

**INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.B. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

PI/PD Name: Thomas Moritz

Gender: Male Female
Ethnicity: (Choose one response) Hispanic or Latino Not Hispanic or Latino

Race:
(Select one or more)
 American Indian or Alaska Native
 Asian
 Black or African American
 Native Hawaiian or Other Pacific Islander
 White

Disability Status:
(Select one or more)
 Hearing Impairment
 Visual Impairment
 Mobility/Orthopedic Impairment
 Other
 None

Citizenship: (Choose one) U.S. Citizen Permanent Resident Other non-U.S. Citizen

Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):

REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project

Ethnicity Definition:

Hispanic or Latino. A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

Race Definitions:

American Indian or Alaska Native. A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

Asian. A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Black or African American. A person having origins in any of the black racial groups of Africa.

Native Hawaiian or Other Pacific Islander. A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

White. A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

WHY THIS INFORMATION IS BEING REQUESTED:

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

**INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.B. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

PI/PD Name: Norman F Johnson

Gender: Male Female
Ethnicity: (Choose one response) Hispanic or Latino Not Hispanic or Latino

Race:
(Select one or more)
 American Indian or Alaska Native
 Asian
 Black or African American
 Native Hawaiian or Other Pacific Islander
 White

Disability Status:
(Select one or more)
 Hearing Impairment
 Visual Impairment
 Mobility/Orthopedic Impairment
 Other
 None

Citizenship: (Choose one) U.S. Citizen Permanent Resident Other non-U.S. Citizen

Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):

REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project

Ethnicity Definition:

Hispanic or Latino. A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

Race Definitions:

American Indian or Alaska Native. A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

Asian. A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Black or African American. A person having origins in any of the black racial groups of Africa.

Native Hawaiian or Other Pacific Islander. A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

White. A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

WHY THIS INFORMATION IS BEING REQUESTED:

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

List of Suggested Reviewers or Reviewers Not To Include (optional)

SUGGESTED REVIEWERS:

P. Bryan Heidorn
GSLIS , UIUC

James Quinn
ICE, UC, Davis

Dr. David Grossman
Illinois Institute of Technology
U.S.A. <grossman@iit.edu>

Vipul Kashyap, Ph.D.
NLM, U.S.A.
<kashyap@nlm.nih.gov>

Prof. Dr. Joachim W. Schmidt
Technical University of Hamburg-Harburg
Germany
<j.w.schmidt@tu-harburg.de>

Dr. Arjen de Vries
CWI, Netherlands
<arjen@acm.org>

Prof. Dr. Gerhard Weikum
University of the Saarland
Department of Computer Science
Germany <weikum@cs.uni-sb.de>

REVIEWERS NOT TO INCLUDE:

Luis Gravano, Columbia University

List of Suggested Reviewers or Reviewers Not To Include (optional)

SUGGESTED REVIEWERS:

Not Listed

REVIEWERS NOT TO INCLUDE:

Not Listed

COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE <i>if not in response to a program announcement/solicitation enter NSF 02-2</i>					FOR NSF USE ONLY	
NSF 02-085			04/15/03		NSF PROPOSAL NUMBER	
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) <small>(Indicate the most specific unit known, i.e. program, division, etc.)</small>						
IIS - SPECIAL PROJECTS (IIS)						
DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# <small>(Data Universal Numbering System)</small>	FILE LOCATION	
				061202768		
EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)		SHOW PREVIOUS AWARD NO. IF THIS IS <input type="checkbox"/> A RENEWAL <input type="checkbox"/> AN ACCOMPLISHMENT-BASED RENEWAL		IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES <input type="checkbox"/> NO <input checked="" type="checkbox"/> IF YES, LIST ACRONYM(S)		
136162659						
NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE			ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE			
American Museum Natural History			American Museum Natural History			
AWARDEE ORGANIZATION CODE (IF KNOWN)			Central Park West at 79th Stre			
4000790000			New York, NY. 100245192			
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) <small>(See GPG II.C For Definitions)</small> <input type="checkbox"/> FOR-PROFIT ORGANIZATION <input type="checkbox"/> SMALL BUSINESS <input type="checkbox"/> MINORITY BUSINESS <input type="checkbox"/> WOMAN-OWNED BUSINESS						
TITLE OF PROPOSED PROJECT Collaborative Research: Development of new digital library applications in the context of a basic ontology for biosystematics information using the literature of entomology (ants)						
REQUESTED AMOUNT \$	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PREPROPOSAL NO., IF APPLICABLE			
756,276	36 months	05/01/03				
CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW						
<input type="checkbox"/> BEGINNING INVESTIGATOR (GPG I.A)			<input type="checkbox"/> HUMAN SUBJECTS (GPG II.C.11)			
<input type="checkbox"/> DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C)			Exemption Subsection _____ or IRB App. Date _____			
<input type="checkbox"/> PROPRIETARY & PRIVILEGED INFORMATION (GPG I.B, II.C.6)			<input type="checkbox"/> INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.9)			
<input type="checkbox"/> HISTORIC PLACES (GPG II.C.9)						
<input type="checkbox"/> SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.C.11)						
<input type="checkbox"/> VERTEBRATE ANIMALS (GPG II.C.11) IACUC App. Date _____			<input type="checkbox"/> HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.E.1)			
PI/PD DEPARTMENT		PI/PD POSTAL ADDRESS				
Library		American Museum of Natural History				
PI/PD FAX NUMBER		Central Park West @ 79th Street				
212-769-5009		New York, NY 10024				
		United States				
NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address		
PI/PD NAME						
Thomas Moritz	MLS	1974	212-769-5417	tmoritz@amnh.org		
CO-PI/PD						
CO-PI/PD						
CO-PI/PD						
CO-PI/PD						

CERTIFICATION PAGE

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 debarment and suspension, drug-free workplace, and lobbying activities (see below), as set forth in Grant Proposal Guide (GPG), NSF 02-2. Willful provision of false information in this application and its supporting documents or in reports required under an ensuing 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. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Appendix A of the Grant Proposal Guide.

Debarment and Suspension Certification

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

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

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Appendix B of the Grant Proposal Guide.

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 of Lobbying Activities," 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			
TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER	

*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.

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 02-2				FOR NSF USE ONLY	
NSF 02-085		04/15/03		NSF PROPOSAL NUMBER	
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.)					
IIS - SPECIAL PROJECTS (IIS)					
DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
				071650709	
EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)		SHOW PREVIOUS AWARD NO. IF THIS IS <input type="checkbox"/> A RENEWAL <input type="checkbox"/> AN ACCOMPLISHMENT-BASED RENEWAL		IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES <input type="checkbox"/> NO <input checked="" type="checkbox"/> IF YES, LIST ACRONYM(S)	
316401599					
NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE			ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE		
Ohio State University Research Foundation			Ohio State University Research Foundation		
AWARDEE ORGANIZATION CODE (IF KNOWN)			ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE		
0030908001			Ohio State University		
NAME OF PERFORMING ORGANIZATION, IF DIFFERENT FROM ABOVE			ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE		
Ohio State University			Ohio State University		
PERFORMING ORGANIZATION CODE (IF KNOWN)			COLUMBUS, OH 43210-1016		
0030908000			Columbus, OH 43210-1016		
IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions) <input type="checkbox"/> FOR-PROFIT ORGANIZATION <input type="checkbox"/> SMALL BUSINESS <input type="checkbox"/> MINORITY BUSINESS <input type="checkbox"/> WOMAN-OWNED BUSINESS					
TITLE OF PROPOSED PROJECT Collaborative Research: Development of new digital library applications in the context of a basic ontology for biosystematics information using the literature of entomology (ants)					
REQUESTED AMOUNT \$	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PREPROPOSAL NO., IF APPLICABLE		
50,577	36 months	05/01/03			
CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW					
<input type="checkbox"/> BEGINNING INVESTIGATOR (GPG I.A)		<input type="checkbox"/> HUMAN SUBJECTS (GPG II.C.11) Exemption Subsection _____ or IRB App. Date _____			
<input type="checkbox"/> DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C)		<input type="checkbox"/> INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.9) _____			
<input type="checkbox"/> PROPRIETARY & PRIVILEGED INFORMATION (GPG I.B, II.C.6)		<input type="checkbox"/> HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.E.1)			
<input type="checkbox"/> HISTORIC PLACES (GPG II.C.9)					
<input type="checkbox"/> SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.C.11)					
<input type="checkbox"/> VERTEBRATE ANIMALS (GPG II.C.11) IACUC App. Date _____					
PI/PD DEPARTMENT Department of Entomology		PI/PD POSTAL ADDRESS The Ohio State University			
PI/PD FAX NUMBER 614-292-7774		1315 Kinnear Road			
		Columbus, OH 432121192			
		United States			
NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address	
PI/PD NAME Norman F Johnson	Ph.D.	1981	614-292-6595	Johnson.2@osu.edu	
CO-PI/PD					
CO-PI/PD					
CO-PI/PD					
CO-PI/PD					

CERTIFICATION PAGE

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 debarment and suspension, drug-free workplace, and lobbying activities (see below), as set forth in Grant Proposal Guide (GPG), NSF 02-2. Willful provision of false information in this application and its supporting documents or in reports required under an ensuing 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. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Appendix A of the Grant Proposal Guide.

