experience gained during a cooperative study between the LTER Network and NCEAS. The results of that collaboration [15-16][32][37][52][58-59][62] provide an ideal starting point for evaluating and improving the technology that drives the integrated knowledge network for ecology. Four key issues for knowledge networking arise from that research. First, difficulties in discovering, retrieving, and standardizing data severely limited the scope of the previous investigation to include only net primary productivity (NPP) of grasslands and vascular plants from LTER data sets. Second, the collection and integration of the data on both NPP and species richness was extremely labor intensive, despite the full assistance of LTER data managers. Third, the relationship between biodiversity and productivity was scale-dependent. Fourth, better access to more data would have improved the results of the study through enhanced statistical power.

We will test the utility of the Knowledge Network by assessing the degree to which it identifies, retrieves, standardizes and integrates data from the same LTER sites (Arctic, Cedar Creek, Kellogg, Konza, Short-grass Steppe, and Toolik Lake) that led to the biological conclusions derived from our previous, more conventional approach to the problem. Moreover, we anticipate that the Knowledge Network will provide additional useable data from those same sites, which we were unable to identify and incorporate in our initial assessment. This will increase the power of statistical analyses, enhancing confidence that non-significant results are the result of the absence of a relationship, and provide additional observations to enable the detection of associations that appear weak because of inherent environmental heterogeneity.

Expansion—Once the knowledge network has the capability to identify and retrieve data comparable to that in our previous work as determined in the validation phase, we can expand the domain of analyses to include other LTER sites, especially those representative of other biomes. In addition, we can obtain data from sites outside the LTER network, such as those in the Organization of Biological Field Stations (OBFS) and the California Nature Reserve System. Such expansive data will allow us to take two important steps forward in understanding the relationship between diversity and productivity. First, we will be able to see if the relationship is scale-dependent in similar fashions *within* different biomes (e.g., grasslands, deserts, and forests). Second, using the combined data from all sites, we will be able to define the relationship between productivity and diversity as the dominant life forms change, e.g. from annual grasses to perennial grasses and herbs, to shrubs, and to hardwood trees. This is important because factors controlling productivity or diversity *within* a biome may be different than those *among* biomes. Patterns in the relationship may change because the underlying mechanisms affecting one or both variables change. This expansion phase of the project will be critical in testing our ability to discover, interpret, and integrate heterogeneous data from a variety of distributed data sources (See Sections IIIA.1 and IIIA.2).

Amplification—Once an effective, efficient, and validated knowledge network is in place, we can amplify perspectives from which we assess the relationship between biodiversity and ecosystem function. Species richness is not the only index of biodiversity, and productivity is not the only measure of ecosystem function. Moreover, the relationship between different aspects of biodiversity and each of a number of ecosystem functions need not be the same. We will obtain data from as many sites as possible, in order to estimate species richness, species evenness, and species dominance [33]. Similarly, from as many sites as possible, we will obtain measures of ecosystem function (e.g., decomposition rates, rates of nitrogen mineralization, turnover times for phosphorus). These data will allow us to answer two important questions. First, do different measures of biodiversity have similar associations (i.e., the form, parameters, and strength of the relationship) with a single measure of ecosystem function? Second, does the same measure of biodiversity respond to different measures of ecosystem function in the same way? This phase of the project will directly test the hypothesis modeling capabilities that were described in Section IIIA.3 because it deals with the expansion of the hypothesis into multiple alternative formulations that can be directly tied to data resources from the Knowledge Network.

Extrapolation—To study the relationship at the broadest of spatial scales (i.e., continental or global), data on biodiversity and ecosystem function must represent salient environmental gradients occurring over broad ranges of latitude, longitude, and elevation. Data derived from the current confederation of sites that contribute to the knowledge network may not be sufficiently ample to generate site-specific data on *both* measures of ecosystem function and biodiversity. Nonetheless, a variety of sites may provide data on one measure of biodiversity (e.g., species richness) while aspects of ecological function (e.g., productivity) may be estimated from broad-scale climatic data (e.g., temperature and precipitation) in conjunction with information about soil characteristics. Hence, we will use the entire arsenal of knowledge networking tools to access a broad array of information from selected governmental and public databases, as well as from climatic and earth science computer models (WWW sites, see [59]), to estimate critical missing data. We can validate this approach for estimating parameters by comparing the estimated values from modeling algorithms to the values for sites at which empirical measurements have been made.

B.3. Working Groups as Integrative Activities

The NCEAS model for working group activities has proven to be highly effective at catalyzing synthetic research in ecology [49]. We will use this model of collaborative research to enhance multidisciplinary interactions and to diversify the expertise of scientists addressing questions concerning the relationship between ecosystem function and biodiversity. The **Biocomplexity Working Group** (P.I. Willig, Coordinator) will comprise representatives of the LTER and OBFS sites who contribute data to the knowledge network, as well as other ecologists, earth and atmospheric scientists, informatics specialists, and computer scientists. Because of the labor-intensive nature of data management for a large multi-site project, as well as the need to provide scientific linkages between biocomplexity research and the modeling efforts required for the automated hypothesis testing, a post-doctoral fellow will assist the coordinator in executing these tasks and serve as liaison with the **Informatics Working Group**. Similarly, the expected site-specificity of scale-dependence in the relationship between biodiversity and ecosystem function will require considerable modeling efforts to develop appropriate transformations for biocomplexity information to be useable for cross-scale, cross-site analyses. These findings will then be incorporated into the knowledge representation format of the Knowledge Network. A second post-doctoral fellow will assist the coordinator in executing this task and serve as liaison with the **Graduate Student Working Group**, which will be confronting similar challenges in their seminar activities.

C. Education, Outreach and Training

To train *young* scientists in the use of biocomplexity information and its associated technology, we will initiate a series of integrated graduate training seminars at 5-10 universities during each year of the award. The goals of the seminars are threefold: 1) to train graduate students in organizing and synthesizing multi-scale data on biodiversity and productivity; 2) to provide an interactive environment in which students have the opportunity to use and provide input to the design of the software tools constituting the knowledge network; and 3) to demonstrate the power of a multi-scale, multidisciplinary approach to biocomplexity research.

In this research training model, successfully employed at NCEAS for other studies, faculty at participating universities will conduct graduate seminars on the relationship between ecosystem function and biodiversity. The seminars will be highly interdisciplinary, including students in ecology, natural resource management, informatics, and statistics. The students will gather and integrate diverse information, using the tools developed in this project, and thereby test the efficacy of the tools. Each of the seminars will synthesize and analyze data from their local LTER or OBFS site on biodiversity and ecosystem function. After the local seminars are completed graduate students from each seminar will participate in an NCEAS working group that will compare results across the sites under consideration. In order to standardize and coordinate the collection and analysis of data we will develop a digital, web-based syllabus, web-based training materials and an electronic collaboratory. The net results of this distinctive training effort will be the generation of a cohort of young investigators well versed in ecological knowledge networking tools and approaches.

To foster the adoption of the knowledge networking tools and approaches developed during this research project, we will collaborate with the Ecological Society of America (ESA) to conduct symposia at annual ESA meetings and develop a web-based training manual for the use of metadata and other knowledge networking tools. In addition, the LTER Network Office will develop a training module from the results of this project for individual LTER and OBFS sites. This module will demonstrate to data managers how to provide reciprocal access to data at multiple sites and field sites. Individuals receiving the training will be expected to make their data available through the knowledge network. When the prototype knowledge network is fully developed, we will extend training efforts to other sites through workshops at annual meetings of the ESA and the LTER All-Scientists meeting.

IV. PROJECT SIGNIFICANCE

Information is most valuable when it is easily accessible for the purpose of knowledge creation and decision making. In 1998, the President's Committee of Advisors on Science and Technology concluded that "We need [...] research to produce mechanisms that can […] efficiently search through [...] biodiversity and ecosystems data sets, make correlations among data from different sources, compile those data in new ways, analyze and synthesize them, and present the resulting information in an understandable and usable manner." [45]. The science of ecology urgently requires a concerted effort to develop an informatics and collaboratory framework to provide access to information from the many disciplines contributing to our understanding of biocomplexity in the context of global ecosystems. Moreover, the advances in informatics and computer science research that are accomplished in addressing this challenge will have direct application to related disciplines in earth and atmospheric sciences, as well as to other disciplines examining the phenomena related to complexity in general. Our research in using structured metadata for the development of higher-order statements and their application to hypothesis modeling has broad applicability to any domain dealing with complex data.

Intellectual Merit

The cornerstone of this proposal involves development of a comprehensive prototype knowledge network for biocomplexity informatics through a unique partnership of ecologists, computer scientists, software engineers, and informatics researchers. This partnership incorporates national leaders in ecological informatics (LTER), synthesis of ecological data (NCEAS), and development of computational infrastructure (SDSC), who collectively have access to extensive ecological data sets, closely-connected and extensive networks of scientists specializing in ecology and informatics, and state-of-the-art computational facilities. We propose to integrate the diffuse and heterogeneous information sources required for the development and testing of theory in ecology and its sister fields into a standards-based, open architecture, knowledge network. The existence of such a network will lead to broadened understanding of biocomplexity and ecological systems, and allow the application of that understanding to societal issues. In developing this network, we will create a new community of environmental scientists who will be able to focus on complex, multi-scale problems that, to date, have proven to be intractable. We will perform foundational research in computer science and informatics to create new tools for discovering, retrieving, interpreting, integrating, and analyzing data from these diverse sources. Our prototype network will be useful across a variety of disciplines and will provide a basis for the growth of multidisciplinary research groups focused on biocomplexity.

Broader Impacts

The results of the proposed research will have broad implications for our ability to understand and manage sustainably the complex ecological systems and biological resources on which all humans depend. Ecological and biodiversity information is voluminous and complex, but currently is inaccessible to research scientists and policy makers. The intellectual advances in information science that we propose will, for the first time, provide an accessible infrastructure for identifying, integrating, managing, and, ultimately, synthesizing the nation's ecological and biodiversity information resources. Key components of the proposal address education (graduate seminars involving over 150 students per year), outreach (to OBFS facilities via the LTER Network Office), and training (at NCEAS, the annual meeting of ESA, and the LTER All Scientists Meeting). The University of New Mexico is a minority-serving institution, and NCEAS and SDSC actively pursue increased involvement of underrepresented groups in their programs. A plan for dissemination of the products created in this proposal (see below) will ensure broad distribution of the results throughout the suite of scientific disciplines that will be the beneficiaries of the proposed research. Ultimately, the existence of a knowledge network for ecological information will serve directly to provide information to scientists, students, policy planners, resource managers, and the public, and will allow them to make science-based decisions regarding the future of our society.

V. APPROPRIATENESS FOR KDI & ROLES OF PROJECT PERSONNEL

This project directly addresses several of the research emphases of the Knowledge Networking focus of KDI. Overall, the proposed research is a foundational effort to build a system that enables researchers to locate , manage, and utilize heterogeneous, distributed, and complex data, and transform these into knowledge that can be shared within and among disciplines. The project involves specific research goals and products in two extremely broad fields, informatics and ecology, but the methods employed will be applicable to any disciplines needing access to large amounts of data in formats suitable for scientific analysis. The domain-specific research that we will use to test and validate the prototype knowledge network involves a multi-scale, multidisciplinary investigation of the vitally important and complex relationship between ecosystem processes and biodiversity.

The informatics research explores the utility of recent advances in metadata languages for expressing key structural (syntax) and conceptual (semantics) aspects of scientific data, in order to facilitate data discovery, access, and integration. We use open standards in this work, thereby creating the potential for constructing complex and interoperable ontologies for describing domain-specific data. We will initially develop a vocabulary for ecological data that will be used by our proposed hypothesis modeling system. This application, based on adaptive probability networks, will provide a framework for the automated testing of formalized representations of hypothesized relationships, using data within the prototype knowledge network. This work will require foundational research in terms of modeling domain-specific knowledge and processes.

Using these approaches, we will develop a network-accessible, prototype knowledge network that is intrinsically multidisciplinary, with data contributions from, e.g., ecologists, meteorologists, and earth system scientists. A team of scientists will then use the prototype to accomplish a comprehensive analysis of the relationship between ecosystem processes and biodiversity. Robust investigation of this issue is currently impossible due to severe limitations in access to relevant data, and subsequent difficulties in integrating these into a common framework.

With regard to other goals of KDI, this project will enhance education by involving graduate students and knowledge networking specialists in the development of informatics tools; promote a culture shift within ecology and other disciplines toward comprehensive data access; enhance the ability of scientists to transform information into useful knowledge for use by society at large, and resource managers in particular; and advance scientists' abilities to discover and retrieve vast stores of disparate data through knowledge networking tools.

Roles of Senior Personnel

- Reichman: PI, *NCEAS Director*; responsible for planning, coordination and implementation
- Jones: co-PI, *Information Specialist*; lead on metadata and interoperability layer
- Brunt: co-PI, *Information Specialist*; lead for distributed data network
- Helly: co-PI, *Computer Scientist*; responsible for adaptive probability network research
- Willig: co-PI, *Ecologist*; will lead ecological research and the ecology working group
- Andelman: *Ecologist; NCEAS Deputy*; work with education, outreach, and training
- Arzberger: *Executive Director*; responsible for coordination of activities at SDSC.
- Rajasekar: *Computer Scientist*; SRB system support at SDSC
- Schildhauer, *Director of Computing*; coordination of technology research between partners
- Sutton, *Developer*; software engineering on hypothesis modeling engine
- Waide: *Executive Director*, *LTER Network Office*; coordination of LTER activities

VI. RESULTS FROM PRIOR NSF SUPPORT

O. J. Reichman and M. Jones, NCEAS, UCSB, Results of Prior Support

National Center for Ecological Analysis and Synthesis (DEB-94-21535)

***May 01, 1995 - April 30, 2000*:** O. J. Reichman is Director of the National Center for Ecological Analysis and Synthesis (NCEAS), and Professor in the Department of Ecology, Evolution, and Marine Biology at UC Santa Barbara. NCEAS provides the opportunity and facilities for intense collaborative research on major fundamental and applied problems in ecological and environmental science. Since May 1995 the Center has hosted over 1400 visiting scientists and sponsored 2 major symposia, 25 workshops, 24 working groups, 22 visiting scientists and sabbatical fellows, and 16 postdoctoral fellows. Details of these activities can be found under the Center's web site at http://www.nceas.ucsb.edu. NCEAS' scientific computing staff maintain a high-performance networked computing environment for these projects, many of which require the integration of large and diverse data sets. NCEAS' mission includes leadership in the development of new methods appropriate to advancing synthetic and integrative work in the ecological sciences. Through experience with our research groups, the scientific computing staff at NCEAS have identified a critical need for better tools and standards for the integration and preservation of disparate ecological data, and have become involved in a number of informatics initiatives to address these issues. NCEAS' Database and Information Specialist and Director of Computing are responsible for planning and coordinating these data and informatics initiatives and strategies.