Debarment and Suspension Certification

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

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

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Appendix B of the Grant Proposal Guide.

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 of Lobbying Activities," 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			
TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER	

*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.

Summary

We are proposing to design and test approaches to mark-up and extraction of scientific data from a corpus of texts drawn from the biosystematics literature of entomology (ants) and to develop a set of applications informed by a basic ontology for biosystematics information, including powerful search and retrieval strategies to operate on this corpus.

Biosystematics is the science that provides the definitional foundations for organismic biology and for the applied science of biodiversity conservation. Within zoological systematics, insects are a uniquely important group of organisms with major impacts on human health and economics; ants constitute a particularly important group of social insects accounting for a major part of the biomass in tropical rainforests and displaying remarkable diversity and behavioral variation.

Building on digital library work already well underway at AMNH, a strong history of biological informatics at OSU and the demonstrated strengths of computer science collaborator at Otto-von-Guericke-Universität Magdeburg, we propose a basic ontology of natural history information defined by time, geography, taxonomy and selected descriptors and supported by a developing array of Web-based services.

We will test approaches to automated XML mark-up using our hypothesized “*Taxon-X Schema*” (derived from the implicit structure of scientific publications in biological systematics and already under development at AMNH). We will use this corpus of marked-up literature to explore the automated extraction and application of imbedded scientific data specifically:

- Extraction of *scientific names* (for possible inclusion in biological names management systems);
- Extraction and searching of *morphological characters*;
- Extraction of *species distribution data* (for plotting in GIS and contribution to the emerging world database of biological diversity)
- Extraction of *collection locales/events* for inclusion in a gazetteer of collecting events.

New search & retrieval strategies and a new powerful search engine will be applied to searching the corpus of marked-up literature. Accomplishment of these goals will result in the extension of worldwide access and increasing the usefulness of an extensive body of scientific literature that is now generally restricted to users who have access to major research libraries. Relatively few computer scientists have become involved with the “problem set” of biological systematics and biodiversity, the collaboration of German and US IM/IT specialists will build on informal collaborations already in place and should lead to broader and more ecumenical design and international dissemination.

The core work of this project will contribute to the design and implementation of key ontological services that will *support biological and conservation education* at all levels. The proposed project will *fundamentally contribute to research and education* in biology by prototyping a model to extend the practical usefulness of the enormous body of legacy literature in biosystematics and contributing to the completion and continuing updating of international biodiversity databases. The corpus of literature to be made available on the Web -- as well as the methodology and protocols for managing them -- occurring in the context of the “Commons” philosophy – will create a new model for international access to this literature – *those who have never before had access to the great biosystematics libraries* will be given access and *society will be benefited* by strengthening of the information resource base serving the international conservation of biological diversity.

TABLE OF CONTENTS

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

Section	Total No. of Pages in Section	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
A Project Summary (not to exceed 1 page)	1	_____
B Table of Contents	1	_____
C Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) (Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	15	_____
D References Cited	3	_____
E Biographical Sketches (Not to exceed 2 pages each)	6	_____
F Budget (Plus up to 3 pages of budget justification)	9	_____
G Current and Pending Support	1	_____
H Facilities, Equipment and Other Resources	1	_____
I Special Information/Supplementary Documentation	5	_____
J Appendix (List below.) (Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	_____	_____
Appendix Items:		

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

TABLE OF CONTENTS

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

Section	Total No. of Pages in Section	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
A Project Summary (not to exceed 1 page)	1	_____
B Table of Contents	1	_____
C Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) (Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	15	_____
D References Cited	3	_____
E Biographical Sketches (Not to exceed 2 pages each)	2	_____
F Budget (Plus up to 3 pages of budget justification)	5	_____
G Current and Pending Support	1	_____
H Facilities, Equipment and Other Resources	1	_____
I Special Information/Supplementary Documentation	5	_____
J Appendix (List below.) (Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	_____	_____
Appendix Items:		

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

Introduction

We are proposing to design and test approaches to mark-up and extraction of scientific data from a corpus of texts drawn from the biosystematics literature of entomology and to develop a set of applications informed by a basic ontology for biosystematics information, including powerful search and retrieval strategies to operate on this corpus.

We will create a digital corpus of current and retrospective entomological literature drawing from the literature of ants -- both monographs and serials. We will test approaches to automated XML mark-up using our hypothesized “*Taxon-X Schema*” (derived from the implicit structure of scientific publications in biological systematics and already under development at AMNH). We will use this corpus of marked-up literature to explore the automated extraction and application of imbedded scientific data focusing on the development and testing of four key applications:

- Extraction of *scientific names* (for possible inclusion in biological names management systems);
- Extraction and searching of *morphological characters*;
- Extraction of *species distribution data* (for plotting in GIS and contribution to the emerging world database of biological diversity)
- Extraction of *collection locales/events* for inclusion in a gazetteer of collecting events.

The corpus of marked-up literature will be searched using a powerful search engine – already under development at Magdeburg -- in the framework of a basic unifying ontology for managing the information and data of systematic biology.

Accomplishment of these goals will result in the extension of worldwide access and increasing the usefulness of an extensive body of scientific literature that is now generally restricted to users who have access to major research libraries. Relatively few computer scientists have become involved with the “problem set” of biological systematics and biodiversity, the collaboration of German and US IM/IT specialists will build on informal collaborations already in place and should lead to broader and more ecumenical design and international dissemination. It will also result in development of technical innovations that will serve as a prototype for all the disciplines of biological systematics and contribute to the emergent standards for international scientific digital libraries.

The literature of biological systematics

Biological systematics is the science that lays the definitional foundations for organismic biology and for biodiversity conservation; the published literature (monographs and serials) of biological systematics is the detailed record of this science.

“*Species*”—together with genes and ecosystems -- comprise one of the three fundamental levels of biological diversity as defined by the Convention on Biological Diversity (1992) and “*species*” is the most universally understood unit of biological classification. More than 1.2 million species have so far been described, and estimates vary from 2 to over 100 million species

The *description* of zoological species is highly regulated through the Code of the International Commission of Zoological Nomenclature (ICZN, 2000) (comparable codes exist for botany and microbiology); description of a new species requires compliance with a formally prescribed set of rules (and the ICZN designates the 10th edition of the *Systema Naturae*, published by Carolus Linnaeus in 1758, as the starting point of zoological nomenclature).

The most fundamental rule prescribed by the ICZN is that each name must be a unique Latin *binomen*, existing of a Genus name and species “specific epithet” -- thus the red wood ant has its scientific equivalent *Formica rufa*. Since this name is used throughout the sciences, each publication or artifact of this species can be found by use of this unique *binomen*. Further requirements to make a species name valid are: publication of the original description in print (or more recently in digital form) and designation of a *type specimen*, normally with auxiliary collection event data attached, (such as collector, collection locality and date, biological notes and location of the type specimen).

The ICZN Code stipulates that, in the absence of other considerations, the earliest valid name assigned to a species is its correct name. This principle of priority means that all subsequent treatments of the taxonomy of any group must cite the original description. Thus in assessing the value of the legacy of literature in biological systematics, it is important to note that the cited half-life (an indicator of the relative obsolescence of scientific journals) of systematics literature is much longer than in most scientific disciplines; citation of original descriptive literature and of the succeeding lineage of published descriptions is required as part of formal revisions of biological names. (Hence, for any given organism, it would not be unusual to cite work by Linnaeus from the 18th Century or other first authors of scientific names and to cite all successive revisions of those names.) Thus in making digital library of systematic literature requires a commitment to provision of sustained access to all original descriptive literature.