NCEAS-sponsored Workshop on "Data Management for the Ecological Sciences": In recognition of the key element that data management plays in facilitating integrative and synthetic research within ecology, NCEAS convened a workshop consisting of executives and informatics specialists from a number of institutions that have similar commitments and concerns about data management. The intent was to foster communication among these institutions, to identify specific informatics needs, and to develop a plan for addressing these issues.

NCEAS-sponsored meetings of the Ecological Society of America's "Committee on Data Archiving and Sharing": NCEAS hosted meetings and maintains the Web site of this special Committee that was formed to address issues of preservation and access to ecological data. NCEAS provided technical consultation to the Committee through participation of its informatics staff. The Committee provided oversight for the Web-based data management project described below, and developed specific recommendations to the governing board of the Ecological Society of America that are leading to the foundation of a peer-reviewed publication dedicated to the description and preservation of particularly noteworthy ecological data sets.

Training and Outreach: NCEAS' scientific computing staff participated as invited instructors to a Data Management Training Workshop, jointly sponsored by the Organization for Biological Field Stations (OBFS) and the Long-Term Ecological Research Network (LTER) in 1997. They have also contributed two chapters to a planned publication growing from this event. The technical staff have been invited speakers to a session on 'Advances in Ecological Software', at the 1997 Ecological Society of America's annual meeting. NCEAS' scientific computing staff provide ad hoc training to dozens of visiting scientists and analysts associated with research efforts at NCEAS that require demanding informatics solutions in order to succeed.

Automation of Ecological Data Management using Structured Metadata: Supplement to NCEAS (DEB-94-21535) August 1997-August 1999. This project has two goals: 1) to formalize the suggested ecological metadata content standard from Michener et al.(1997) using extensible markup language, XML; 2) to develop a platform-neutral software tool that enables researchers to create XML documents that conform to the ecological metadata content standard developed under goal (1). See www.nceas.ucsb.edu/ecoinformatics for a summary of progress and software developed to date, including the following milestones.

- Completion of a comprehensive review of proposed metadata standards relevant to ecological sciences (NBII, FGDC, LTER, FLED).
- Design and creation of a Ecological Metadata Language (EML), an XML document type declaration (DTD) for ecological metadata.
- Implementation of the capability to automatically convert XML-structured metadata to HTML using style sheets, which is useful for automatic generation of web pages. Moreover, it illustrates the process of automatic conversion from one metadata standard to another.
- Design and implementation of a metadata editor in Java that enables a scientist to create and modify a metadata document using valid XML-based Document Type Definitions (DTD). The editor provides validation of the resulting XML-based document with the DTD. The design of the editor is flexible in that it supports multiple, diverse, and evolving content standards for metadata that are represented by DTDs. Current focus is on the EML DTDs.
- Creation of a prototype tool for automated quality assurance processing for heterogeneous data that is described using EML metadata. The implementation uses a perl-based XML parser to process XML metadata files and automatically generate analytical processing code for SAS analytical software. Results are formatted in HTML.

Peer reviewed publications from this research include:

Frondorf, A., M. B. Jones, and S. Stitt. 1999. Linking the FGDC geospatial metadata content standard to the biological/ecological sciences. Proceedings of the Third IEEE Computer Society Metadata Conference. Bethesda, MD. April 6-7, 1999.

Nottrott, R., M. B. Jones, and M. Schildhauer. 1999. Using XML-structured metadata to automate quality assurance processing for ecological data. Proceedings of the Third IEEE Computer Society Metadata Conference. Bethesda, MD. April 6-7, 1999.

Invited Presentations from this research include:

Jones, Matthew B., M. P. Schildhauer, and R. Nottrott. 1998. The Role of Informatics in Ecological Synthesis. 1998 Ecological Society of America Annual Meeting. Baltimore, Maryland.

Jones, Matthew B. 1997. Web-Database Integration. NSF Training Workshop for Ecological Data Management. Albuquerque, New Mexico.

Jones, Matthew B. 1997. Data Management at the National Center for Ecological Analysis and Synthesis. NSF Training Workshop for Ecological Data Management. Albuquerque, New Mexico.

Schildhauer, Mark P. and M. B. Jones. 1997. Software Choices for a High Performance Analytical Environment. 1997 Ecological Society of America Annual Meeting. Albuquerque, New Mexico.

Web-based Data Management for Ecological Analysis and Synthesis (DBI 9631091)

September 01, 1996 - August 31, 1999. Frank Davis, past Deputy Director of NCEAS, was the PI for NCEAS on this grant. This grant supports the development of a web-based data management system following the recommendations of the FLED committee report. It addresses mechanisms for encouraging high-quality collections of ecological data, design and operation issues for the web-based data management system, and mechanisms being considered to manage intellectual property rights and provide long-term funding. NCEAS provided technical review of the Web-based front-end for this prototype archive of ecological data.

J. Brunt, LTER Network Office, University of New Mexico, Results of Prior Support

The PI is currently Co-PI and Associate Director for Information Management at the Long-Term Ecological Research (LTER) Network Office (DEB – 9634135). This project supports the communication, data management, data dissemination, and technological advances of the LTER Network of research sites. The PI developed the design plan for the LTER network information system and has the responsibility of coordinating its development with the LTER Network Informatics group and LTER collaborators. The LTER network information system plan was developed out of a need for an information system that seamlessly facilitates and integrates both cross-site and intra-site data exchange, with the mission of meeting the research needs of LTER scientists. The strategy is to design and develop a distributed, LTER-wide information system using a modular approach while maintaining and building on present functionality. The system is currently composed of a variety of prototype modules that are being evaluated. The results of this evaluation will be developed into a framework, or set of specifications, that describe the interoperability of discrete units of the system, into which additional, future modules can be plugged. Other support relevant to this proposal falls into 2 categories: 1) Software Development, and 2) Computational Infrastructure Development.

Software Development

DBI-8905996 - The NMSU Science Workbench Project: An Integrated Approach to Research Data Management. The PI worked on the NMSU Science Workbench Project in graduate school at New Mexico State University producing a number of papers, technical reports, and this proposal which was funded after leaving for the University of New Mexico. The Science Workbench was a computational environment that supported a philosophy of data management and group research. The environment consisted of tools that broadly support data management and manipulation; statistics, graphics, and numerical analysis; and development and analysis of technical text. The Workbench supported both individual research and group demonstrations and workshops. This work was groundbreaking in its use of the internet and client/server technology to support scientific interaction among ecologists.

BIR-9123090 – Collaboration Software for Ecological Scientists: In 1991, the PI received a grant w/ Dr. John Rasure of UNM Computer Engineering for development of collaboration software for ecological scientists. The objective of this multidisciplinary project was to develop software that provided the mechanisms and methods that enabled and encouraged scientists to exchange data, ideas, and results across geographically distributed research sites. The motivation was twofold: the need for better methods of processing and visualizing data in environmental research; and the need for collaboration across and within disciplines. The work extended the Khoros 1.0 software, which was later developed into a highly successful commercial software (K2), provided a working environment designed to motivate scientists from different locations to interact over a computer network.

Computational Infrastructure Development

The PI contributed to and/or served as PI on a number of supplement grants to develop a basic data management capability for the Sevilleta LTER project (BSR-8811906, DEB-9411976), facility grants to develop a similar capability at the Sevilleta Research Field Station (BIR-9311867, BIR - 9512658), at the Mapimi Biosphere Reserve Station in Mexico (DEB-9527083), and the Organization of Tropical Field Stations in Costa Rica (DBI-9601194). These proposals demonstrate a wide-ranging use of technology in solving data management and communication problems from T-1 WAN technology to Satellite data systems.

J. Helly, SDSC, Results of Prior Support

San Diego Supercomputer Center. Cooperative Agreement 10/1/95-9/30/97

This support provided for the establishment of an ecological outreach program at SDSC. As a direct result of this funding:

- San Diego Bay Project was established at SDSC (1),
- First Workshop on Computational Ecology (2) was prepared and held,
- Ecological Society of America's FLED (Future-of-Long Term Ecological Data) was enabled (3), (4),
- numerous workshops and presentations including those at the NSF LTER (Long-Term Ecological Research) Coordinating committee, National Center for Ecological Analysis and Synthesis, White House Conference on Marine Monitoring.

Web-based Data Management for Ecological Analysis and Synthesis 10/1/96-9/30/97 This grant from BIR/DBA is supporting the development of a web-based data management system to support ecological analysis and synthesis following on the recommendations of the FLED committee report. We have also been investigating the means by which high-quality collections of ecological data can be encouraged and supported. We discuss the issues involved in design and operation of the web-based data management system and the mechanisms being considered to manage intellectual property rights and provide long-term funding (5).

National Partnership for Advanced Computing Infrastructure (NPACI) 10/1/97-9/30/02 The Earth systems science area, led by Freeman Gilbert of the Scripps Institution of Oceanography (SIO) and John Helly of the San Diego Supercomputer Center (SDSC) at UCSD, is advancing the study of the Earth's natural systems and the complex interactions between humans and those systems. To do so, NPACI partners will focus on earth systems modeling and orchestrating multiscale models across distributed computers; ecological and environmental modeling that incorporates remote-sensing data; and an Earth systems digital library that will create new opportunities for scientists, educators, students, and policy makers.

In early efforts, NPACI is supporting climate researchers investigating the 1997-1998 El Niño phenomenon. To develop better predictions of conditions provoked by El Niño at global, regional, and watershed scales, NPACI researchers have assembled a set of compute- intensive multiscale, multiresolution models to be run on NPACI machines. One result will be a collection of global and regional climate and weather models residing within NPACI.

The Earth systems science team will build on work at UCSD (including SIO and SDSC), the University of Texas at Austin, UCLA, UC Irvine, UC Berkeley, UC Santa Cruz, the University of Maryland, the University of Kansas, the Long Term Ecological Research Network at the University of New Mexico, and partners from the technology thrust areas.

LTER Network Office 2/1/98-1/30/02 Funding for this activity has just begun and will provide support for support of the LTER system by SDSC.

Citations

1. San Diego Bay Project; http://sdbay.sdsc.edu

2. J. Helly, S. Levin, W. Michener, F. Davis, T. Case, The State of Computational Ecology. (1996) http://www.sdsc.edu/compeco_workshop/report/helly_publication.html

3. K. Gross, *et al*, Report of the Committee on the Future of Long-term Ecological Data (FLED). (1995) http://esa.sdsc.edu/FLED/FLED.html

4. W.K. Michener, J.W. Brunt, J.J. Helly, T.B. Kirchner, S.G. Stafford, Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications* (1997). **7**: (1):330-242.

5. J. Helly, T. Elvins, D. Sutton, D. Martinez, Controlled Publication of Digital Scientific Data: An Example in Ecology. *In preparation* (1999).

Michael R. Willig—Texas Tech University, Results of Prior Support

Long-Term Ecological Research in the Luquillo Mountains of Puerto Rico (DEB-9705814); 10/94-9/00—As co-PI, my responsibilities involve long-term monitoring of selected populations and communities of animals, including snails, walking sticks, and bats. I have documented their response to a variety of disturbances, including Hurricane Hugo, and have evaluated trajectories of recovery over the subsequent 9 years. Current efforts are continuing with respect to Hurricane Georges, for which the previous years of data provide the background context for assessment of immediate and short-term effects. Synthetic efforts involve 1) comparison of trajectories of recovery by populations, communities, and biogeochemical parameters, 2) development of a model to reflect the dynamics of recovery and interactions among elements of a disturbance regime, and 3) assessment of scale-dependence in the relationship between productivity and diversity. As a consequence, I have been an author on 14 abstracts at scientific meetings, and have 17 published or *in press* articles. In addition, I have been the major advisor of five graduate students working in the Luquillo Mountains**:** Secrest (MS, 1995), Alvarez (PhD, 1997), Yee (MS, 1999), Cox (PhD), and Bloch (PhD).

PUBLICATIONS

I. Gannon, M.R., and **M.R. Willig**. 1995. Ecology of ectoparasites from tropical bats. Environ. Entomology 24:1495-1503.

II. Walker, L.R., J. K. Zimmerman, **M.R. Willig**, and W.L. Silver, Eds. 1996. Long-term responses of Caribbean ecosystems to disturbance. Biotropica 28:414-614.

III. Walker, L.R., J. K. Zimmerman, **M.R. Willig**, and W.L. Silver, Eds. 1996. Long-term responses of Caribbean ecosystems to disturbance. Biotropica 28:414-614.

IV. Garrison, R.W., and **M.R. Willig**. 1996. Arboreal invertebrates. Pp. 183-245, in: The Food Web of a Tropical Rain Forest (R.B. Waide and D.P. Reagan, Eds.). Univ. of Chicago Press, Chicago, Illinois , 616 pp.

V. **Willig, M.R.**, and M.R. Gannon. 1996. Mammals. Pp. 397-431, in: The Food Web of a Tropical Rain Forest (R.B. Waide and D.P. Reagan, Eds.). Univ. of Chicago Press, Chicago, Illinois, 616 pp.

VI. **Willig, M.R.**, D.L. Moorhead, S.B. Cox, and J.C. Zak. 1996. Functional diversity of soil bacterial communities in the tabonuco forest: Interaction of anthropogenic and natural disturbance. Biotropica 28:471-483.

VII. Secrest, M.F., **M.R. Willig**, and L.L. Lind. 1996. The legacy of disturbance on habitat associations of terrestrial snails in the Luquillo Experimental Forest, Puerto Rico. Biotropica 28:502-514.

VIII. Zimmerman, J.K, **M.R. Willig**, L.R. Walker, W.L. Silver. 1996. Introduction: Disturbance and Caribbean ecosystems. Biotropica 28:756-765.

IX. Gannon, M.R., and **M.R. Willig**. 1997. The effect of lunar illumination on movement and activity of the red fig-eating bat (Stenoderma rufum). Biotropica 29:525-529.

X. **Willig, M.R.**, E.A. Sandlin, and M.R. Gannon. 1998. Structural and taxonomic correlates of habitat selection by a Puerto Rican land snail. Southwest. Natur. 43:70-79.

XI. **Willig, M.R.**, M.F. Secrest, S.B. Cox, G.R. Camilo, J.F. Cary, J. Alvarez, and M.R. Gannon. 1998. Long-term monitoring of snails in the Luquillo Experimental Forest of Puerto Rico: Heterogeneity, scale, disturbance, and recovery. Pp. 293-322, in: Forest Biodiversity in North, Central, and South American and the Caribbean: Research and Monitoring (F. Dallmeier and J.

Comisky, eds.). Man and the Biosphere Series, Volume 21, UNESCO and The Parthenon Press, Carnforth, Lancashire, UK.

XII. Gannon, M.R., and **M.R. Willig**. 1998. Long-term monitoring protocol for bats: Lessons from the Luquillo Experimental Forest of Puerto Rico. Pp. 271-291, in: Forest Biodiversity in North, Central, and South American and the Caribbean: Research and Monitoring (F. Dallmeier and J. Comisky, eds.). Man and the Biosphere Series, Volume 21, UNESCO and The Parthenon Press, Carnforth, Lancashire, UK.

XIII. **Willig, M.R.**, and M.A. McGinley. 1999. Animal responses to natural disturbance and roles as patch generating phenomena. Pp., 667-689, in: Ecology of Disturbed Ground, Elsevier Science, The Netherlands (In Press).

XIV. Walker, L.R. and **M.R. Willig**. 1999. An introduction to terrestrial disturbance. Pp. 000-000, in Ecology of Disturbed Ground, Elsevier Science, The Netherlands (In Press).

XV. **Willig, M.R.** and L.R. Walker. 1999. Disturbance in terrestrial ecosystems: salient themes, synthesis, and future directions. Pp. 000-000, in Ecology of Disturbed Ground, Elsevier Science, Amsterdam, The Netherlands (In Press).

XVI. Waide, R.B., **M.R. Willig**, G. Mittelbach, C. Steiner, L. Gough, S.I. Dodson, G.P. Juday, and R. Parmenter. 1999. The relationship between primary productivity and species richness. Ann. Rev. Ecol. & Syst. (In Press).

Paraguayan Mammals and Their Ectoparasites (DEB-9400926); 7/94-6/99—I am one of two PIs on this project to conduct a biotic survey and inventory of small mammals and their ectoparasites. We have completed extensive surveys of bats, marsupials, and rodents at 25 sites distributed in a manner to parallel environmental gradients of temperature and precipitation, and to sample all major biomes and regions designated for protection. We have collected over 10,000 mammals and their associated biological objects (e.g., karyotypes, heart, kidney, and lung tissues, skeletons, ectoparasite assemblages). Sorting and identifying ectoparasites is a major focus of current efforts, in collaboration with systematic colleagues throughout the country. A detailed monograph on the bats of Paraguay is ready for submission, and research is in progress concerning a similar tome for rodents. We are working on systematic research concerning sexual dimorphism and intraspecific variation in two common bats, *Sturnira lilium* and *Artibeus lituratus*. In addition, we are finalizing research concerning species abundance distributions, body size distributions, and dominance-diversity relationships of bats at the 25 sites. As a consequence, I have been an author on 9 abstracts, and 6 others will be presented in June. Four graduate student projects will emerge from this research, including the work of Lopez-G (PhD, 1998), Presley (PhD), Stevens (PhD), and Gorresen (MS).

Publications

1. Lopez-Gonzalez, C., S.J. Presley, R.D. Owen, and **M.R. Willig**. 1998. Noteworthy records of bats (Chiroptera) from Paraguay. Mastozoologia Neotropical 5:41-45.
2. **Willig, M.R.,** S.J. Presley, R.D. Owen, and C. Lopez-Gonzalez. Composition and structure of bat assemblages in Paraguay: A subtropical – temperate interface. J. Mamm. (In Press).

VII. DISSEMINATION OF RESULTS AND INSTITUTIONAL COMMITMENT

The proposed research will create multiple deliverables that will be disseminated to appropriate end users as indicated below. Project personnel have a strong and shared commitment to the widespread distribution and implementation of the products of their research.

Informatics and Biocomplexity Working Groups - Peer-reviewed journals will be the principal means for publication of scientific results from the activities of the Working Groups. Project participants will give presentations at scientific meetings, especially targeting special symposia on ecological data management to be organized at the annual meetings of the Ecological Society of America.

Graduate Seminar – Participation in graduate seminars on enabling biodiversity research through informatics will provide a major avenue for disseminating results to students throughout the nation. Over 100 students are projected to take part in these seminars. A special symposium dedicated to the integration and presentation of these results will take place at the LTER All Scientists meeting scheduled for August 2000. Multiple peer-reviewed publications are anticipated from this activity.

Software implementing desktop, repository and network services; manuals and documentation for end-users and systems personnel; and tutorials – These will be distributed directly to researchers affiliated with the project and will be freely available through the Web. The availability of these tools will be made known through the Ecological Society of America, whose publications reach the majority of the ecologists in the United States.

Operational knowledge network prototype - The network prototype, including the distributed data repository framework with advanced network level services, will be broadly available and easily accessible through the Internet.

NCEAS facilities, located off-campus in downtown Santa Barbara, house a permanent staff of 11, including the Director and Deputy Director, four administrative support personnel, and five scientific computing personnel. The Center sponsors approximately 30 resident scientists and over 400 visiting scientists per year.

NCEAS maintains a heterogeneous, networked computing environment for its resident and visiting scientists. The internal backbone consists of 10 and 100 Mbps switched ethernet. Internet connectivity is achieved via a dedicated T1 connection to the UCSB campus. NCEAS' technical staff are participating in campus planning for the CalRen-II network, which will bring OC-3 (155 Mbps) and eventually OC-12 (622 Mbps) connectivity to UCSB as part of the NSF-sponsored vBNS. NCEAS maintains two high-end database and analytical servers: a Silicon Graphics Origin 2000, with four R10K CPU's, and 1.5 GB of RAM; and an SGI Power Challenge L, with four R10K CPU's, and 1.5 GB of RAM. Storage includes over 45 GB online disk storage and a 75 GB robotic DLT tape backup system. All servers are archived nightly, and are on uninterruptible power supplies. High-end PCs, Macintoshes, and UNIX workstations are available to resident and visiting researchers. Planned upgrades with direct relevance to supporting an informatics venture include addition of a large, high-availability storage system, RAM upgrades for the servers, and an automated high-volume, high-speed tape backup system. NCEAS supports a number of scalable software packages, including SAS, MATLAB, Splus, CPLEX, ArcInfo and ArcView, and parallelizable compilers for Fortran 90, Fortran 77, C++, and C. Oracle, running on our SGI Power Challenge, is our primary database server.

The **LTER Network Office** occupies a suite of seven offices, two computer laboratories, and a conference room on the University of New Mexico South Campus. The Network Office houses staff and computer facilities for the LTER Network Information System Infrastructure. The backbone of the network information system is the network office data center. The center has recently installed new scalable servers and increased the network bandwidth to better serve the LTER and ecological community. Two Sun 450 Enterprise servers (4-300 MHz UltraSPARC cpu's, 1GB memory, 20 GB local disk, tape backup including DDS3- 12/24 GB, redundant power supplies and uninterruptible power) and two Dell Poweredge servers (2-300 MHz Pentium II, 256 MB memory, DDS3 tape backup, a 45 GB RAID5 disk, redundant power supplies and UPS) are the primary computers serving the lternet domain. Unix and NT 4.0 versions of Erdas/Imagine and Arc/Info are used for GIS support. MS SQL server is used for databases, with plans to implement Unix Oracle (UNM site license). The office is connected to the Internet and vBNS via ATM (155 Mbps) and 100 Mb/s UNM fiber backbone In 1999, the UNM fiber backbone will be upgraded to 1 Gb/s. The LTER Network Office server interfaces will be simultaneously upgraded to take advantage of this new connectivity.

SDSC will provide resources via the National Partnership for Advanced Computational Infrastructure (NPACI). These include supercomputers, archival storage systems, data handling platforms, and advanced visualization systems. During the next five years, the capabilities of the center will be upgraded to include peak teraflops-capable systems, petabyte archives, and high-performance data-handling systems that provide a robust test bed for information management research. Major hardware resources currently at SDSC include a 14-processor CRAY T90, a 256-processor CRAY T3E, a 128-node IBM SP, over 2 TB of disk, and over 100 TB of tape storage. These are being augmented by a novel multithreaded supercomputer, a Tera MTA. During the next five years, we plan to upgrade an IBM SP system to provide a data-handling environment with a petabyte of storage, 10 TB of disk, and a sufficiently large number of nodes to sustain data movement rates in excess of 3 GB/s. A peak teraflops parallel compute engine will be connected to the data-handling environment using emerging Super HIPPI LAN technology. SDSC is located on several large networks that will be essential to moving data within our proposed data network. First, the NSF-sponsored very High Speed Backbone Network Service (vBNS) interconnects UCSD/SDSC with its national NPACI partners. It operates at 155 Mbps (OC-3) and parts of the backbone network run at 622 Mbps (OC-12). Connection to UCSB is provided through the CalREN-2 network.

Associated data-intensive computing software includes a distributed data handling system developed at SDSC (Storage Resource Broker), digital library technology acquired through collaborations with IBM, University of California, Santa Barbara and University of California, Berkeley, parallel object-relational database technology acquired in collaboration with IBM, and the High Performance Storage System (HPSS) archival storage software that is being developed and tested with IBM and LLNL. The archival storage system at SDSC sustains over a terabyte of data movement per day, with peak access rates over 2000 transactions per hour.

The NPACI program is creating scientific data collections for disciplines including Neuroscience, Molecular Science, Earth Systems Science, and Astronomy. Access to these data collections is provided through the SDSC Storage Resource Broker. The combination of information management technology, scientific data collections, and the data handling platforms that support rapid access to the data provides an excellent test bed for evaluating new infrastructure for managing scientific data and scientific algorithms.

VIII. PERFORMANCE GOALS

Research and development activities that will lead to the construction of our knowledge network prototype have been modularized to allow our multiple institutions to work simultaneously on different components of the network. Although the components will be ultimately integrated, each will have well-specified APIs that allow relatively independent progress. This process will be facilitated by the use of component-oriented software design practices (e.g., use of CORBA interfaces). Table 1 describes the major performance goals for each phase of the project.

TABLE 1: Performance Goals

DATA—Unified Access to Distributed Data (LTERNet, NCEAS)	
Year 1:	Install SRB on LTER, NCEAS systems, develop CORBA interface to SRB
	Extend SRB deployment to 6 core LTER sites
	Develop, deploy, and test metadata catalog based on Extended EML
Year 2:	Continue SRB deployment to extended set of field stations
	Develop query interface for metadata catalog
Year 3:	Develop metadata conversion tools (site-specific)
	Test and refine metadata and data client interfaces
	System integration with other components
INFORMATION—Resolving Semantic Heterogeneity via Schema Integration (NCEAS, LTERNet)	
Year 1:	Formalize extensible metadata structure in RDF
	Develop prototype metadata editor / query interface to catalog
	Develop metadata schemata for dataset structure
	Develop metadata schema for data set semantics
	Research technologies for concept representation
Year 2:	Continue research on concept representation
	Develop data integration engine
	Specification of concept ontologies for case study
Year 3:	Refine conceptual query capabilities and data integration engine
	System integration with other components
KNOWLEDGE—Automated Modeling of Scientific Hypotheses (SDSC, NCEAS)	
Year 1:	Develop Hypothesis Specification Language
Year 2:	Develop APN and Conditional Probability estimators
Year 3:	Test hypothesis modeling with biocomplexity case study
	System integration with other components
BIOCOMPLEXITY RESEARCH—Multi-scale Research on Biodiversity and Ecosystem Function (TTU, UNM, NCEAS)	
Year 1:	Validation, test data discovery and access
Year 2:	Expansion and Amplification, test data integration and query
Year 3:	Amplification and Extrapolation, test hypothesis modeling
EDUCATION, OUTREACH, and TRAINING—(NCEAS, LTERNet, ESA)	
Year 1:	Conduct graduate training seminars, outreach via ESA
Year 2:	Conduct graduate training seminars, outreach via ESA
Year 3:	Conduct graduate training seminars, outreach via ESA

IX. MANAGEMENT PLAN

This project is organized to provide leadership through vision, decision making, and accountability. The management structure reflects the strengths and commitments of the institutions to ensure success of the project, and is modeled on the successful management structure of the Protein Data Bank, led by Rutgers University and distributed over the National Institution for Standards and Technology and UCSD/SDSC.

There are two levels of management: a Core Management Team and the Project Team. **The Core Management Team**, consisting of Jim Reichman, PI, Bob Waide, Peter Arzberger, and Mike Willig will be responsible for coordinating activities at their respective sites, ensuring the institutional commitments to this project, and allocating resources to ensure project success. The PI is responsible for developing and articulating the goals, implementing plans to meet goals, and ensuring success of the overall project.

The **Project Team** will consist of the PI, all the co-PIs, and senior personnel. The project team is involved in all functions of management: maintaining project focus, integrating activities, allocating resources, continuously assessing the project, and planning for future activities. Monthly teleconferencing and at least semi-annual meetings of the project team, chaired by the PI, will ensure coordination of the research and development activities, meeting the deadlines for deliverables, and planning new activities. Table 2 indicates the principal research and development activities and the responsible lead.

TABLE 2: Project Team and Activities (* Indicates Activity Lead)

Principal activities	**Institution**	**Core Team**	***Members***
DATA Unified Access to Distributed Data	LTERNet/ UNM	Waide	Brunt*
INFORMATION Resolving Semantic Heterogeneity via Schema Integration	NCEAS/ UCSB	Reichman	Jones*, Schildhauer
KNOWLEDGE Automated Modeling of Scientific Hypotheses	SDSC/ UCSD	Arzberger	Helly*, Rajasekar, Sutton
BIOCOMPLEXITY RESEARCH Multi-scale Research on Biodiversity and Ecosystem Function	TTU, UNM, NCEAS	Willig	Willig*, Waide, Andelman

Each of these individuals has a strong record in related work, and in many cases individuals have worked together on prior projects (see results from Prior Support). We will constitute an Advisory Committee of approximately 10 individuals to ensure that our efforts will impact the ecological sciences and beyond, to assist us in important strategic decisions, and to monitor our progress. This committee will have members from groups directly affected by the advances from this project: data providers and users (OBFS); community best practice promulgators (ESA); technological developers (Computer Science and bioinformatics community); and the broader biocomplexity researchers (Museum Collections, NBII). A number of individuals have agreed to serve on the Advisory committee, which will be formalized should an award be made. We plan to meet with the committee four times during the three years, the first meeting being held within six months of the start of the award.