In this systematic literature, a standard unit of reporting, at any taxonomic level (e.g. subspecies, species, genus, family or higher), may be described as a “taxonomic treatment” (or “taxon treatment”). Biologist-systematists, produce such a descriptive element, typically after extensive fieldwork and comparison of specimens, in order to accurately describe or revise the description of a specimen or group of specimens. When this work is accepted for publication, it may become a source of authority for the species names and identifying characteristics and is typically listed in one of several recognized bibliographic services specialized to the task (SEE for example: http://www.biosis.org/htmls/free_resources/name_counts.html). Typically, such taxon treatments conform to generally discernible, regular standards. (AMNH has carried out preliminary definitional research on this “taxon treatment” concept involving both systematists and information specialists and this hypothesis will be rigorously tested in the course of this proposed research.) The taxon treatment may occur in monographs or journal publications and this proposal, by operating on diverse formats and age-classes of literature will test the utility of this concept.

Of course, some researchers working at institutions that do not have extensive collections of literature may be able to acquire copies of certain works through interlibrary loans or by requesting photocopies from colleagues, but these methods are time consuming and require delays in research. In the case of very rare works, however, travel to a library holding the work may be the only option available, since libraries will generally not loan originals. But before requesting a loan or visiting a library to study a rare work, researchers must know (or at least suspect) that the work contains the information they need. A comprehensive digital library would make it possible to rapidly locate numerous references on taxa of interest, even those that may have escaped inclusion in previous bibliographies. A digital library of scientific literature also has potential benefit for preservation of literature. With nearly all of the information needed by students of the early literature available online, the requirement to handle the fragile original materials for simple reference work will almost certainly decrease.

Why Entomology?

Insects comprise 80% of the animal species on earth and play a prominent role in the research of many biological disciplines, including developmental biology and molecular genetics. Moreover, insects have a huge impact globally in nutrient cycling, agriculture, forestry, and human and animal health. Insects are our most important competitors for food, while at the same time providing crucial ecological services through their roles as pollinators, predators, and parasites. Many of the most severe insect pests are species introduced into a new area; thus, effective quarantine procedures demand access to knowledge from beyond the known local concerns to a worldwide perspective.

Why Ants?

The full literature of entomology is too large to be treated within the scope of this proposal we therefore propose to concentrate our research on one well-treated and technically well-served group.

Among insects groups ants are uniquely suited for specialized treatment. Ants are one of the few exceptions for which online digital catalogues of all their species exist. They included (in January 2002) 10079 valid species world wide, based on over 18,000 originally published names (antbase.org). The primary taxonomic literature, that is the literature that includes descriptions of new species or revisions affecting the status of a species, includes 3,800 publications. These publications average 22 pages, and thus the full body of literature comprises 80,000 pages. If secondary descriptions of taxa of different life stages are included, the number of publications increases to 8,000 (Ward et al., 199?). Finally, if all the publications on ants alone are included, the number of publications raises to well over 30,000 online for ants (Formis, 2002).

Ants are not only one of the ecologically most important animal taxon globally, but also one of the taxa, which are the most advanced in terms of digital access. A complete, continually updated digital list of species exists as backbone of a, which includes the provision of a page for each taxon from which links are given to most of the currently available online taxonomic information (DNA sequences: GenBank, regional databases, imagery, bibliography, distribution data, etc.), and an increasing amount of full text publications of the original descriptions in form of pdf documents.

Methodology, Standards, and Outputs

A Proposed Basic Ontology for the Information of Systematic Biology

Informatics specialists have given considerable thought to the normal structure of information in biological systematics. The Darwin Core model¹ developed primarily by Viegals at the University of Kansas – in conjunction with the Taxonomic Data Working Group (TDWG) -- expresses a proposed core standard for a biological specimen. The Darwin Core model suggests a basic ontology for systematic information that can support and inform searching across our full corpus of entomological literature. (Moreover, we believe that this ontology has potentially broader applicability to most types of digital objects in the natural history information environment.) The ontology here described provides a theoretical framework for the subsequently proposed technical research.

Key services in this ontological system will be:

- a selected *digital gazetteer* [Alexandria Project www.alexandria.ucsb.edu/; TGN: <http://www.getty.edu/research/tools/vocabulary/tgn/>; GEOnet Names Server <http://164.214.2.59/gns/html/index.html>];
- a *biological names resolver* [ITIS: <http://www.itis.usda.gov/>; Species 2000: <http://www.sp2000.org/>; Ubio: < <http://www.ubio.org/>>];

- *a chronological authority system* [AGI – under development];
- *a macro-thesaurus of entomological and zoological descriptors*. [including: Foote, Richard H. *Thesaurus of Entomology*. College Park, MD: Entomological Society of America, 1977. / *The Torre-Bueno glossary of entomology* / compiled by Stephen W. Nichols; managing editor, Randall T. Schuh; including Supplement A by George S. Tulloch. . [New York, N.Y.] : New York Entomological Society in cooperation with the American Museum of Natural History, c1989 / Zoological record: http://www.biosis.org/free_resources/zr_taxhier.html; Pennak, Robert W. *Collegiate dictionary of zoology*. New York, Ronald Press Co. [1964]].

These designated services will be invoked interoperatively to resolve native ambiguity in given information objects and to support coherent integration of searching across our full corpus thus addressing the need to resolving variability in names, dates, geographical locales and subject descriptions.

Technically development of this ontological framework will be modeled on design already under development by University of Massachusetts, Boston. The UMB group has been exploring mechanisms for semantically based resource discovery. (See the link on “Semantic resource discovery” at www.cs.umb.edu/efg. This work expresses ontologies in RDF and RDF-Schema and registers web services, such as those we mention above, in public UDDI registries (see www.uddi.org). Clients seeking services with keywords not listed in the metadata of the registered services, explore the ontology to decide which service description may or may not be a likely target. In this way, client applications can avoid querying data sources that have a high likelihood of having no answer to the client’s query, conversely reduce the likelihood of *failing* to query a source which might provide an answer. For example, if the metadata of a certain service expresses that it can answer queries about plants, but the client query is expressed in terms of flowers, then an ontological expression to the effect that a flower is-a-part-of a plant will lead to a query against the plant database. The Electronic Field Guide (EFG) Group (UMB/ AMNH) will express our ontologies in *RDF*; write service wrappers around the participating databases using the *Web Service Description Language* (WSDL) [<http://lionhead.cs.umb.edu/wsdemo/jsp/index.jsp>] and register the results in a public UDDI server. We will also provide demonstration clients and an API by which others can write such ontologically informed discovery mechanisms. The search engine to be developed by the Magdeburg group will make use of these ontologies as well, evaluating whether it can productively use this API, or must opt for tighter integration.

Contributions to the Development of Ontological Services

Digital Gazetteers:

All specimen data is geo-referenced but geo-referencing occurs with extreme variations in resolution. (SEE http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/session1/moritz.htm) A long term goal for data management in the museum setting is the development of digital gazetteers that appropriately resolve the dilemmas of weakly expressed geo-spatial resolution of specimen data (and that permit expressions of confidence in the expressed resolution); (work to address this problem has been undertaken at the California Academy of Sciences – SEE below).

A major problem digital gazetteers is the current lack of inclusive, retrospective names with appropriate time-qualifications; most gazetteers for most uses seek to have the most current available names, hence there is secondary interest in providing data for places that may no longer

exist or that may have moved. A goal for this project will be to sample the geo-spatial referencing imbedded in the “*material studied*” section of the *taxon treatment* and to test the feasibility of validating and introducing those data into digital gazetteer systems like the UC Santa Barbara Alexandria Digital Library (ADL) Project in conformance with already specified content standards (http://www.alexandria.ucsb.edu/gazetteer/gaz_content_standard.html). These extracted geo-names can be run against the digital gazetteers to yield both enriched geo-references for those names identified and a subset of excluded names. This identified subset can in turn be run against other gazetteer sources (Getty Thesaurus of Geographic Names: (TGN) or the US NIMA Geonames Server (<http://gnpswww.nima.mil/geonames/GNS/index.jsp>) The resultant residue of untreated names can be manually enhanced by comparison with sources such as field notes or other related digitized archival materials and submitted.

Published locality names are linked to collectors, and thus we will develop locality names that correspond with collecting events of specific collectors and expeditions by comparing field notes and gazetteers provided by the scientists.

Biological Names Resolver: The potential complexity of scientific names has been discussed above. A number of initiatives are underway to create inclusive systems for the validation of scientific (and common) names and to associate valid synonyms and identified orthographic variants with those names. AMNH and OSU have been collaborating with IT IS for two years and in addition, AMNH has been working with the Woods Hole Oceanographic Institution/ Marine Biological Laboratory initiative (UBIO) to work toward its effective acceptance by the systematics community.

A Chronological Authority System: Variations in expressions of recent time are largely resolvable with existing tools but there is a need for harmonization of time elements with respect to geologic time. AMNH is currently cooperating in an effort by AGI to resolve this problem by creating a Web-based standard by community consensus.

A Macro-Thesaurus of Entomological and Zoological Descriptors: The NLM has invested significant resources in development of their UMLS system. We propose to explore a development of this model focusing on the controlled vocabularies of entomology and zoology (as noted above).

Analysis of the entomological literature

Analysis of biosystematics texts -- XML Schema: We propose to apply rules derived from an *XML Schema* (<http://www.w3.org/XML/Schema>) to refine a mark-up standard (both retrospectively and in future publication) for scientific literature; this mark-up standard is informed by analysis of both the structural and semantic conventions of disciplinary scientific literature (in this case, biological systematics/entomology).

Semantic Analysis: Within the sciences, publishing tends to adhere to more formally procedural (if not tabular) standards for publication. Within the biological systematics community, published descriptive treatments of taxa (“*taxonomic* /or *taxon treatments*”) must conform to generally recognized standards to constitute a valid original description or revision. (SEE Above) Such treatments are described, among other places, in J. Winston’s *Describing Species*ⁱⁱ and as noted above, the ICZNⁱⁱⁱ closely regulates the naming of animals. The taxon treatment object describes, among others, the following kinds of information, each marked in a way that can be extracted by an application:

taxon name. The accepted or valid scientific name. Without this name it is impossible to compare information across publications or data sources, because other nomenclature, such as common names, has no agreed upon structure, independence of local usage or source of authority.

authority. The details of the publication of the currently accepted (valid) name of the taxon. This allows the systematists to verify that the name is being correctly used.

history/synonymy. References to publications of the previous names if this is a revision. This establishes the context, the research lineage describing the specimen under revision.

materials studied. The actual specimens examined for the treatment, with brief descriptive data (which may include: place of collection, date of collections, caste /sex /developmental stage).

diagnosis. A treatment typically contains the key diagnostic features, or *characters* of the organisms, that allow the taxon to be distinguished from others. Scientists use these diagnostic characters to identify a specimen in hand. The diagnostic element is eligible for finer resolution, because if the individual characters are marked up, then it is possible to automatically enable interactive identification software for the species involved. UMass-Boston has implemented such use of XML (<http://www.cs.umb.edu/~ram/efg/>). An important issue to be studied in such increased resolution is the need for subdisciplinary-specific mark-up standards for diagnoses, since the diagnostic characters vary significantly among different taxa.

descriptions. Taxon treatments often contain additional information about the organisms being treated; this information may include descriptive details that are not immediately relevant to diagnosis (such as ecological descriptions about habitat, behavior or relationships with other organisms). This material provides yet another potential derivative of the data extraction process. We will have the capacity to extract the essential descriptive text from the publications being treated and as described below, we propose to test their usefulness by offering views of these texts designed for different audiences using the informal education and public outreach experience of the AMNH National Center for Science Literacy, Education and Technology.

Keys. Identification aids, usually linked to taxon names at a variety of hierarchic levels;

figures. Illustrations of morphology of other information usually associated with taxon names;

biotic associations. Information on host associations, e.g., plants.

Formal Structural Analysis: Analysis of the conventions of publishing reveals remarkable uniformity in structural publishing formats across the contemporaneous literature and stability within individual series over time. This is largely an artifact of the above-described rigorous standards (ICZN) for systematic literature. This relative consistency limits the problem set of formatting challenges to be coped with in structural-analytical work. In the case of AMNH's approximately 220,000 pages of scientific publications, the editorial/publishing standards have remained remarkably stable over time. Discussions with colleagues at BIOSIS^{IV}, working with 5000+ scientific serial titles, suggest that as few as 400 structural formats may actually be in use among these publications.

Deriving the Taxon-X Schema for Systematics: In October 2000 a 3-day working session was held at AMNH involving both biological systematists and information technologists. The session introduced basic digital library concepts and then proceeded to explore information-using behaviors among participants individually and in focus groups. The workshop concluded with a close analysis of the structure of taxon treatments within the literature of systematics. The result of this session was an analytical outline for a standard taxon treatment – which has subsequently been expressed as a draft *XML Schema* (here referred to as the *TAXON-X Schema*).

Mark-up of the entomological literature

XML Schemas: We propose to use XML as the most suitable encoding scheme for our proposal. XML offers several advantages: widely accepted standards exist, supporting a breadth of open source and commercial tools; an XML Schema permits strong data-typing that can promote robust data exchange and analysis; most widely used database systems now can generate XML, making it ideal for our documents to participate in database federations; and, finally, XML is particularly amenable to diverse data descriptions.

XML Schemas (and XML DTD's) provide a specification for the controlled grammar of a collection of documents marked in XML. Schemas provide for strong data-typing, which implies that mark-up errors can be detected by a validating XML parser. Having strong typing can also guide the development of applications that read the document when those applications are themselves written in a strongly typed language such as Java. In what follows, "*schemas*" will refer to both DTD's and XML Schemas when the distinction is unimportant.

We propose to extend and develop a markup system using XML, which offers several well-known advantages including widely accepted standards and support by a breadth of open source and commercial tools. XML will allow for multiple uses and presentations of the same documents in different systems. XML markup can be done at any level of resolution. The AMNH digital library initially chose to use a very low-resolution approach—a subset of the *TEI Lite* tag set (<http://www.uic.edu/orgs/tei/>; <http://www.uic.edu/orgs/tei/lite/>). This tag set was used to capture only the broad structural elements that allow efficient navigation through the document, but limited manipulation of the content. In contrast, this project will develop a higher resolution tag set that will provide better access to the data within the target texts. We propose to develop and apply the above-described *Taxon-X Schema* to this body of marked up literature.

Production and evaluation of automated mark-up guided by the *Taxon-X Schema*

Experimental corpus of text: As a part of the work here proposed, we will compile an experimental corpus of digitized text. This corpus will include ant literature *already digitized* as a part of the AMNH Digital Library using a reduced version of the *TEI-Lite* standard (<http://www.uic.edu/orgs/tei/lite/>). This collection will then be augmented with additional texts to be digitized by vendors under this proposal. We expect that the resultant corpus of texts will approach 30,000 pages.

Selection criteria for the inclusion of the publications are as follows. The publication will be in English, permission for use the publication must be obtained. A preliminary list of publications with the highest value to the current understanding (i.e. synoptic studies (monographs)) has been provided by Bolton (1994) and including 474 publications until 1995, plus 55 additional publications published since then. An additional number of publications will be selected to represent a set of individual description of species, (not to exceed 500 publications) and thus totaling approximately 1000 publications, with an average of ca 22 pages (estimate based on calculating page number of 3,800 publications covering ant systematics since 1758, Agosti and Schultz, Atherton-Seidal proposal to the Smithsonian Institution).

Digital Capture of Texts: It is our assessment of current technology that the most effective and efficient method for creating an extensive digital library is to contract with a third party vendor to scan documents in uncompressed TIF format (either at 600 dpi bi-tonal or 400 dpi grayscale, depending on the presence of halftone illustrations) for archiving. Derivatives files in JPG and/or GIF format will then be made for web-based delivery. Additionally, the vendor will perform a single pass re-keying of the material and tag all documents in an XML-DTD schema as devised by the AMNH digital library. Deliverables will be in CD-R format and eventually uploaded for

archiving onto a Storage Area Network maintained by the AMNH Systems Division. This approach will achieve a level of accuracy approaching 99.95 percent. For documents that cannot be "guillotined", or for other reasons cannot leave the premises of the Museum, we will create digital documents in house using a Minolta digital publication scanner.