X. LETTERS OF COOPERATION / COMMITMENT

Archbold Biological Station – Swain

Robert B. Waide
Executive Director
LTER Network Office
Department of Biology
University of New Mexico
Albuquerque, NM 87131-1091

5/4/99

Dear Bob,

Supporting letter from Archbold Biological Station, Florida for:

A Knowledge Network for Biocomplexity: Building and Evaluating a Metadata-based Framework for Integrating Heterogeneous Scientific Data

Archbold Biological Station is interested in cooperating in this new venture to develop software tools for ecologists at field stations to simplify searches for ecological data. We would like to formally participate in this proposed program.

Archbold Biological Station is a privately endowed research facility located in south-central Florida devoted to long-term ecological research and conservation, part of the global effort to understand, interpret, and preserve the earth's natural diversity. Archbold Biological Station would contribute to this partnership access to extensive ecological data sets, and a network of scientists specializing in ecology and synthesis of ecological data. Station facilities http://www.archbold-station.org incorporate a library, laboratories, reference collections, plus dining and housing. A state-of-the-art computational infrastructure includes Internet, Intranet, GIS and Local and Wide Area Networking technologies. Staff and visiting scientists conduct research on the unique scrub habitats of the Lake Wales Ridge and environs. The Station owns and manages a 2,000-ha natural preserve, a relict sand dune habitat with one of the highest concentrations of threatened and endangered species in the USA. Study and management of endemic, endangered species and communities form dominant research themes. Ecological research programs are also conducted in other scrub ecosystems regionally, and intensive conservation activities have been focused on the establishment of protected areas throughout the Lake Wales Ridge. Educational programs emphasize research opportunities for graduate and undergraduate students, providing facilities for teaching field ecology to visiting classes, and environmental instruction for schoolchildren.

A major division of Archbold Biological Station is the MacArthur Agro-ecology Research Center (the Ranch), a 4,170-ha working cattle ranch and citrus grove managed at commercial production levels for research purposes. The Ranch provides staff and visiting scientists an opportunity to measure and monitor ecological effects of agricultural practices at real world scales of space and numbers. It serves as a working laboratory to explore the relationships among the ecological, economic, and socio-political functions of this agricultural ecosystem.

Archbold Biological station could play three important roles in this project:

- First, we would make available Station data collected on species richness/biodiversity, ecosystem function, and climate - thus contributing to the validation of the data retrieval system proposed.
- Second we can provide feedback on the functionality of the knowledge network.
- Finally, we would like to participate in a scientific working group at NCEAS involved in a study of the relationship between biodiversity and ecosystem function

We would wish to cooperate with staff from the LTER Network Office to configure new software for data management and would be very interested in technical support for our data management activities under this proposal. We would be willing to send a representative from Archbold to participate in four workshops over a three-year period, with expenses covered by the project.

In conclusion Archbold Biological Station is interested in this ambitious, worthwhile proposal and willing to support this project.

Sincerely,

Hilary Swain

Hilary Swain, Ph.D.
Executive Director
Archbold Biological Station
PO Box 2057
Lake Placid, FL 33862
telephone: 941-465-2571
voice mail: x251 for business, x241 for personal
fax: 941-699-1927
email: hswain@archbold-station.org

The University of New Mexico

Department of Biology
167 Castetter Hall
Albuquerque, NM 87131-1091
(505) 277-3411

May 11, 1999

Mr. Matt Jones
National Center for Ecological Analysis and Synthesis
735 State St., Suite 300
Santa Barbara, CA 93101-3351

Dear Matt:

As Chair of the Coordinating Committee of the Long Term Ecological Research (LTER) Network, I am pleased to endorse your proposal titled: A Knowledge Network for Biocomplexity: Building and Evaluating a Metadata-based Framework for Integrating Heterogeneous Scientific Data. The intellectual advances in information science that you propose will, for the first time, provide an accessible infrastructure for identifying, integrating, managing, and, ultimately, synthesizing the nation's ecological and biodiversity information resources. Ultimately, the existence of a knowledge network for ecological information will serve directly to provide information to scientists, students, policy planners, resource managers, and the public, and will allow them to make science-based decisions regarding the future of our society.

We believe that the LTER Network of 21 sites can play several important roles in this project. By making available data collected at LTER sites, we can contribute to the validation of the data retrieval system you will construct. By using beta versions of the software you will develop, we can provide important feedback on the functionality of the knowledge network. Finally, by participating in scientific working groups at NCEAS and the planned graduate seminars, we can contribute to the study of the relationship between biodiversity and ecosystem function.

We wish you the best of luck in your proposal and look forward to the resulting collaboration.

Sincerely,

James R. Gosz
Department of Biology
University of New Mexico

From: Jack Stanford <stanford@selway.umt.edu> 5/1/99 15:19
Subject: participation in LTER proposal
To: jones@nceas.ucsb.edu
CC: "Robert Waide" <rwaide@lternet.edu>

Dear Dr. Jones,

I offer our full support for the proposal entitled "A Knowledge Network for Biocomplexity: Building and Evaluating a Metadata-based Framework for Integrating Heterogeneous Scientific Data," which is being prepared for NSF under the KDI initiative.

We at Flathead Lake Biological Station are pleased to be invited to participate
and will make our expertise and long term data fully available to the project.

Sincerely,
Jack A. Stanford.

Jack A. Stanford
Jessie M. Bierman Professor of
Ecology and Director
Flathead Lake Biological Station
The University of Montana
311 Bio Station Lane
Polson, MT 59860-9659
Phone: 406/982-3301
FAX: 406/982-3201
Check out our Web Page: <http://www.umt.edu/biology/flbs>

From: "Mark R. Stromberg" <stromber@socrates.berkeley.edu> Thu 13:23
Subject: Re: Joint project
To: <rwaide@lternet.edu>
CC: jones@nceas.ucsb.edu

Dear Bob,

I have been away from my e-mail for a few days.

We at Hastings would be pleased to join in your project and participate and offer any assistance possible. It looks like a timely and useful effort. As you may know, Dr. Walter Koenig has been using large databases in ecological studies in relation to smaller data sets in California.

Thank you for thinking of Hastings.

Sincerely,

Mark

Mark R. Stromberg, Ph.D.
Resident Director
Hastings Natural History Reservation
Museum of Vertebrate Zoology
University of California, Berkeley
38601 E. Carmel Valley, CA 93924
Carmel Valley, CA 93924
Office: 831 659-2664 fax: 659-7208
stromber@socrates.berkeley.edu

Hastings Reserve Information:
http://nrs.ucop.edu/reserves/hastings/hastings.html

Organization of Biological Field Stations
http://www.obfs.org

From: Michael Hamilton <director@jamesreserve.edu> Fri 10:24
Subject: Joint project
To: jones@nceas.ucsb.edu
CC: rwaide@lternet.edu, amckee@lternet.edu

Dear Matt,

I am writing you, at the request of Robert Waide, to initiate my collaboration in your proposal to the NSF KDI program. My UC-NRS field Station, the James San Jacinto Mountains Reserve http://www.jamesreserve.edu , has databases and a data management infrastructure which I feel will lend itself nicely within the objective of your proposal. I will be willing to share data, post new data to our web sites, and participate in metadata protocol development that would be appropriate for our station's needs. My own research involves ecological metadata support systems, and so I would be willing to share my ideas as an exchange within your consotium.
I have a somewhat old CV located at http://www.jamesreserve.edu/staff/hamilton.html, and I encourage you to explore our web site to gain a sense of what resources and capabilities I can offer to your proposal. I am pasting into this email a short description of my recent work with regards to data management and field station applications which we submitted in a pending proposal. I hope it is helpful.

Good luck with your proposal,

Best regards,

Michael Hamilton
Director

The James San Jacinto Mountains Reserve: The Biodiversity Visualization Laboratory (BVL) was established at the James San Jacinto Mountains Reserve in 1985 by Dr. Michael Hamilton to explore applications of remote sensing, digital media, and database technologies. For two decades, Dr. Hamilton, the James Reserve's manager and an internationally recognized expert in ecological monitoring, has directed biodiversity inventory projects in California, Venezuela, Russia, Mexico, South Africa, Namibia and Botswana, and has provided technical advice and training for more than 150 International conservation groups. He is a founding board member of the International Society for Conservation Geographic Information Systems http://www.scgis.org. For additional information about the activities of Dr. Hamilton and the James Reserve, please refer to the Reserve's web site: http://www.jamesreserve.edu.

For the past decade, the BVL has focused on the most effective methods to:

… Radically improve our ability to accurately quantify and analyze hundreds of species within complex, dynamic ecological systems without an enormous investment in human manpower

… Increase global collaboration among conservation biologists, ecologists and other field scientists in an entirely new, yet cost effective manner

… Accurately translate field research conducted within the framework of one discipline into appropriate parameters for use across disciplines.

These challenges formed the basis for the **Macroscope,** a unique information management technology that enhances environmental research. The term "**MACROSCOPE**," is an acronym for **MAC**hine **R**epresentations **O**f **SC**aleable **O**bjects and **P**rocesses in **E**cosystems. Early prototypes of the award-winning Macroscope demonstrated how digital hypermedia information structures could integrate multiple scales of organismal, ecological and environmental data. Applications from this system have been multi-disciplinary and global in context. For example, Macroscope projects conducted by Dr. Hamilton were developed to; a) quantify and predict the flow of radioactive nuclides through farmlands downwind from the Chernobyl reactor, b) model the interaction between vegetation, wildfire, and rare and endangered species in California and Mexico; c) teach conservation science in southern Africa; d) catalogue and integrate research collections at biological field stations in Venezuela.

Success with these projects demonstrated opportunities for providing rapid and efficient access to vast quantities of environmental information, for evaluating research on more complex systems, and for increasing our understanding of complex living systems.

The Macroscope Ecological Observatory: The science of biodiversity is impressively complicated to define and categorize. It attempts understanding of living systems across extremes of spatial scale (molecular to global proportions), of temporal scale (physiological response times to evolutionary change), and of levels of biological organization spanning DNA to global biomes. Scientists who study biodiversity come from diverse academic fields that do not share a common scale to observe and describe biodiversity. A central activity of the Macroscope Ecological Observatory will address this problem by demonstrating how convergent digital information structures and technologies can unify a new range of biological scales and environmental phenomena.

Our current research involves:

… New tools for collecting ground-based remote sensing data, down to centimeter resolution, which automatically update the virtual ecosystem models, and provide quantitative measurements of change at multiple scales

… New techniques to create web-based virtual ecosystems -- realistic, spatially explicit graphical representations of biodiversity which are accurate to the individual scale of objects in space and time

… Simulation models driven by real-time automatic data acquisition, that are linked to virtual ecosystems for visualization and interpretation

----

Michael P. Hamilton
Director, James San Jacinto Mountains Reserve
Adj. Asst. Professor, Department of Biology, UCR
Mailing Address:
UC JAMES RESERVE
54445 North Circle Drive
PO BOX 1775
IDYLLWILD, CA 92549-1775
Phone: (909) 659-3811
Fax: (909) 659-2182
E-mail: director@jamesreserve.edu
WWW: http://www.jamesreserve.edu

"Not all who wander are lost"

J. R. R. Tolkien

From: "Jorge A. Jimenez" <jjimenez@cro.ots.ac.cr> 5/3/99 16:52
Subject: Joint Project
To: jones@nceas.ucsb.edu
CC: <amckee@lternet.edu>, "Robert Waide" <rwaide@lternet.edu>, "rmatlock@sloth.ots.ac.cr" <rmatlock@sloth.ots.ac.cr>, "Gary S. Hartshorn" <ghartsho@duke.edu>

Dear Dr. Jones

This message is to indicate our desire to make La Selva Biological Station one of the participants in the project: "A Knowledge Network for Biocomplexity:
Building and Evaluating a Metadata-based Framework for Integrating Heterogeneous Scientific Data".
We are in agreement to make available, for the purpose of this project, the available data on species richness/biodiversity, net primary productivity, and climate, from our data bases.
We are very enthusiastic about our participation in this project and wish you the best luck on the proposal submission.

Sincerely,

Dr. Jorge A. Jimenez
Director in Costa Rica
Organization for Tropical Studies
San José, Costa Rica
P.O. Box 676-2050
San Pedro, Costa Rica
Ph: (506) 240-6696
Fax: (506) 240-6783

From: Marjorie Holland <mholland@olemiss.edu> Mon 10:36
Subject: Joint project
To: jones@nceas.ucsb.edu
CC: rwaide@lternet.edu

The University of Mississippi Field Station is interested in participating in significant activities that will improve the informational infrastructure of the biological sciences, particularly when those activities complement our own efforts to develop visualization and presentation systems for integrated ecological data. There is much work to be done, and cooperation across the OBFS, LTER, and greater ecological community is critical. We are interested in participating in the workshop you mention, and look forward to receiving details of the proposed project.

Thank you for contacting us.

M.M. Holland, director, The University of Mississippi Field Station
Shoemaker Hall, The University of Mississippi
University, MS 38677 USA
phone= 662-232-5874

May 12, 1999

Dr. O. J. Reichman, Director
NCEAS
735 State Street, Suite 300
Santa Barbara, CA 93101

Dear Jim:

The Ecological Society of America is pleased to write in support of your proposal entitled "A Knowledge Network for Biocomplexity: Building and Evaluating a Metadata-based Framework for Integrating Heterogeneous Scientific Data". This proposal is a natural complement to our previous collaboration related to the publication of ecological data which grew out of the ESA Future of Long-term Ecological Data (FLED) committee. If successful, this proposed effort will result in standards and methods which the ESA would be pleased to consider advocating under the Education, Outreach and Training component of the grant. In that regard, we appreciate the support you are requesting to ensure that this component of your project can be facilitated by the ESA. We look forward to a successful proposal and to a continued collaboration.

Sincerely,

Katherine Gross, President, ESA
Kellogg Biological Station
Michigan State University
Hickory Corners, MI 49060

SiliconGraphics
Computer Systems

614 Chapala Street
Santa Barbara. California 93 101
Telephone (805) 884-7812
FAX (805) 963-0410
E-mail: janes@losangeles.sgi.com
http://www.sgi.com

May 12, 1999

Dr. Mark Schildhauer
National Center for Ecological Analysis and Synthesis
735 State Street Suite 300
Santa Barbara, CA 93 101

Dear Dr. Schildhauer,

Thank you for the opportunity to partner with the National Center for Ecological Analysis and Synthesis (NCEAS) on your proposal to the National Science Foundation's "Knowledge and Distributed Intelligence in the Information Age" program, specifically "Ecological Informatics". We at Silicon Graphics have a company wide commitment to environmental applications and we are intrigued with your idea. We have been working closely with the chemistry and biological sciences on solutions to bioinformatics to accelerate time to discovery. The notion that informatics can accelerate research projects in ecological research seems natural and technically appealing. We would welcome the opportunity to join you in this research.