Automating mark-up: The regularity imposed by editorial standards of this literature should permit simple PERL scripts based on prescribed rules, sensitive to the discrete structural formats of the text, to find document object boundaries defined by required titles and format. The PERL can then output the text with appropriate mark-up, guided by the schema. An iterative cycle of mark-up and evaluation will be run on a representative subset of treatments from the corpus; this process will refine this PERL script and once it is stable, we will systematically apply a validating XML parser and record the proportion of tags that are not compliant with the Schema and therefore need hand editing using a Schema-aware editor such as XMLSpy. This will allow us to audit the accuracy of the procedure to make estimates of the cost of future mark-up endeavors. Our target is to require no more than 5-10 minutes per treatment. We expect the refined mark-up process to optimize the required human effort.

In each of the above instances, integral association of attributive metadata (specifying provenance of the data and legal rights) with the data elements will be critical to establishing the success of methodology. Having produced a normalized set of texts for testing, these texts will be operated upon by the collaborating institutions. This approach should enable users to extract only those elements upon which they seek to operate.

Intended Products of the Research:

A refined Taxon-X Schema: An early product of this work should be a refined version of the Taxon-X schema for application in digitization of other biosystematics literature and for possible prospective application in current biosystematics publishing.

Taxonomic Applications:

Prototypical links to ITIS and services such as GENBANK and other services: OSU will focus on the development of directly links to the ITIS system with a view to prototyping the delivery of current taxonomic names and synonyms. This service will provide a model application for deployment by distributed taxon-specific databases (there are several at AMNH alone).

Taxonomic Revision Discovery and Notification Agent (RDNA): The *Integrated Taxonomic Information System (ITIS)* is a collaboration among the U.S., Canadian and Mexican governments (*ITIS* (<http://www.itis.usda.gov/plantproj/itis/>); *ITIS*ca* (<http://sis.agr.ca/itis/>)). Its purpose is to manage taxonomic names, providing a common authority for scientific names, synonyms, common names, and other taxonomic data. *ITIS* holds over XXXX records, of which slightly over half have had the authority of their taxon names verified by experts (<http://www.itis.usda.gov/plantproj/itis/contrib.html>).

The AMNH Digital Library Project and Ohio State have worked closely with ITIS to provide a global list of valid ant names. At this writing, the final work of populating ITIS with synonymous names and combinations is being accomplished. Since the scientific name is *the* primary key to most biological data, revision of names can obscure reference to those publications. ITIS, in turn, provides the link to specialist-level databases on the taxonomy and systematics of taxa and the connections effected by these resources to other on-line resources.

For animals, there is a source whose role is to be a *de facto* repository of current names-- *Zoological Record* (http://www.biosis.org/htmls/products_services/zoorecord.html). However, *Zoological Record* is essentially passive; that is, a client—human or software—seeking to learn if a particular name is current and valid must take action on its own to find out. Furthermore, data sources such as *ITIS* are ignorant of any name revisions affecting their data unless their operators or cooperators take specific action to update records.

AMNH, Magdeburg, OSU and UMB will collaborate on development of Web-accessible RDNA will seek to address these issues with several kinds of web interfaces available both with a browser interface for human use and a public application-programming interface for distributed application use. In the simplest interface, the human or software client will specify a taxon, typically a species, whose current name status is desired. The RDNA will find and examine the corresponding treatment in our corpus and also query a known collection of taxonomy servers to decide which reveals the latest accepted taxon name, and return the data for that name, including the authority for the name, together with optional treatments if the distributed sources have them available. The project will also explore possible ways of integrating such notification mechanisms with Magdeburg's search engine development (see below), which is another envisioned outcome of the project

The second service that the RDNA will address focuses on new publication. A person wishes to be notified of the publication of a new revision can register the taxa s/he wishes tracked. The Notification Agent will daily poll its data sources to inquire about those taxa and send mail to the registrant when a revision is published. The system will also allow for the registration of a URL at which a remote agent is providing a CGI program that will accept notification in the form of a SOAP (Simple Object Access Protocol: <http://www.w3.org/TR/SOAP/>) message of published specification. That remote agent is then free to make whatever use it chooses of the notice, such as forwarding it to a human administering the database or perhaps automatically updating the remote database directly from the message.

Morphological Application and EFG:

The Magdeburg team will develop a taxonomically informed search engine for morphological queries. Users can query this search engine, by entering characteristics of the taxon sought, e.g., 'yellow head', 'size 12.5 mm'. The main objective is to provide a system that is operational, and that suits the biologists' needs.

Contractors at UMASS Boston (UMB) will extend their exploration the application of their Electronic Field Guide (EFG) generation techniques (www.cs.umb.edu/efg) to the data marked up in a taxon treatment. The EFG tools accept descriptive characters of taxa in XML form (or across an ODBC bridge) valid for an XML-Schema of the project's design, and from that produce an object-oriented, web accessible identification tool for use in the field or the laboratory. Therefore, the main task of this subproject will be the extraction and transformation of *Taxon-X* valid data into data suitable for EFG generation. The likely route for this will be Path for extraction and XSLT for transformation.

Geo-referencing : a Gazetteer of Collecting Events and Derivation of Species Distributions

As noted above, one of the primary elements of modern taxonomic treatments is the “material examined” section: a summarization of the extrinsic data associated with the specimens that were studied by the authors. Essentially, this is a listing of the collecting events - the place, time, people, and methods employed -, which produced the evidentiary basis for the scientific work. The data are important not only in the context of the scientific publication itself, but also in

documenting the time and place of surveys of biodiversity by specialized collectors. Access to a comprehensive resource of who collected where, and when, would be of great value in understanding the details of the distribution of biodiversity. In addition, it would greatly facilitate the process of digitization of the billions of specimens in natural history collections.

Two major issues inhibit this process. First, the transcription of data is tedious and error-prone. In entomology, this is compounded by the large numbers of specimens involved; (even a modest university collection typically contains between 1 to 4 million specimens; AMNH holds more than 18 million; major national collections have tens of millions of specimens). A second major impediment is geo-referencing, placing latitude, longitude, and error estimation on the place name at which a specimen was collected. The California Academy of Sciences – following on work initiated by Moritz while based in SF -- has developed an application to address this problem: [<http://www.calacademy.org/research/informatics/GeoRef/index.html>] and this application will be tested. With such data, biodiversity records can be correlated with a huge array of environmental data; without them, the specimens are mere historical curiosities. The mark-up work proposed here can contribute significantly to reducing the cost in time and money associated with geo-referencing.

The goal is to share the task of geo-referencing by the creation of a gazetteer of collecting events. Many of the localities cited in the literature and on specimen labels are, logically, not within the populated political units that are typically recorded in gazetteers. A significant number of collecting events are shared among collections and across taxa because the scientists share and exchange specimens and information on productive sites. It is also the case that places named as recurrent collecting sites for biosystematics may not hold sufficient interest for non-systematists to support inclusion in major gazetteers. There are two primary sources of the collecting event data: existing collection databases and the published literature. The Ohio State group has started work on a prototype gazetteer for entomological collecting events, building on available on-line databases. The number of such resources, however, is relatively small at present, fewer than a dozen institutions. This gazetteer is one of the work products that emerged from the *Scientific and Technical Advisory Group* meeting for the *Digitization of Natural History Collections* (June 2002, Washington, DC) of the *Global Biodiversity Information Facility*. As part of this proposal, OSU and AMNH will seek to supply to that effort the data on collecting events extracted from the marked-up literature, thus substantially increasing both the temporal and geographic range of data. The geo-referencing work will follow the guidelines published by the Mammal Networked Information System (<http://elib.cs.berkeley.edu/manis/GeorefGuide.html>).

The result of the gazetteer effort will be an application that will enable a user to find the geographic coordinates associated with a collecting event, either by simple string matches of place names or by following a political hierarchy. Collecting events may be found within political units or bounding boxes specified by coordinates. The goal, then, is to harness and employ the mass of collecting event data that has been published in the taxonomic literature toward the eventual goal of documenting the world distribution of biodiversity. Access will be effected both through Web-interfaces for people as well as XML-mediated web services based on the Darwin Core.

Search and Retrieval Strategies and Search Engine Development.

In a preparatory activity, the Magdeburg team has built a prototype based on Microsoft SQL Server^Y. It mimics most of the functionality envisioned and can be demonstrated. However, it does not yet meet the following requirements. The Magdeburg team will seek to accomplish:

Search results: While this point is obviously important, there are certain aspects that deserve special mention in the biosystematics context. The first one is that technical terms and proper nouns occur frequently. A standard thesaurus does not contain them. Nevertheless, the system should deal with them in a meaningful way. Another issue is that biosystematics documents contain numbers. For instance, a document might state that the length of a certain species ranges from 8 to 12 mm. Even though full-text search mechanisms typically ignore them, it must be feasible to search this kind of information just as well.