We envision our partnership as twofold: (1) access to technical resources and (2) contribution towards workstations and servers.

The focal point for technical support will come from the Environmental Sciences Group (ESG), part of the Applications Division of Silicon Graphics. This is a multi-disciplinary group with expertise in weather forecasting, climate, ocean, air pollution, and groundwater analysis packages and visualization of model-produced environmental data. This would be an ideal opportunity to expand the scope of the ESG to tackle challenging data integration and modelling issues emerging from interactions of biotic and abiotic processes. The ESG is engaged in joint projects that help SGI improve application performance and understand the high performance computer market for environmental sciences. The ESG provides benchmarking, code porting and optimization, consultation and participation in commercial and academic forums.

We have been long time advocates of integrating the power to manage and analyze very large datasets with leading visualization capabilities. For the innovative new user interfaces that you describe, Silicon Graphics offers these technologies today to save valuable software development time: OpenGL, Cosmo Worlds for VRML creation, OpenGL Optimizer, Imagevision, Performer, Mineset and Cosmo software for web development. Access to these product groups can be coordinated by a representative in the ESG or through a technical system engineer in our local office. We also have a close relationship with many independent software vendors (ISVs) that enable scientific visualization. We will encourage our ISVs to provide evaluation software and working

614 Chapala Street
Santa Barbara California 93101
Telephone (805) 884-7812
FAX (805) 963-0410
E-mail: janes@losangeles.sgi.com
http://www.sgi.com

models from demos to accelerate your development efforts. Our award winning video streaming product, Mediabase, which is rapidly being embraced for distance learning in education may prove valuable for your education and outreach program.

Silicon Graphics will initially provide 100 hours of access to our product and industry experts and specialists. The hourly rate for such consulting averages $200 per hour. This contribution represents $20,000 in consulting services. We will also offer the opportunity to enlarge this commitment through mutual consent of the participating parties. This offer will be open for a two year period.

Finally, Silicon Graphics has much software and the high bandwidth workstations and servers ideally suited for this endeavor. Because of our interest in being part of this research, we would like to extend an additional contribution to the project in the form of an increased discount, beyond the educational discount, of 15% not to exceed a total 50% discount for our Unix based workstations and servers. This contribution is an indication of partnership and our desire to push the envelope of technology on the Silicon Graphics platform. We will extend this offer for a two year period. We wish to be a visible partner in this project which promises far-reaching benefits for ecological research.

Sincerely,

Debbie James

Debbie
Account Manager
Silicon Graphics

XI. References Cited

[1] Baru, C., R. Frost, R. Marciano, R. Moore, A. Rajasekar, and M. Wan. 1997. Metadata to Support Information-based Computing. IEEE Conference on Metadata, September 1997, Greenbelt, MD.

[2] Baru, C., R. Moore, A. Rajasekar, W. Schroeder, M. Wan, R. Klobuchar, D. Wade, R. Sharpe, and J. Terstriep. 1998. Data Handling Architecture for a Prototype Federal Application. IEEE Conference on Mass Storage, March 1998, Silver Spring, MD.

[3] Baru, C., R. Moore, A. Rajasekar, and M. Wan. 1998. The SDSC Storage Resource Broker. Proceedings of CASCON'98 Conference, Nov.30-Dec.3, 1998, Toronto, Canada.

[4] Batini, C., M. Lenzerini, and S.B. Navathe. 1986. A comparative analysis of methodologies for database schema integration. ACM Computing Surveys 18(4): 323-364.

[5] Beech, D., S. Lawrence, M. Maloney, N. Mendelsohn, and H. S. Thompson. 1999. XML schema part 1: structures. World Wide Web Consortium Working Draft 06 May, 1999. Available at: "http://www.w3.org/1999/05/06-xmlschema-1".

[6] Biron, P. V. and A Malhotra. 1999. XML schema part 2: datatypes. World Wide Web Consortium Working Draft 06 May, 1999. Available at: "http://www.w3.org/1999/05/06-xmlschema-2".

[7] Bray, T., J. Paoli, and C. M. Sperberg-McQueen. February, 1998. Extensible Markup Language (XML) 1.0. Available at "http://www.w3.org/TR/REC-xml".

[8] Brickley D., R.V. Guha, eds. 1999. Resource Description Framework (RDF) Schema Specification. W3C Proposed Recommendation 03 March 1999. Available online at http://www.w3.org/TR/PR-rdf-schema.

[9] Chapin, F.S., III, A.E. Giblin, J.A. Laundre, K.J. Nadelhoffer, and G.R. Shaver. 1995. Responses of arctic tundra to experimental and observed changes in climate. Ecology 69:694-7111.

[10] Daniel, R. Jr., and C. Lagoze. 1997. Extending the Warwick Framework: From Metadata Containers to Active Digital Objects. D-Lib Magazine. ISSN 1082-9873. Volume: November, 1997.

[11] Daniel, R. Jr., C. Lagoze, and S. D. Payette. 1998. A metadata architecture for digital libraries. Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries. IEEE Computer Society, Los Alamitos, California. pp. 276-288.

[12] FGDC. 1998. FGDC-STD-001-1998. Content Standard for Digital Geospatial Metadata (revised June 1998). Federal Geographic Data Committee. Washington, D.C.

[13] Frondorf, A., M. B. Jones, and S. Stitt. 1999. Linking the FGDC geospatial metadata content standard to the biological/ecological sciences. Proceedings of the Third IEEE Computer Society Metadata Conference. Bethesda, MD. April 6-7, 1999.

[14] Gallagher, R. and T. Appenzeller. 1999. Special Feature "Complex Systems—Beyond Reductionism." *Science* 284:79-109

[15] Gough, L, C.W. Osenberg, K.L. Gross, and S.L. Collins. Fertilization effects on species density and primary productivity in herbaceous plant communities. Ecology (Submitted).

[16] Gross, K.L., C.E. Pake, E. Allen, C. Bledsoe, R. Colwell, P. Dayton, M. Dethier, J. Helly, R. Holt, N. Morin, W. Michener, S.T.A. Pickett, and S. Stafford. 1995. Final report of the Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED). Volume I: Text of the report. 122 pp.

[17] Gross, K.L., M.R. Willig, L. Gough, R. Inouye, and S.B. Cox. Patterns of species diversity and productivity at different spatial scales in herbaceous plant communities from LTER sites. Ecology (Submitted).

[18] Hartvigsen, G., A. Kinzig, and G. Peterson. 1998. Special Feature " Complex Adaptive Systems: Use and Analysis of Complex Adaptive Systems in Ecosystem Science." *Ecosystems* 1:409-463.

[19] Heckerman, D. E. 1995. Learning Bayesian Networks. Microsoft Corporation, Redmond, WA, Technical Report MSR-TR-95-02, December (Revised 1995).

[20] Heckerman, D. E. 1996. A Comparison of Scientific and Engineering Criteria for Bayesian Model Selection. Microsoft Corporation, Redmond, WA MSR-TR-96-12.

[21] Heckerman, D. E. and B. N. Nathwani. 1992. "An evaluation of the diagnostic accuracy of pathfinder. *Computers and biomedical research* 25(1): 56-74.

[22] Heckerman, D. E. and D. Geiger. 1995. Likelihoods and Parameter Priors for Bayesian Networks. Microsoft Corporation, Redmond, WA, Technical Report MSR-TR-95-54, November 1995.

[23] Heckerman, D. E., E. J. Horvitz, and B. N. Nathwani. 1992**.** Toward normative expert systems .1. The pathfinder project. M*ethods of Information in Medicine*. 31(2): 90-105.

[24] Helly, J. J. 1984. A Representational Basis for the Development of a Distributed Expert System for Space Shuttle Flight Control. NASA Johnson Space Center, Houston, TX, NASA Flight Control DivisionTechnical Memorandum 58258, June 1984.

[25] Jordan, M. 1999. Learning in Graphical Models. ISBN 0-262-60032-3. MIT Press: Cambridge, MA. pp. 648.

[26] Kent, R. L. 1999. RDF Schema in OML/CKML. Available at "http://asimov.eecs.wsu.edu/CKRMI/RDF-Schema.html".

[27] Lagoze, C. and D. Fielding. 1998. Defining Collections in Distributed Digital Libraries. D-Lib Magazine. ISSN 1082-9873. Volume: November, 1998.

[28] Lassila, O. 1998. Web Metadata: A matter of semantics. IEEE Internet Computing 2: 30-37.

[29] Lassila, O. and R. R. Swick. 1999. Resource Description Framework (RDF) Model and Syntax Specification: W3C Recommendation 22 February 1999. World Wide Web Consortium. Available at "http://www.w3.org/TR/REC-rdf-syntax".

[30] Leung, Y. 1997. Intelligent Spatial Decision Support Systems. Heidelberg: Springer-Verlag.

[31] Likens, G. E. 1992. Excellence in Ecology. 3. The Ecosystem Approach: Its Use and Abuse. Ecology Institute, Oldendorf/Luhe, Germany.

[32] Lyons, K.A., and M.R. Willig. 1999. A hemispheric assessment of scale-dependence in latitudinal gradients of species richness. Ecology (In Press).

[33] Magurran, A.E. 1988. Ecological Diversity and Its Measurement. Princeton University Press, Princeton, New Jersey.

[34] Manola, F. 1999. Technologies for a Web Object Model. IEEE Internet Computing 3: 38-47.

[35] Michener, W. K., A. B. Miller and R.W. Nottrott. 1990. LTER Core Data Set Catalog. Belle Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, South Carolina.

[36] Michener, W. K., J. W. Brunt, J. Helly, T. B. Kirchner, and S.G. Stafford. 1997. Non-geospatial metadata for the ecological sciences. Ecological Applications 7:330-342.

[37] Mittelbach, G.G., C.F. Steiner, K.L. Gross, H.L. Reynolds, S.M. Scheiner, R.B. Waide, M.R. Willig, and S.I. Dodson. The relationship between species richness and productivity depends on scale. Ecology (Submitted).

[38] Moore, R. 1998. Data Caches Form Framework for Distributed Data-handling Testbed, NPACI & SDSC EnVision 13:1 10-11.

[39] Morgenstern, M. 1997. Metadata for Heterogeneous Databases. In: The Second IEEE Metadata Conference, Silver Springs MD. Available at "http://www.llnl.gov/liv_comp/metadata/md97.html".

[40] Morik, K., S. Wobel, J.-U. Kietz, and W. Emde. 1993. Knowledge Acquisition and Machine Learning: Theory, Methods, and Applications. San Diego: Academic Press.

[41] Nottrott, R., M. B. Jones, and M. Schildhauer. 1999. Using XML-structured metadata to automate quality assurance processing for ecological data. Proceedings of the Third IEEE Computer Society Metadata Conference. Bethesda, MD. April 6-7, 1999.

[42] Ober, J. 1999. The California Digital Library. D-Lib Magazine. ISSN 1082-9873. Volume: March, 1999.

[43] Odum, E. P. 1971. Fundamentals of Ecology. 3rd Edition. Saunders, Philadelphia, Pennsylvania.

[44] Paepcke, A., M.Q.W. Baldonado, C.K. Chang, S. Cousins and H. Garcia-Molina. 1999. Using Distributed Objects to Build the Stanford Digital Infobus. IEEE Computer 32: 80-87.

[45] PCAST. 1998. Teaming with Life: Investing in Science to Understand and Use America's Living Capital. President's Committee of Advisors on Science and Technology: Panel on Biodiversity and Ecosystems. Available at: "http://www1.whitehouse.gov/WH/EOP/OSTP/Environment/html/teamingcover.html".

[46] Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, California: Morgan Kaufmann.

[47] Pickett, S.T.A., J. Kolasa, and C.G. Jones. 1994. Ecological Understanding: The Nature of Theory and the Theory of Nature. Academic Press, San Diego, California.

[48] Rajasekar, A., R. Marciano, R. Moore. 1999. Collection-based Persistent Archives. IEEE Symposium on Mass Storage Systems, San Diego, California, February 1999, 176-185.

[49] Reichman, O. J. 1999. Activities and opportunities at the National Center for Ecological Analysis and Synthesis. Bulletin of the Ecological Society of America. Volume 80 (2): 99-102.

[50] Russell, S. J. and P. Norvig. 1995. Artificial Intelligence: A Modern Approach: Prentice-Hall.

[51] Russell, S. J., J. Binder, and D. Koller. 1994. Adaptive probabilistic networks. University of California, Berkeley, CA, Technical Report UCB CSD-94-824, July 25 1994.

[52] Scheiner, S.M., S.B. Cox, M.R. Willig, G.G. Mittelbach, C. Osenberg, and M. Kaspari. Species diversity: scale effects and species area curves. Ecology (Submitted).

[53] Schroeder, W. 1997. The SDSC Encryption/Authentication (SEA) System. Available at: "http://www.sdsc.edu/~schroede/sea.html".

[54] Schulze, E.D., and H.A. Mooney. 1994. Ecosystem function of biodiversity: a summary. Pp. 497-510, in Biodiversity and Ecosystem Function (E.D. Schulze and H.A. Mooney, Eds.), Springer-Verlag, Berlin.

[55] Sheth, A., and J. Larson. 1990. Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. ACM Computing Surveys 22: 183-236.

[56] Srivastava, J. 1999. Warehouse creation—a potential roadblock to data warehousing. IEEE Transactions on Knowledge and Data Engineering. Volume 11(1): 118-126.

[57] Varis, O. 1998. A belief network approach to optimization and parameter estimation: application to resource and environmental management. Artificial Intelligence, vol. 101:135-163.

[58] Waide, R.B., and M.R. Willig. Species diversity and productivity: An overview. Ecology (Submitted, 1998).

[59] Waide, R.B., M.R. Willig, C.F. Steiner, G.G. Mittelbach, L. Gough, S.I. Dodson, G.P. Juday, and R.R. Parmenter. 1999. The relationship between productivity and species richness. Annual Review of Ecology and Systematics (In Press).

[60] Wardle, D.A. 1999. Biodiversity, ecosystems and interactions that transcend the interface. TREE 14:125-127.

[61] Weibel, S. 1995. Metadata: the foundations of resource description. D-Lib Magazine, July 1995. Available at "http://www.dlib.org".