Usability: The user interface must follow the biologists' intuition, and both technical project members and biologists will collaborate closely to achieve this. Given that other relevant information on ants and other species is available online as well, search results will contain hyperlinks to such related information. This requirement also includes online documentation, use cases made available to prospective users etc. A further issue is minimization of user input: A biologist classifying a specimen of ant. should have to record and type in only a few distinctive characteristics of the ant in order to determine its species. The system should guide the user and ask for such distinctive information such that user input requested is minimal.

Flexibility and extensibility. Modern search engines feature user-adaptivity in one way or the other. This means that user input alters the search technique, resulting in new, typically better search results. Some examples of user input and respective adaptation schemes are as follows: Extensions to the thesaurus, sense disambiguation of query terms 'by hand', and relevance feedback regarding preliminary query results. This gives rise to several research questions that the project wants to investigate:

- **Relevance-feedback models**, i.e., ways how such feedback may be interpreted. The project will evaluate which models should be applied.
- **System implementation** incorporating a large number of adaptation schemes requires the capacity for dynamic selection of the most appropriate and most efficient in context.

Robustness and fault tolerance. Briefly, the project intends to attain robustness and fault tolerance by using off-the-shelf database technology. This reduces application-development complexity and allows leaves aside implementation of generic features, notably storage management, query processing, and recovery. Furthermore, it is relatively easy to distribute the system, once workloads become too high for a single node (Grabs, Böhm, Schek, 2001a). Using a DBMS is an important design decision that significantly affects the content of the project, at least its technical part. The project must use the database interface, i.e., data must have a relational structure, and access is declarative via SQL.

Performance quality. This requirement has several facets -- performance should be up in the following circumstances as well: scalability to include large number of documents as well as large number of concurrent users; high flexibility in the sense described above, i.e., the user has several possibilities to adapt the search; support for complex queries, i.e., complex queries that are the outcome of query expansion, not necessarily the native queries as issued by the user.

Spatial information. The system envisioned would include spatial information, i.e., information where and when a certain species has been encountered. Search and classification will use such information. Spatial proximity of the habitat of the current ant and the living space according to the document will be a search criterion. The project will implement such a 'spatial similarity measure' by integrating existing geo-thesauri, e.g., Geonames. To this end, it will address the following points: investigate physical design alternatives for the geo-thesaurus, e.g., materialization of the entire thesaurus in the database, access via service interface, or new, 'intermediate' solutions; design of mechanisms that allow the user to disambiguate search results

from the geo-thesaurus by hand, and to adapt the distance metric. This is a special case of the work package 'flexibility' described earlier; record such user input as feedback and make use of it when evaluating queries in the future; development of reporting and visualization tools for spatial information.

Reporting and online documentation. Given that other information on ants has been made available online already, search results will contain hyperlinks to related information. A related issue is that access to the system should be easy and intuitive. This includes online documentation. The search engines will be set up on the US servers in collaboration with the Magdeburg team. Preparation of the existing data servers will be done by the US teams to implement requests from the German search engine development team.

Auditing and evaluation: measurement of necessary levels of mediation/data enhancement
In all aspects of the project, careful measurements will be kept of requirements for investment of time and resources in human modifications and data enhancement. This auditing will include close accounting of all unit-investment of human effort, particularly with respect to interventions in the mark-up process where careful distinctions will be maintained with costs associated with original OCR work, one time efforts involved in the refinement of our *TAXON-X Schema*, and any residual requirements of effort following revision. Our goals will be to provide empirical assessments of the tested approach and to offer guidance for prospective efforts to mark up legacy publications and future publishing.

Evaluation of results: In addition to this auditing and evaluation process, iterative testing and ongoing feedback from Museum scientists and educators will provide evaluative data throughout the project cycle. We will in addition solicit evaluations through Internet discussions (such as TAXACOM, Entomo-L, and Ornith-L) as well as through Web-based surveys and Website specific solicitations of feedback.

Related research proposals:

As noted above, the ICZN provides that the conclusions of taxonomic treatments retain their validity. Since the publication of the *Systema Naturae* (10th ed.), there has been a dramatic change in the quality and quantity of data that accompany taxonomic treatments. Modern works typically include the examination and documentation of hundreds of specimens, anatomical descriptions and illustrations, phylogenetic analyses, and often molecular sequence data. The challenges of marking up the "classical" literature and "modern" literature (at least in the 20th century) are quite different. Our focus is on this newer more elaborated work, and a parallel effort is being undertaken to digitize the old rare entomological monographs (a collaboration between Cornell and the Deutsches Entomologisches Institut). We view these two efforts as complementary, and both will contribute to completion of access to the published literature.

Broader Impacts of Project:

The core work of this project will contribute to the design and implementation of key ontological services that will *support biology and conservation education* at all levels. The proposed project will *fundamentally contribute to research and education* in biology by prototyping a model to extend the practical usefulness of the enormous body of legacy literature in biosystematics and contributing to the completion and continuing updating of international biodiversity databases. The corpus of literature to be made available on the Web -- as well as the methodology and protocols for managing them -- occurring in the context of the "Commons" philosophy -- will create a new model for international access to this literature -- *those who have never before had access to the great biosystematics libraries* will be given access and *international society will be*

benefited by strengthening of the information resource base serving the conservation of biological diversity.

Work Plan and Management Plan

Management will be achieved by creation of closed Internet lists, conference calling and yearly meetings of all collaborators. We will conduct bi-weekly conference calls and communicate by e-mail on a regular basis. The first meeting will be in Germany within the first 6 month's of the project to consolidate plans for collaboration (subsequent meetings in Ohio and NY). Dr. Agosti, as a Swiss-based research Associate of AMNH, who has a history of collaboration with all teams, will coordinate work between the German and US teams. Standard project management software will be used conjointly to monitor progress and revise schedules as necessary.

Three-Year Work Plan

Pre-award ongoing work:

Spring 2002: OSU /AMNH collaborate to load world list of valid ant names to Integrated Taxonomic Information System (ITIS)

Sept-Dec 2002: first prototype of entomological collecting event gazetteer based upon existing on-line databases at OSU, University of Kansas, REMIB, ENHSIN, Agriculture Western Australia, Museum Victoria, INBio

Aug-Dec 2002: Communication of synonyms, invalid combinations, unavailable names of ants to ITIS

Year One

Establishment of project communications protocols and of technical interoperability between collaborating institutions. AMNH/ Magdeburg / OSU.

Capture, Collation and Mark-up of Literature: (AMNH)

Digital corpus of ant literature created combining *existing* collection of marked up literature (at AMNH digital library) with additional texts captured by vendor

- Refine analysis and test approaches to automated XML mark-up using hypothesized *Taxon-X schema*.
- Production and evaluation of automated mark-up guided by the *Taxon-X schema*
- Dissemination and community evaluation of refined *Taxon-X schema*.

Ontological Services: (AMNH/OSU)

Develop liaison with existing ontological services (digital gazetteers, biological names resolvers, chronological authority systems) to develop contributory mechanisms (AMNH/OSU)
AMNH will explore development of a modest version of the UMLS system for Web-based implementation as a "macro-thesaurus" of entomological/zoological terms (AMNH)

Development of Search/Retrieval Mechanisms: (Magdeburg)

Study, design and testing of query mechanisms (using existing prototype). (Magdeburg)
Perform user adaptive study for creation of effective search engine (Magdeburg)
Implement results of user study (Magdeburg)

Applications: (OSU)

(Ongoing) Maintenance and elaboration of linkages between OSU database and existing on-line resources at National Center for Biotechnology Information, Museum of Comparative Zoology Types database, Japanese Ant Image database, etc.

(Ongoing) Update of OSU taxonomic database to incorporate newly described species and taxonomic decisions, including annotations from new literature.