[62] Willig, M.R., and R.B. Waide. Species diversity and productivity: A synthesis. Ecology (Submitted, 1998).

O. J. Reichman
CURRICULUM VITAE

OFFICE ADDRESS:
National Center for Ecological Analysis and Synthesis
735 State Street, Suite 300
Santa Barbara, CA 93101-5504
(805) 892-2504; reichman@nceas.ucsb.edu

HOME ADDRESS:
5906 Via Lemora
Goleta, CA 93117
(805) 692-9454

EDUCATION:

1966	A.A., Broward Junior College, Ft. Lauderdale, FL
1968	B.A., Texas Tech University, Lubbock (zoology major, history minor)
1970	M.S., Texas Tech University
1974	Ph.D., Northern Arizona University, Flagstaff
1975	Post-doctoral appointment (with James H. Brown), University of Utah

RECENT PROFESSIONAL EMPLOYMENT:

1996-	Director, National Center for Ecological Analysis and Synthesis
1996-	Professor, University of California, Santa Barbara
1995-96	Assistant Director - Research, National Biological Service, Dept. of Interior (Senior Executive Service)
1993-95	Director, Konza Prairie Research Natural Area
1992	Professor, Kansas State University
1991-93	Associate Vice Provost for Research, Director of the Office of Research and Sponsored Programs (Interim), Kansas State University
1990-91	Director, Ecology Program, National Science Foundation

OTHER RECENT PROFESSIONAL ACTIVITIES:

1998 -	Editor, Ecological Applications
1997 -	President, American Society of Mammalogists
1997 -	Advisory Board, *Ecosystems*
1993 -	Editor, Special Features, *Journal of Mammalogy*
1992 -	Board of Trustees, BIOSIS Corporation (Vice Chair; Chair)
1988 -	Board of Directors, American Society of Mammalogists
1993-95	Council (Board of Directors), Ecological Society of America
1992-95	Board of Editors, Ecological Society of America (Ecology)
1991-92	Assistant Director - Kansas Experimental Program to Stimulate Competitive Research
1989 -	National Science Foundation Advisory Panels (Ecology, Dissertations, NCEAS, Field Stations and Marine Labs)
1989	Research Fellowship - Blandy Experimental Farm, University of Virginia

RECENT RESEARCH GRANTS:

1982-85	N.S.F. - Resource management strategies involving desert rodents, stored seeds, and domesticated molds (with D. Wicklow) - $45,000
1985-89	N.S.F. - An analysis of the impact of a fossorial herbivore, the plains pocket gopher, on plant survival, growth, and reproduction using simulated and natural burrow - $78,800 (Research Experience for Undergraduate Supplement for handicapped student - $3,800)
1986-87	National Geographic Society - Effects of competition and predation on the behavior and ecology of rodents on islands in the Sea of Cortez - $3,800

1992-95 N.S.F. - Collaborative research - Spatially explicit, animal-generated patterning of the environment and plant community structure (with D. Hartnett, N. Huntly, and R. Inouye) - $470,000 ($10,900 supplement for handicapped student)

1992-95 N.S.F. - Spatial distribution of microbes in desert rodent dens - $49,589 ($3,800 supplement for 2 minority high school students)

1993-95 N.S.F. - Development of Master Plan for Konza Prairie - $15,000

1994-97 The Nature Conservancy/Mellon Foundation - Relationship between landscape patterns, ecological processes, and biological diversity in native Great Plains grasslands (with several collaborators) - $285,000

1994-96 N.S.F. - Phase II development of Konza Prairie Research Natural Area - $192,000 ($507,000 including all matching support)

1996-01 N.S.F. - National Center for Ecological Analysis and Synthesis - $12,500,000 (replaced original P.I. after becoming Director)

1998-99 N.S.F. - An Analysis of Biotic and Physical Factors in El Niño-Induced Landslides - $29,854

1998-01 NSF - CRB: Dynamic interactions between plant distributions and animal-generated disturbance: Trajectories of restored communities. (with Eric Seabloom) - $191,000

FIVE PUBLICATIONS PERTINENT TO THIS PROPOSAL:

Reichman, O.J. 1991. Desert mammal communities, *In* Desert Communities, G. Polis, ed. University Arizona pg. 311-347

Reichman, O.J., J.H. Benedix, Jr., and T.R. Seastedt. 1993. Animal generated disturbances and size hierarchies in a tallgrass plant community. Ecology 74:1281-1285

Reichman, O.J., and H. Ronald Pulliam. 1996. Ecosystem management research at the National Biological Service. Ecological Applications 6:694-696

Reichman, O.J., and E. Roberts. 1994. Computer simulation analysis of heteromyid rodent foraging in relation to seed distributions: Implications for coexistence. Aust. J. Ecol. 42:467-477

Gendron, R.P., and O.J. Reichman. 1995. Food perishability and inventory management: A comparison of three caching strategies. American Naturalist 145:948-968

FIVE ADDITIONAL PUBLICATIONS:

Reichman, O.J. 1988. A comparison of the effects of crowding and pocket gopher disturbance on mortality, growth, and seed production of *Berteroa incana*. American Midland Naturalist 120:58-69

Reichman, O.J. 1988. Konza: A tallgrass natural history. University Press of Kansas, Lawrence, KS 226 pp.

Huntly, N., and O.J. Reichman. 1994. Effects of subterranean herbivores on vegetation. J. Mammalogy 74:852-859

Post, D., O.J. Reichman, and D. Wooster. 1992. Characteristics and significance of the caches of eastern woodrats (*Neotoma floridana*) J. Mammalogy 74:688-692

Reichman, O.J. 1995. Living Landscapes of Kansas. University Press of Kansas 154 pp.

JAMES W. BRUNT
Associate Director for Information Management
LTER Network Office
Research Assistant Professor
Department of Biology
University of New Mexico
Albuquerque, New Mexico 87131-1091

EDUCATION

M.S.	New Mexico State University	1988	Computational Ecology/Experimental Statistics
B.S.	New Mexico State University	1986	Botany/Chemistry
B.A.	New Mexico State University	1986	Biology

RELEVANT EXPERIENCE

November 1997 – Present: **University of New Mexico**. Title: **Research Assistant Professor**, Position**: Associate Director for Information Management**. Responsible for development of LTER Network information system and coordinating the information system development activities with LTER network partners.

February 1997-October 1997**: Photon Research Associates, Inc.** Title**: Staff Scientist,** Position**: Senior Systems Engineer.** Responsible for setting direction, policies, and guidelines for all aspects communication and information technology for a company of 100+ with a budget of 1,000.000 + / year. Including, collaborating with 5 divisions management, users, and system administrators to insure timely implementation and maintenance of specified enterprise and desktop solutions.

January 1989 – February 1997: **University of New Mexico**. Title: **Analyst/Programmer II**, Position: **Director of Sevilleta Information Management System -** Responsible for: the design, implementation and administration of the research information management computer system for the Sevilleta Long-Term Ecological Research (LTER) Program - supervising technicians, training project students, technicians, and investigators, and providing liaison with information management groups from the LTER Network and other national and international agencies.

1986 **-** 1988: **New Mexico State University - Computing Research Lab**. Title: GRA Position: **Science Workbench Project**. Worked with programmers and scientists in the development of research data management and analysis tools.

PEER-REVIEWED PUBLICATIONS

Brunt, James W. In Press. **The LTER Network Information System:A Framework For Ecological Information Management.** In: Proceedings of 1998 North American Symposium - Towards a Unified Framework for Forest Ecosystem Monitoring and Research. November 1-6, 1998 Guadalajara, Jalisco, Mexico.

Michener, William K., James W. Brunt, John Helly, Thomas B. Kirchner, and Susan G. Stafford. 1997. **Non-GeoSpatial Metadata for the Ecological Sciences**. Ecological Applications. 7(1):330-342.

Michener, William K., James W. Brunt, and William H. Jefferson. 1995. **New techniques for monitoring Crassostrea virginica (American Oyster) recruitment in the intertidal zone**. ICES Mar. Sci. Symp. 199:267-273.

Brunt, James W. 1994. **Research Data Management in Ecology: A Practical Approach for Long-term Projects**. Pages 272-275 in Proceedings of the Seventh International Working Conference on Scientific and Statistical Databases. IEEE Computer Society Press.

Su Wengui, Chang Yu, James W. Brunt, and Walt Conley. 1994. **Archival and Exchange Files and Their Applications in Ecology**. Advances in Resources and Eco-Environmental Research (Chinese). 5(3):37-39.

Su Wengui, James W. Brunt, Jiang Hong, and Chang Yu. 1994. **Data management and its significance in modern ecology**. Advances in Resource and Eco-Environmental Research (Chinese). 5(2):27-30.

Stafford, Susan G., James W. Brunt, and William K. Michener. 1994. **Integration of Ecological Research and Scientific Information Management**. Pages 3-19 in W. K. Michener, J. W. Brunt, and S. G. Stafford (Eds.), Environmental Information Management and Analysis: Ecosystem to Global Scales. Taylor and Francis, New York.

Conley, Walt, and James W. Brunt. 1991. **An Institute for Theoretical Ecology? - Part V: Practical Data Management for Cross-Site Analysis and Synthesis of Ecological Information**. Coenoses 6:173-180.

O'Neill, Robert V., Sandra J. Turner, Valerie I. Cullinan, Debra P. Coffin, Terry Cook, Walt Conley, James W. Brunt, John M. Thomas, Marsha R. Conley, James R. Gosz. 1991. **Multiple Landscape Scales: An Intersite Comparison**. Landscape Ecology 5:137-144.

Brunt, James W. and Walt Conley. 1990**. Behavior of a multivariate algorithm for ecological edge detection.** Ecological Modelling 49: 179-203.

Brunt, James W., Marsha R. Conley and Gary C. Cunningham. 1988. **Sex in Ephedra trifurca (Ephedraceae) with relation to Chihuahuan Desert Habitats**. American Midland Naturalist, 119:137-142.

BOOKS

W. K. Michener, J. W. Brunt, and S. G. Stafford (Eds.). 1994. **Environmental Information Management and Analysis: Ecosystem to Global Scales.** Taylor and Francis, London. 555 pages.

Michener, William K. and James W. Brunt. In Press. **Ecological Foundations**: **Managing and Transforming Data into Information and Knowledge.** Blackwell Scientific, Ltd., London.

JOHN J. HELLY
University of California, San Diego
San Diego Supercomputer Center
9500 Gilman Drive
La Jolla, CA 92093-0505
619-534-5060
619-534-5117 (fax)
hellyj@sdsc.edu

EDUCATION

1984 Ph.D. in Computer Science, UCLA
1978 M.S. in Biostatistics, UCLA
1975 M.A./B.A. in Biology, Occidental College, Los Angeles

PROFESSIONAL EXPERIENCE

1998-present	Specialist, San Diego Supercomputer Center, UCSD
1994-1998	Senior/Principal Scientist, San Diego Supercomputer Center, General Atomics
1991-1994	Director, Data Management and Scientific Computing, MEC Analytical Systems, Carlsbad, CA
1988-1991	Senior Scientist and Manager, Database Technology, Hughes Space and Communications Group, El Segundo, CA
1982-1988	Systems Director, Advanced Projects, Space Flight Operations, Shuttle and Ground
1978 -1981	Lecturer in Biochemistry, Department of Anesthesiology, UCLA Medical Center
1976-1981	Senior Statistician, Department of Anesthesiology, UCLA Medical Center
1977–1981	Statistical Consultant, Division of Thoracic Surgery, Department of Surgery, UCLA Medical Center
1976–1977	Statistician, Child Trauma Intervention Project, Neuropsychiatric Institute, UCLA Medical Center
1974-1975	Visiting Instructor in Biology, Occidental College, Los Angeles, CA

RELATED PUBLICATIONS

Helly, J., Elvins, T. T., Sutton, D. and Martinez, D. A Method for Interoperable Digital Libraries and Data Repositories (1999) Future Generation Computer Systems, Elsevier, 1999 (to appear)

Helly, J. , New Concept of Publication, (1998) Nature, 393, 107

Helly, J. J., Visualization of ecological and environmental data, in *Data and Information Management in Ecological Sciences: A Resource Guide*, eds. W. Michener, J. H. Porter, S. Stafford, Univ. of New Mexico Press, July 1998.

Michener, W., J. Brunt, J. Helly, T. Kirchener, and S. Stafford, Non-geospatial metadata for ecology, *Ecological Applications*, February 1997.

Helly, J., Levin, S., Michener, W., Davis, F. and Case, T. The State of Computational Ecology, (1996) University of California San Diego, San Diego Supercomputer Center, La Jolla, CA.

OTHER RELEVANT PUBLICATIONS

Helly, J., Elvins, T. T., Sutton, D, Martinez, D, Miller, S., Pickett, S., Davis, F. Controlled Publication of Digital Scientific Data (in preparation) TBD,

Helly, J.J., D. Stow, and A. Eshragi, An ultra-low level remote sensing system (ULLRSS) for environmental monitoring and mapping, *Proc. of Second Thematic Conference on Remote Sensing for Marine and Coastal Environments*, ERIM, pp. II-526, 1994.

Helly, J.J., and A. Carpenter, Visualization of coastal marine water-quality data using parallel coordinates, SAS User Group Conference, SUGI '93 Conference paper, 1993.

Helly, J. J. A A Representational Basis for the Development of a Distributed Expert System for Space Shuttle Flight Control (1984) , Vol. 58258 NASA Johnson Space Center

Helly, J.J. and K. Herbinson, Visualization of a salinity plume from a coastal marine desalination plant, *Water Environ. Res*., Jul/August, 66(5) pp. 753-758, 1994.

GRADUATE STUDENTS
None

POSTDOCTORAL RESEARCHERS
None

GRADUATE ADVISOR
Jaques Vidal, Computer Science, University of California, Los Angeles

Biographical Sketch

MATTHEW B. JONES

National Center for Ecological Analysis and Synthesis
735 State St., Suite 300
Santa Barbara, CA 93101

E-mail: jones@nceas.ucsb.edu
http://www.nceas.ucsb.edu/
Tele: (805) 892-2508

EDUCATION

1994 **M.S., UNIVERSITY OF FLORIDA**, Zoology Department, Gainesville, FL
Research: Co-evolutionary plant-animal interactions and community ecology, focusing on the role of secondary seed dispersal as it affects plant recruitment. Advisor: Dr. D. Levey

1989 **B.A., DARTMOUTH COLLEGE**, Hanover, NH

PROFESSIONAL EXPERIENCE

1996 - present **Database and Information Specialist (CNT III)**
National Center for Ecological Analysis and Synthesis, U. of California Santa Barbara

1994 - 1996 **Programmer Analyst II** Psychology, U. of California Santa Barbara

1995 - 1996 **Software Developer, Tierra project**

1994 **Fiscal Computer Consultant** Zoology, University of Florida

1992 **Academic Computer Consultant** Zoology, University of Florida

1990 **Research Technician** Zoology, Ohio State University

1989 **Field Crew Supervisor** Biological Sciences, Dartmouth College

PROFESSIONAL SOCIETIES

IEEE Computer Society
Ecological Society of America

PROFESSIONAL ACTIVITIES

Committe Member, Biological Data Working Group, Federal Geographic Data Committee (FGDC). Contributing to development of Biological Data Profile of the FGDC Content Standard for Digital Geospatial Metadata.