Year Two

Data Extraction: (AMNH/Magdeburg/OSU)

- Extraction of *scientific names* (for possible inclusion in biological names management systems) AMNH/OSU;
- Extraction and searching of *morphological characters* (Magdeburg);
- Extraction of *species distribution data* (for plotting in GIS and contribution to the emerging world database of biological diversity) (AMNH/OSU)
- Extraction of *collection locales/events* for inclusion in a gazetteer of collecting events (OSU/AMNH)

Search and Retrieval: (Magdeburg)

Implement query expansion, query adaptation mechanisms and relevance feedback models. (Magdeburg)

Applications: (OSU/AMNH/Magdeburg)

(Ongoing) Maintenance and elaboration of linkages between OSU database and existing on-line resources at National Center for Biotechnology Information, Harvard MCZ Types database, Japanese Ant Image database, etc. (OSU)

(Ongoing) Update of OSU taxonomic database to incorporate newly described species and taxonomic decisions, including annotations from new literature; (OSU)

Development and deployment of XML for taxonomic database and collecting event gazetteer using Darwin Core v. 2 XML schema made operational (OSU)

Design and Development of web-accessible RDNA (AMNH, Magdeburg, OSU)

Year Three

(Ongoing as above) Data Extraction: (AMNH/Magdeburg/OSU)

Applications: (OSU)

(Ongoing) Maintenance and elaboration of linkages between OSU database and existing on-line resources at National Center for Biotechnology Information, Museum of Comparative Zoology Types database, Japanese Ant Image database, etc. (OSU)

(Ongoing) Update of OSU taxonomic database to incorporate newly described species And taxonomic decisions, including annotations from new literature. (OSU)

(Ongoing) Extraction of *collection locales/events* for inclusion in a gazetteer of collecting events (OSU/AMNH)

Development and coding of elaborated species-page for taxon name server queries, including links from taxonomic names to digital versions of original literature produced in project (OSU)

Search and Retrieval: (Magdeburg)

Final version of search and retrieval system user-tested to optimize performance and made operational. (Magdeburg)

Development and deployment of internet-based and web-accessible evaluation mechanisms (AMNH, Magdeburg, OSU)

ⁱ <http://digir.sourceforge.net/prot/darwin3.xsd>

ⁱⁱ Winston, J. *Describing species: practical taxonomic procedures for biologists*. NY, Columbia, 1999.

ⁱⁱⁱ International Code of Zoological Nomenclature. International Commission on Zoological Nomenclature; [editorial committee, W.D.L. Ride ... et al.]. 4th ed London: International Trust for Zoological Nomenclature, c/o Natural History Museum, 1999

^{iv} Linda Sacks, BIOSIS, Pers Comm. (E-mail), 1/18/01

^v Note that the prototype does not make use of the XML extensions of SQL server. This is because they currently are not mature enough as the fundament of a system that shall be operational. For instance, the extensions impose unrealistic restrictions on the text size of XML elements, and the querying mechanisms are intricate as well.

Bibliography

(Agosti, Majer, Alonso, Schultz, 2000) Agosti, D., J. D. Majer, L. E. Alonso and T. R. Schultz. 2000. *Ants: standard methods for measuring and monitoring biodiversity*. Smithsonian Institution Press, Washington, DC. 280 pp.

(Agrawal, Chaudhuri, Das, 2002) Sanjay Agrawal, Surajit Chaudhuri, Gautam Das: DBXplore, 2002. A System For Keyword-Based Search Over Relational Databases. Proceedings of the International Conference on Data Engineering (ICDE 2002)

(Baeza-Yates, Ribeiro-Neto, 1999) Ricardo Baeza-Yates, Berthier Ribeiro-Neto, 1999. *Modern Information Retrieval*. Addison Wesley.

(Borror, Triplehorn, Johnson, 1989) Borror, D. J., C. A. Triplehorn and N. F. Johnson. 1989. *An introduction to the study of insects*, 6th edition. Saunders College Publishing, Philadelphia. 875 pp.

(Chaudhuri, Gravano, 1999) Surajit Chaudhuri, Luis Gravano, 1999. Evaluating Top-k Selection Queries. Proceedings of 25th International Conference on Very Large Data Bases (VLDB 1999), 397-410.

(Chaudhuri, Narasayya, Sarawagi, 2002) Surajit Chaudhuri, Vivek R. Narasayya, Sunita Sarawagi, 2002. Efficient Evaluation of Queries with Mining Predicates. Proceedings of the International Conference on Data Engineering (ICDE 2002).

(Cui, Wen, Ma, 2002) Hang Cui, Ji-Rong Wen and Wei-Ying Ma, 2002. Probabilistic Query Expansion using Query Logs. Proceeding of the Eleventh World Wide Web conference (WWW 2002).
<http://www2002.org/CDROM/refereed/558/>

(Cui, Heidorn, Zhang, 2002) H. Cui, P.B. Heidorn, H.Zhang. An Approach to Automatic Classification of text for Information Retrieval. JCDL 2002: Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries. G. Marchionini and W. Hersh (ed.s). NY, ACM, 2002, p. 96-97.

(Deutsch, Fernandez, Suciu, 1999) Alin Deutsch, Mary F. Fernandez, Dan Suciu, 1999. Storing Semistructured Data with STORED. Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1999), 431-442.
<http://citeseer.nj.nec.com/deutsch98storing.html>

(Fiebig, Kanne, Moerkotte, 2001) Thorsten Fiebig, Carl-Christian Kanne, Guido Moerkotte, 2001. Natix - ein natives XML-DBMS. Datenbank-Spektrum 1(1), 5-13.

(Florescu, Kossmann, 1999) Daniela Florescu, Donald Kossmann, 1999. Storing and Querying XML Data using an RDMBS. IEEE Data Engineering Bulletin 22(3), 27-34.

(Fuhr, Großjohann, 2001) Norbert Fuhr, Kai Großjohann, 2001. XIRQL: A Query Language for Information Retrieval in XML Documents. Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001), 172-180.

(Grossman et al., 1997) David A. Grossman, Ophir Frieder, David O. Holmes, David C. Roberts, 1997. Integrating Structured Data and Text: A relational approach. Journal of the American Society of Information Science. <http://citeseer.nj.nec.com/grossman97integrating.html> 122-132

(Gruninger, Lee, 2002) M. Gruninger and J. Lee. Ontology Applications and Design. Communications of the ACM (2002) 45:2: 39-41.

(Gupta, et al., 2002) Amarnath Gupta, Bertram Ludascher and Reagan Moore. Ontology Services for Curriculum Development in NSDL. JCDL 2002: Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries. G. Marchionini and W. Hersh (ed.s). NY, ACM, 2002, p. 219-220.

(Mitra, Singhal, Buckley, 1998) Mandar Mitra, Amit Singhal, Chris Buckley, 1998. Improving Automatic Query Expansion. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1998), 206-214. <http://citeseer.nj.nec.com/121460.html>

(Mandala, Tokunaga, Tanaka, 1999) Rila Mandala, Takenobu Tokunaga, Hozumi Tanaka, 1999. Combining Multiple Evidence from Different Types of Thesaurus for Query Expansion. Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1999), 191-197.

(Moritz, 2002) Moritz, Tom (co-author) "A Framework of Guidance for Building Good Digital Collections" Washington, DC, IMLS. <<http://www.ims.gov/pubs/forumframework.htm>>

(Moritz, 2002) Moritz Tom, "Building the Biodiversity Commons" D-Lib

Magazine, June, 2002 <
<http://www.dlib.org/dlib/june02/moritz/06moritz.html>>

(Moritz, 1995) Moritz, Tom, Conservation Thesaurus Version (1.0) IUCN (The World Conservation Union) Gland, Switzerland, 1995.

(Moritz, Tom, 1989) Moritz, Tom, "Toward community standards in systematics," SPECTRA: International Journal of the Museum Computer Network 16(2):18-20 (1989).

(Seidl, Kriegel, 1997) Thomas Seidl, Hans-Peter Kriegel, 1997. Efficient User-Adaptable Similarity Search in Large Multimedia Databases. Proceedings of the 23th International Conference on Very Large Data Bases (VLDB 1997), 506-515

(Theobald, Weikum, 2002) Anja Theobald, Gerhard Weikum, 2002. The Index-Based XXL Search Engine for Querying XML Data with Relevance Ranking. Proceedings of the 8th International Conference on Extending Database Technology (EDBT 2002), 477-495.

(Ward, Bolton, Shattuck, Brown, 1996) Ward, P. S., B. Bolton, S. O. Shattuck, and W. L. Brown, Jr. 1996. A bibliography of ant systematics. University of California Publications in Entomology 116:1-417.