Advisory Board Member. Oz/Specify Biological Collections Management Software Project, University of Kansas.

Research interests: Data Management, Web-database integration, Evolutionary and genetic programming, Simulation Modeling, Individual-based modeling, Structured language applications (XML)

GRANTS AND FELLOWSHIPS

1997	NSF Grant to NCEAS on "Automation of Ecological Data Management", ($50K), Database Activities Program in the Biological Sciences
1993	Smithsonian Institution Graduate Student Fellowship
1992	Sigma Xi Grant-In-Aid of Research
1992	Smithsonian Tropical Research Institute Graduate Fellowship
1991-1993	NSF Predoctoral Fellowship
1991-1993	Grinter Fellowship, University of Florida
1989	Andrew Mellon Grant for Environmental Research

HONORS AND AWARDS

1989	First Prize, Sigma Xi undergraduate science competition
1989	Rufus Choate Scholar (Top 5% of class)
1989	Academic Citation in Environmental Engineering
1987	Academic Citation in Education
1986	H.L. Rhode III Merit Scholar

INVITED PRESENTATIONS

Jones, Matthew B., M. P. Schildhauer, and R. Nottrott. 1998. The Role of Informatics in Ecological Synthesis. 1998 Ecological Society of America Annual Meeting. Baltimore, Maryland.

Jones, Matthew B. 1997. Web-Database Integration. NSF Training Workshop for Ecological Data Management. Albuquerque, New Mexico.

Jones, Matthew B. 1997. Data Management at the National Center for Ecological Analysis and Synthesis. NSF Training Workshop for Ecological Data Management. Albuquerque, New Mexico.

Schildhauer, Mark P. and M. B. Jones. 1997. Software Choices for a High Performance Analytical Environment. 1997 Ecological Society of America Annual Meeting. Albuquerque, New Mexico.

PUBLICATIONS MOST CLOSELY RELATED TO THE PROPOSED PROJECT

Frondorf, A., M. B. Jones, and S. Stitt. 1999. Linking the FGDC geospatial metadata content standard to the biological/ecological sciences. Proceedings of the Third IEEE Computer Society Metadata Conference. Bethesda, MD. April 6-7, 1999.

Nottrott, R., M. B. Jones, and M. Schildhauer. 1999. Using XML-structured metadata to automate quality assurance processing for ecological data. Proceedings of the Third IEEE Computer Society Metadata Conference. Bethesda, MD. April 6-7, 1999.

Jones, M. B., R. Nottrott and M. P. Schildhauer. 1998. Metadata Editor to Support Ecological Data and Information Management. Prototype software under development at the National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA. Details available at http://www.nceas.ucsb.edu/ecoinformatics.

Jones, Matthew B. 1998. Web-based Data Management. *In* Data and Information Management in the Ecological Sciences: A Resource Guide. *Edited by* Michener, W.K., J.H. Porter and S.G. Stafford. LTER Network Office, University of New Mexico, Albuquerque, New Mexico.

OTHER PUBLICATIONS

Jones, M.B. 1994. Secondary seed removal by ants, beetles, and rodents in a neotropical moist forest. Master's Thesis, University of Florida, Gainesville, FL.

Peart, D.R., N.J. Poage and M.B. Jones. 1992. Winter injury to subalpine red spruce: influence of prior vigor and effects on subsequent growth. Canadian Journal of Forest Research. 22: 888-892.

Jones, M.B., C.L. Folt and S. Guarda. 1991. Characterizing individual, population and community effects of sublethal levels of aquatic toxicants: an experimental case study using Daphnia. Freshwater Biology 26: 35-44.

Peart, D.R., M.B. Jones and P.A. Palmiotto. 1991. Winter injury to red spruce at Mt. Moosilauke, N.H. Canadian Journal of Forest Research 21: 1380-1389.

MICHAEL R. WILLIG

Research Focus:

My research program is broad, quantitative, and focuses on 6 themes: (1) macrogeographic patterns with regard to diversity, species range size, latitude, and productivity, (2) disturbance ecology, (3) functional diversity of microbial communities, (4) community structure, (5) landscape ecology, and (6) tropical ecology and conservation.

Current Title and Affiliation:

Professor, Program in Ecology and Conservation Biology, Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409-3131
Sabbatical Fellow, National Center for Ecological Analysis and Synthesis, University of California at Santa Barbara, 735 State Street, Santa Barbara, CA 93101-5504

Education:

B.S. Biology, University of Pittsburgh, 1974
Ph.D. Biology, University of Pittsburgh, 1982

Professional Experience:

1993-present	Professor of Biological Science, Texas Tech University
1995-1997	Chairperson, Department of Biological Sciences, Texas Tech University
1994-1996	Director, The Institute for Environmental Sciences, Texas Tech University
1989-1993	Associate Professor of Biology, Texas Tech University
1983-1989	Assistant Professor of Biology, Texas Tech University
1981-1983	Assistant Professor of Biology, Loyola University

Graduate and Postgraduate Advisor:

Doctoral Students—7 in last 5 years (9 total)
Masters Students—12 in last 5 years (20 total)
Post-doctoral Fellows—1 in last 5 years (3 total)

Professional Organizations (Selected):

American Association for the Advancement of Science; Ecological Society of America; The Association for Tropical Biology; American Society of Naturalists; American Society of Mammalogists; Animal Behavior Society; Society for the Study of Evolution; The Paleontological Society; The Mexican Society of Mammalogists; The Argentine Society of Mammalogists; Association of Systematic Zoologists; Entomological Society of America; The Society for Conservation Biology; The North American Benthological Society; Society for Environmental Toxicology and Chemistry

Research Grants (Total > 8.0 million; last 5 years itemized):

1993-99	U.S. Department of Agriculture, Forest Service, **$79,332.**
1994-99	N.S.F. Long-Term Ecological Projects (with R.D. Owen). **$265,000.**
1994-00	National Science Foundation LTER (with Co-PDs [R. Waide, A. Lugo, F. Scatena, and J. Zimmerman] & other Co-PIs). **$4,080,000.**
1995	U.S. Department of Defense (via USA-CERL). **$29,779.**
1996	National Center for Ecological Analysis and Synthesis (with R. Waide).
1982-96	Dept. of Energy, Oak Ridge Associated Universities. **$45,000.**
1997	N.S.F. Research Enhancement for Undergraduates (with R.D. Owen). **$11,280.**
1997	N.S.F. Long-Term Ecological Projects Supplement (with R.D. Owen). **$20,826.**
1997-99	U.S.D.A., Forest Service (with S. Cox). **$34,368.**
1998	National Center for Ecological Analysis and Synthesis. **$99,100.**

Publications (> 100 published or in press):

Five Most Relevant to Proposed Activity

Lyons, S.K., & **M.R. Willig**. 1997. Latitudinal patterns of range size: Methodological concerns and empirical evaluations for New World bats and marsupials. **Oikos** 79:568-580.

Willig, M.R., & S.K. Lyons. 1998. An analytical model of latitudinal gradients of species richness with an empirical test for marsupials and bats in the New World. **Oikos**, 81:93-98.

Kaufman, D.M., & **M.R. Willig.** 1998**.** Latitudinal patterns of mammalian species richness in the New World: the effects of sampling method and faunal group. **J. Biogeog.** 25:236-246.

Waide, R.B., **M.R. Willig**, G. Mittelbach, C. Steiner, L. Gough, S.I. Dodson, G.P. Juday, & R. Parmenter. 1999. The relationship between productivity and species richness. **Ann. Rev. Ecol. & Syst.** (In Press).

Lyons, K.A., & **M.R. Willig**. A hemispheric assessment of scale-dependence in latitudinal gradients of species richness. **Ecology** (In Press).

Other Significant Publications of Interest

Zak, J.C., **M.R. Willig**, D.L. Moorhead, & H.G. Wildman. 1994. Functional diversity of microbial communities: A quantitative approach. **Soil Biol. & Biochem.** 26:1101-1108.

Camilo, G.R., & **M.R. Willig**. 1995. Dynamics of a food chain model from an arthropod-dominated lotic community. **Ecol. Model.** 79:121-129.

Willig, M.R., D.L. Moorhead, S.B. Cox, & J.C. Zak. 1996. Functional diversity of soil bacterial communities in the tabonuco forest: Interaction of anthropogenic and natural disturbance. **Biotropica** 28:471-483.

Willig, M.R. & L.R. Walker. 1999. Disturbance in terrestrial ecosystems: salient themes, synthesis, and future directions. Pp. 000-000, in Ecology of Disturbed Ground, Elsevier Science, Amsterdam, The Netherlands (In Press).

Stevens, R.D., & **M.R. Willig**. 1999. Community structure, abundance, and morphology. **Oikos** (In Press).

Waide, R.B., & **M.R. Willig.** 1999. Species diversity and productivity: an overview. **Ecology** (Submitted).

Biographical Sketch: Sandy J. Andelman

National Center for Ecological Analysis & Synthesis
University of California
735 State Street, Suite 300
Santa Barbara, California 93101
Tel 805-892-2505; Fax 805-892-2510 Email andelman@nceas.ucsb.edu

PROFESSIONAL POSITIONS

1/99 – present: Deputy Director, National Center for Ecological Analysis & Synthesis, University of California, Santa Barbara and Adjunct Assistant Professor, Bren School of Environmental Science & Management, University of California, Santa Barbara

7/98 – 12/98 Research Associate, National Center for Ecological Analysis & Synthesis, University of California, Santa Barbara

9/97 - 6/98 Visiting Assistant Professor, Bren School of Environmental Science & Management, University of California, Santa Barbara

9/94 – 9/97 Director of Conservation Science, The Nature Conservancy of Washington

9/90 – 9/994 Research Assistant Professor, Institute for Environmental Studies, University of Washington

EDUCATION

Ph.D., University of Washington, 1985; B.A., Lewis and Clark College, 1977.

FIVE RELEVANT PUBLICATIONS

Andelman, S.J. and E. Meir. 1999. Breadth is better than depth: Biodiversity data requirements for adequate reserve networks. In press, *Conservation Biology*.

Davis, F.W., D.M. Stoms, and S.J. Andelman. 1999. Systematic reserve selection in the USA: An example from the Columbia Plateau ecoregion. In press, *Parks.*

Andelman, S.J. and W. Fagan. The quest for umbrella species: efficient conservation surrogates or expensive mistakes. Ms., in revision, *Conservation Biology*.

Possingham, H. and S. Andelman. 1999. An equity formula for fauna persistence: what happens when conservation biologists have to provide prescriptions for real problems – quickly? Ms., in review, *Conservation Biology*.

Kareiva, P.M., Andelman, S.J., Doak, D.F., Eldard, B., Groom, M., Hoekstra, J., Hood, L., James, F., Lamoreux, J., McCullough, C., Regetz, J., Savage, L., Ruckelshaus, M., Skelly, D., Wilbur, H. and K. Zamudio. 1999. Using science in habitat conservation plans. NCEAS and AIBS. www.nceas.ucsb.edu

RECENT GRANTS AND AWARDS (PAST FIVE YEARS)

1998-1999. Analytical toolbox and decision support framework for selecting and prioritizing conservation sites at a regional scale. ($50,000) (PI, with Frank Davis).

1997-1999. Designing and assessing the viability of nature reserve systems at regional scales: integration of optimization, heuristic and dynamic models. National Center for Ecological Analysis and Synthesis Working Group. PI.

1998. Models for siting nature reserves. The Nature Conservancy ($10,000) PI.

1998. National Science Foundation grant on behalf of the Society for Conservation Biology to fund graduate and postdoctoral student travel stipends ($15,000).

1997-1998. U.S. Fish and Wildlife Service, Ecosystem Restoration Program. Restoration and monitoring prairie ecosystems and endangered species in w. Washington ($50,000) PI.

1997- 1998. U.S. Army, Fort Lewis. The application of matrix projection models to assess the effects vehicular disturbance on the rare plant *Aster curtus* ($25,000) PI.

1996-1997. The Foster Foundation, Pacific Gas & Transmission, and The Nature Conservancy. A pilot effort in ecoregional-scale conservation planning for biodiversity in the Columbia Plateau ($100,000) PI.
1996-1997. Conservation Technology Support Program. Conservation applications of geographic information systems ($89,000) PI
1991-1993. Center for Wildlife Conservation and Graduate School, University of Washington, ($44,000) PI.
1993. Bureau of Land Management. Research, conservation and management priorities for neotropical migratory songbirds in Oregon and Washington ($ 7,000) PI.
1992. U.S. Fish and Wildlife Service, Cooperative Agreement ($125,000) PI.
1991. U.S. Fish and Wildlife Service, Cooperative Agreement ($250,000) PI.
1991-1992. Institute of Museum Services, Conservation Grant ($17,500) PI.
1991. Odessa Economic Development Association ($10,000) PI
1990-1991. Washington Department of Wildlife ($22,700) PI.
1986-1989. Harry Frank Guggenheim Foundation Research Grant ($69,000) PI.

PROFESSIONAL SERVICE
Elected Officer, Board of Governors (Secretary), Society for Conservation Biology, 1996-present.
Review Panel, Conservation Technology Support Program, 1998,1999.
Review Panel, The Nature Conservancy/Mellon Foundation Ecosystem Research Program, 1997.
Steering Committee, Partners in Flight, Oregon-Washington, 1992-1997.

RESEARCH INTERESTS
Developing a theoretical foundation for nature reserve network design; sensitivity of reserve siting models to biases in biodiversity data; models for assessing ecological value; theory for optimal management of ecosystem services; use of stochastic population models for risk assessment and management of threatened populations; computer-aided decision support tools for environmental decision making in the context of uncertainty.

PROFESSIONAL SOCIETIES
Society for Conservation Biology, Ecological Society of America, American Institute of Biological Sciences, Natural Areas Association

PETER W. ARZBERGER
Deputy Director
University of California, San Diego
San Diego Supercomputer Center, 0505
9500 Gilman Drive
La Jolla, CA 92093-0505
619-534-5079
619-534-5056 (fax)
parzberg@ucsd.edu
http://www.sdsc.edu/~parzberg/parzberg.html

EDUCATION

1983 Ph.D. in Mathematics, Purdue University
1979 M.S. in Mathematical Statistics, Purdue University
1974 B.A. in Mathematics, University of Massachusetts

PROFESSIONAL EXPERIENCE

1996-present Executive Director, NPACI; Deputy Director, San Diego Supercomputer Center
1995-1996 Executive Director, San Diego Supercomputer Center
1993-1995 Deputy HPCC Coordinator, National Science Foundation
1988-1995 Program Officer, National Science Foundation
1985-1988 Visiting Assistant Professor, Department of Statistics, University of Wisconsin
1983-1985 Assistant Professor, Department of Mathematics, Rochester Institute of Technology

COMMITTEES

Member, Resource Advisory Committee for NIH Resource for Concurrent Biological Computing, 1995-present

RELATED PUBLICATIONS

Arzberger, P. (one of 15 co-authors), Position paper on molecular evolution, *Molecular Phylogenetics and Evolution*, vol. 1, no. 1, pp. 84-85, 1992.

Arzberger, P., J. Sunley, N. Sedransk, and K. Crank, Readers' comments to the new researchers' committee report: Comments from the National Science Foundation, *Statistical Sciences*, vol. 7, no. 2, pp. 256-259, 1992.

Arzberger, P., *Mathematics and Biology: The Interface, Challenges and Opportunities*, Simon Levin, editor, (NSF liaison to the committee writing the report) 1992.

Arzberger, P., Results for generalized inbreeding systems, *J. Mathematical Biology*, 26:535-550, 1988.

Arzberger, P., A probabilistic and algebraic treatment of regular inbreeding systems, *J. Mathematical Biology*, 22:175-197, 1985.

GRADUATE AND POSTGRADUATE ADVISOR:

Stanley Sawyer, Department of Mathematics, Washington University

ARCOT K. RAJASEKAR

San Diego Supercomputer Center, University of California at San Diego,
9500 Gilman Drive, La Jolla CA 92093-0505 USA
619-534-8378 FAX: 619-534-5077 sekar@sdsc.edu

Education:

- Ph.D., Computer Science, University of Maryland, College Park (1989) Thesis Advisor, Dr. Jack Minker
- M.S., Computer Science, Indian Institute of Technology, Madras, India (1983) Thesis Advisor: Dr. R. Kalyanakrishnan
- B.E. (Honors), Electronics and Communications, University of Madras, Madras, India (1979)

Employment Experience:

- 1998- Principal Scientist, Enabling Technologies, SDSC. Working on projects in digital libraries, data repositories, massive data analysis systems and archival systems.
- 1996-1998 Staff Scientist, Enabling Technologies, SDSC. Working on projects in digital libraries, data repositories, massive data analysis systems and archival systems.
- 1990-1996 Assistant Professor, University of Kentucky at Lexington. Teaching and research work in artificial intelligence and databases.
- 1989-1990 Research Associate, Post-Doc, University of Maryland Institute for Advanced Computer Studies. Research work in logic programming and non-monotonic reasoning.

Areas of Interest:

- Areas of interest include research and development technologies for building digital library systems, archival systems, data mining and database-oriented reasoning, and logic programming and default reasoning.
- Research activities at SDSC include development of a Storage Resource Broker for integrating distributed data repositories, cloning of digital library systems and development of meta data catalog system for handling system-level and domain-specific meta data

Professional Contributions and Honors:

- Annals of Mathematics and Artificial Intelligence, guest editor, *Logic in Databases, Knowledge Representation and Reasoning*, Volume 14, Nos. 2,3 & 4, September 1995.
- Third International Conference on Logic Programming and Nonmonotonic Reasoning, Program Chair, Poster Session, Lexington, June 1995.
- Annals of Mathematics and Artificial Intelligence, guest editor, *Disjunctive Logic Programming*, Volume 12, Nos. 1 & 2, December 1994.
- Workshop on Logic Programming with Incomplete Information, co-organizer, Vancouver, October, 1993.
- Joint International Conference and Symposium on Logic Programming, session chair, Washington, D.C., October, 1992.
- Workshop on Disjunctive Logic Programs, co-organizer, San Diego, November, 1991.

Support received from the following funded projects:

- NSF - National Partnership for Advanced Computational Infrastructure
- NSF - Molecular Structures Database
- NARA - Email Archives

Biographical Sketch

MARK P. SCHILDHAUER

National Center for Ecological Analysis and Synthesis
735 State St., Suite 300
Santa Barbara, CA 93101

E-mail: schild@nceas.ucsb.edu
http://www.nceas.ucsb.edu/
Tele: (805) 892-2509

EDUCATION

1991 **Ph. D.**, Department of Biological Sciences, University of California, Santa Barbara. Santa Barbara CA
Research: Dynamics of territorial acquisition in the polygynous reef fish, Thalassoma bifasciatum. Advisor: Dr. R. R. Warner

1976 **B.A., HARVARD COLLEGE**, CAMBRIDGE, MASS.

PROFESSIONAL EXPERIENCE

1995 - present **Director of Computing**
National Center for Ecological Analysis and Synthesis, U. of California Santa Barbara

1993 - 1995 **Computer Resource Manager and Technical Coordinator,**
Division of Social Sciences, U. of California Santa Barbara

1990 – 1993 **Progammer Analyst II**
Social Sciences Computing Facility, U. of California Santa Barbara

1986-1989 **Ecological data analyst and laboratory computer coordinator**
Marine Review Committee and Marine Sciences Institute, U. of California Santa Barbara

PROFESSIONAL SOCIETIES

IEEE Computer Society
Ecological Society of America

PROFESSIONAL ACTIVITIES

Data Advisory Panel Member, Multi-Agency Rocky Intertidal Network group, coordinated by the Minerals Management Service of the U.S. Dept. of the Interior.

Technical Advisor, National Index Site Committee

Research interests: Scientific computing, Management of networked computing systems, Data visualization, Computer-supported collaborative work

GRANTS AND FELLOWSHIPS

1997 NSF Grant to NCEAS on “Automation of Ecological Data Management using Structured Metadata” ($50K), Database Activities Program in the Biological Sciences

HONORS AND AWARDS

1989 Regents’ Fellowship, UCSB
1972 National Merit, Merit Scholarship, Harvard College

INVITED PRESENTATIONS

Jones, Matthew B., M. P. Schildhauer, and R. Nottrott. 1998. The Role of Informatics in Ecological Synthesis. Ecological Society of America Annual Meeting. Baltimore, Maryland.

Schildhauer, Mark P. 1997. Virtual Working Groups at NCEAS: Using the Web to Facilitate Scientific Collaboration. NSF Training Workshop for Ecological Data Management. Albuquerque, New Mexico.

Schildhauer, Mark P. and M. B. Jones. 1997. Software Choices for a High Performance Analytical Environment. Ecological Society of America Annual Meeting. Albuquerque, New Mexico.

Schildhauer, Mark P. 1993. Training people to use the SAS system under UNIX operating systems. Presentation at the First Regional Conference, Western Users of SAS Software

PUBLICATIONS MOST CLOSELY RELATED TO THE PROPOSED PROJECT

Nottrott, R., M. B. Jones, and M. Schildhauer. 1999. Using XML-structured metadata to automate quality assurance processing for ecological data. Proceedings of the Third IEEE Computer Society Metadata Conference. Bethesda, MD. April 6-7, 1999.

Jones, M. B., R. Nottrott and M. P. Schildhauer. 1998. Metadata Editor to Support Ecological Data and Information Management. Prototype software under development at the National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA. Details available at http://www.nceas.ucsb.edu/ecoinformatics.

Schildhauer, Mark P. 1998. Virtual Working Groups at NCEAS: Using the Web to Facilitate Scientific Collaboration. *In* Data and Information Management in the Ecological Sciences: A Resource Guide. *Edited by* Michener, W.K., J.H. Porter and S.G. Stafford. LTER Network Office, University of New Mexico, Albuquerque, New Mexico.

Schildhauer, Mark P. 1993. Training people to use the SAS system under UNIX operating systems. *In* Western Users of SAS Software, Proceedings of the First Regional Conference.

OTHER PUBLICATIONS

Schildhauer, M. P. 1991. Dynamics of territorial acquisition in a polygynous reef fish, Thalassoma bifasciatum. Ph. D. dissertation, University of California, Santa Barbara..

Bence, J. R., and M. P. Schildhauer. 1989. Technical Report to the California Coastal Commission. (E). Metals and Radiation. Marine Review Committee, Inc.

Hoffman, S. G., M. P. Schildhauer, and R. R. Warner. 1985. The costs of changing sex and the ontogeny of males under contest competition for mates. Evolution 39: 915-927.

DON W. SUTTON
University of California, San Diego
San Diego Supercomputer Center
9500 Gilman Drive
La Jolla, CA 92093-0505
619-534-5085
619-534-5152 (fax)
suttond@sdsc.edu

EDUCATION

1987 Ph.D. in Applied Mechanics and Engineering Sciences (Bioengineering), UC San Diego
1983 M.S. in Applied Mechanics and Engineering Sciences (Bioengineering), UC San Diego
1981 B.A. in Biophysics (College of Letters and Science), UC Berkeley

PROFESSIONAL EXPERIENCE

1996-present	Research Scientist, San Diego Supercomputer Center
1992-1996	Adjunct Professor, Physics Department, San Diego State University
1991-1992	Lecturer, Mathematics Department, San Diego Community College
1989-1991	Postdoctoral Fellow, Marine Physical Lab, Scripps Institution of Oceanography
1987-1989	Postdoctoral Fellow, Applied Mechanics and Engineering Science, UC San Diego

AFFILIATIONS, GRANTS, AND AWARDS

Member, American Geophysical Union, American Society of Civil Engineers, Society of Environmental Toxicology and Chemistry
Co-investigator, DERA Project Grant (US Navy) "San Diego Risk Assessment/Technology Demonstration," 1993-1995
Co-investigator, SERDP Applied Research Grant (US Navy) "Contaminant Dispersal Model for San Diego Bay," 1993-1995
National Research Council Training Award, 1992-1993
National Research Service Award, 1988-1989
NIH Predoctoral Traineeship, 1983-1987

RELATED PUBLICATIONS

Cheng, R.T., J.W. Gartner, E.S. Gross, K. Richter, D.W. Sutton, P.F. Wang, Modeling tidal hydrodynamics of San Diego Bay, California, *Jour. of Am.Water Resources Association*, 1997 (submitted).

Sutton, D.W., B. Chadwick, and K. Richter, Mass loading and subsequent baywide transport of sediment resuspended during tug assisted ship movements at the Naval Station San Diego, California and the World Ocean '97, *Am. Soc. of Civil Engineers Conference Proceedings*, 1997.

Sutton, D.W., K. Richter, Resuspension of sediments during ship docking and berthing at Naval Station San Diego, CA, *Am. Geophys. Union Fall Conference Proceedings*, 1996.

Sutton, D.W., B. Chadwick, K. Richter, Computer model predictions of baywide transport and fate of creek storm inputs to San Diego Bay, Navy Technical document, 1996.

Sutton, D.W., K. Richter, Computer model study on changes in water currents and sediment transport rates due to proposed dredging of the shipping channel, Navy Technical document, 1995.

Sutton, D.W., B. Chadwick, K. Richter, Computer model simulations of fuel spills at Ballast Point, San Diego Bay, CA, *Conference Proceedings Oceans '95*, vol. 3, pp. 1,714-1,721, 1995.

Richter, K., D.W. Sutton, L. Reidy, R. Cheng, Comparison of water velocity field data and model predictions in San Diego Bay, *Conference Proccedings Oceans '95*, vol. 3, pp. 1,751-1,755, 1995.

GRADUATE STUDENTS ADVISED
None

POSTDOCTORAL RESEARCHERS ADVISED
None

GRADUATE AND POSTGRADUATE ADVISORS
Jules Jaffe, UC San Diego/SIO
Geert Schmid-Schonbein, UC San Diego

ROBERT B. WAIDE

CURRENT POSITION

Executive Director, Long-Term Ecological Research Network Office
801 University Blvd., Suite 104
Albuquerque, NM 87106
SS: 328-40-7640
Telephone: 505/272-7311
Fax: 505/277-7080
Electronic mail: rwaide@lternet.edu

EDUCATION

1978 Ph.D. in Zoology, University of Wisconsin – Madison
Dissertation: Interactions between tropical resident and north temperate migrant birds in southern Campeche, Mexico.

1973 M.S. in Zoology, University of Wisconsin – Madison
Thesis: Seasonal changes in the avifauna of a tropical wet forest.

PREVIOUS EXPERIENCE

Director of the Institute for Tropical Ecosystem Studies (University of Puerto Rico-Rio Piedras). 1982-97.

Principal Investigator for the Luquillo Experimental Forest Long-Term Ecological Research site. 1987-97.

Director of the Puerto Rico Minority Research Center of Excellence. 1995-97.

Direction of a five-member research team studying mineral and energy cycling and transport in a subtropical wet forest, sponsored by the U. S. Department of Energy. 1982-88.

RECENT RESEARCH PUBLICATIONS

Waide, R. B. M.R. Willig, C. F. Steiner, G. Mittelbach, L. Gough, S. I. Dodson, G.P. Juday, and R. Parmenter. 1999. The relationship between productivity and species richness. Annual Review of Ecology and Systematics (in press).

Waide, R.B., J.K. Zimmerman, and F.N. Scatena. 1998. Controls of primary productivity in a montane tropical forest: Lessons from the Luquillo Mountains in Puerto Rico. Ecology 79:31-37.

D. P. Reagan and R. B. Waide, eds. 1996. The Food Web of a Tropical Rain Forest, University of Chicago Press.

Zou, X., C. Zucca, R. B. Waide, and W. H. McDowell. 1995. Long-term influence of deforestation on tree species composition and litter dynamics of a tropical rain forest in Puerto Rico. Forest Ecology and Management 78:147-157.

Taylor, C.M., S. Silander, R. B. Waide, and W. J. Pfeiffer. 1995. Recovery of a Tropical Forest after Gamma Irradiation: A 23-Year Chronicle. In Tropical Forests: Management and Ecology. A. E. Lugo and C. Lowe, eds. Ecological Studies, Volume 112. Springer Verlag.