Biographical Sketch: Thomas Daniel Moritz

a. Professional Preparation

Pratt Institute, *M.L.I.S.*, December, 1974
Graduate School of Library and Information Science
Brooklyn, New York

Georgetown University *B.S.F.S.(International Affairs)*, September, 1972
School of Foreign Service
Washington, D.C.

b. Appointments

1999-present Boeschstein Director of Library Services (American Museum of Natural History, New York)
1986-1999 Academy Librarian (California Academy of Sciences, San Francisco)
1983-1986 Fisheries-Oceanography Librarian (Univ. of Washington, Seattle)
1981-1983 Friday Harbor Librarian (Univ. of Washington, Seattle)
1979-1981 Bibliographer, WESTFORNET Forest Resource Library (Univ. of Wash., Seattle)
1976 Bibliographer (Contract Employment) National Academy of Sciences, Wash., DC
1975-1976 Reference Librarian, National Natural Resources Library, USDI, Wash., DC

c. Publications

- (i) A list of up to 5 publications most closely related to the proposed project.

June, 2002 "Building the Biodiversity Commons" D-Lib Magazine, June, 2002
< <http://www.dlib.org/dlib/june02/moritz/06moritz.html>>

November, 2001 (co-author) "A Framework of Guidance for Building Good Digital Collections" Washington, DC, IMLS. < <http://www.imls.gov/pubs/forumframework.htm>>

March, 1996 *Conservation Thesaurus Version 1.0* IUCN (The World Conservation Union) Gland, Switzerland, 1995.

May, 1992 "Reprint collections in the culture of science," in International sharing of polar information resources: Proceedings of the 14th Polar Libraries Colloquy: 3-7 May, 1992, Byrd Polar Research Center, Ohio State University, Columbus, Ohio. Report (Byrd Polar Research Center); no. 4.

June, 1989 "Toward community standards in systematics," in SPECTRA: International Journal of the Museum Computer Network 16(2):18-20 (1989).

- (ii) A list of up to 5 other significant publications

March, 1996 *The Sierra Club Wetlands Reader: A Literary Companion*
Co-editor with Sam Wilson. Sierra Club Books, San Francisco.

October, 1986 "Puget Sound Access: the case history of a 'local database'." Paper presented at the Annual Meeting of the IAMSLIC, Newport, Oregon. Published in "Marine science library networks: national and international" (Proceedings of 12th IAMSLIC Conference Summer 1988).

March, 1993 "Report on diagnostic procedures and a definition of minimum requirements for providing information services on a national and/or regional level," with David Moulder. Special report prepared for UNESCO/IOC/IODE/ Group of Experts on Marine Information Management. (IOC Manuals and Guides No. 30: MIM Publication Series No. 1).

d. Synergistic Activities

Digital Library Development: PI on a \$2 Million, 5-year foundation grant (July, 1999--) to develop digital library at the American Museum of Natural History [AMNH], with goal of integrating all Museum information resources. Accomplishments to date include building technical and staff infrastructure and completing pilot involving digitized information resources in diverse formats.

- ❖ October, 2001. National Science Digital Library CIS National Visiting Committee
- ❖ May, 2000 Panel Presentation, "Information in the New Millennium: Linking Resources," Association of Systematic Collections Annual Meeting: Baltimore
- ❖ March, 2000 Presentation, "Integration of metadata and data from multiple sources in a digital environment," *The Biota of Canada Information Network: Documenting and Analysing Canada's Living Capital for Science and Society*, Ottawa (Sponsored by Agriculture and Agri_Foods Canada, Biodiversity Section, Research Branch)
- ❖ September , 1999, "Geo-referencing the Natural and Cultural World, Past and Present: Toward Building a Distributed, Peer-Reviewed Gazetteer System" Digital Gazetteer Information Exchange, Smithsonian Institution, Washington, DC
<http://www.alexandria.ucsb.edu/gazetteer/dgie/DGIE_website/session1/moritz.htm>

IUCN (The World Conservation Union)/ WCPA (World Commission on Protected Areas)/ BCIS (Biodiversity Conservation Information System): June-Nov.,1994, Sabbatical at IUCN, Gland, Switzerland. Dec. 1995, Appointed to World Commission on Protected Areas IUCN; Chairman, Information Management Task Force. WCPA representative to Biodiversity Conservation Information System (BCIS) (April, 1996); BCIS Steering Committee (Dep. Chair, Feb., 1999 -)

- ❖ Completed Conservation Thesaurus (which has been made available in a hypermedia application by the California State Resources Agency on its CERES Website).
- ❖ Visited IUCN National Programmes in Pakistan and Nepal, completed draft plans for information management within context of both countries' National Conservation Strategies.
- ❖ Conducted Internet and WWW Training Sessions in Europe, Asia and the Americas.
- ❖ Chair/ Deputy Chair BCIS -- Design, development, and leadership consortium of 12 international organizations, including four IUCN Commissions (www.biodiversity.org)
- ❖ Chair IMTF IUCN/WCPA Designed and developed the global PARC (Protected Areas Resource Centres) proposal in conjunction with World Conservation Monitoring Centre (WCMC) (Cambridge, UK). <<http://wcpa.iucn.org/taskforce/info/info.html>>

Database Design and Implementation: Puget Sound Institute (University of Washington, Seattle)(1986-1987) Project Director, SOUND ACCESS (As private consultant: Tom Moritz & Associates, Inc.) Designed, developed, and produced a comprehensive Puget Sound bibliographic database (co-funded by the University of Washington and US EPA). Staff of 18.

Gulf of the Farallones US National Marine Sanctuary (San Francisco) (1988-1989)

Directed design, revision, and editing of geographically defined bibliographic database, including resident geographic and taxonomic authority files (based on the Puget Sound Access model).

Recent Collaborators:

Liz Bishoff, Colorado Digitization Alliance; Priscilla Caplan, Florida Center for Library Automation; Tim Cole, University of Illinois Urbana-Champaign; Anne Craig, Illinois State Library; Daniel Greenstein, Digital Library Federation; Doug Holland, Missouri Botanical Garden; Ellen Kabat-Lensch, Eastern Iowa Community College; Bob Morris, University of Massachusetts, Boston; Ginger Ogle, University of California, Berkeley; John Saylor, Cornell University; Neil Thomson, NHM, London

Biographical Sketch: Donat Agosti

(July 22, 2002)

Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA. email agosti@amnh.org

a. Professional Preparation.

Federal Institute of Technology, Zürich, SWITZERLAND	Ecology & Systematics	Dipl.natw.ETH, 1983
Federal Institute of Technology, Zürich, SWITZERLAND	Systematics	Ph.D., 1988
The Natural History Museum, London, UK	Systematics	1989-1990
Australian National Insect Collection, CSIRO, Canberra, Australia	Systematics	1990-91
University of Zurich, Switzerland / Natural History Museum London	Systematics / Genetics	1991-1993
American Museum of Natural History, New York	Systematics	1993-1994

b. Appointments.

2001 - Research Associate, Smithsonian Institution, Washington DC, USA

1998 - Research Associate, American Museum of Natural History, New York, USA

1994 - 1997 Research Scientist, American Museum of Natural History, New York, USA

1993 - Conservation Officer of the International Union for the Study of Social Insects;

1993 - Chairperson, Social Insects Specialists Group, Species Survival Commission, The World Conservation Union.

2001 - Global Taxonomy Initiative: African Regional Working Group member

c. Publications.

Agosti, D., Majer, J., Alonso, E. and Schultz, T., (eds.). 2000. *Ants: Standard methods for measuring and monitoring biodiversity*. Biological Diversity Handbook Series. Smithsonian Institution Press. Washington D.C., 20+280pp. (including several authored papers)

Agosti, D., Majer, J., Alonso, E. and Schultz, T., (eds.). 2000. Sampling Ground-dwelling Ants: Case Studies from the Worlds' Rain Forests. *School of Environmental Biology Bulletin* **18**, 118pp..

Grimaldi, D., & Agosti, D., 2000. A Formicine in New Jersey Cretaceous amber (Hymenoptera: Formicidae), and early evolution of the ants. *Proceedings of the National Academy of Sciences* **97**: 13678-13683

Brandão, C. R. F., J. L. M. Diniz, D. Agosti and J. Delabie. 1999. Revision of the Neotropical ant subfamily Leptanilloidinae. *Systematic Entomology* **24**, 17-36.

Agosti, D., Grimaldi, D. and Carpenter, J.M., 1998. Oldest known ant fossils discovered. *Nature* **391**, 447.

Related publications, submitted or in preparation.

Agosti, D, and Johnson, N. F., 2002. Taxonomists need better access to published data. *Nature* **417**: **222**.

Saatchi S.S., D. Agosti, K. Alger, J. Delabie, and J. Musinsky. 2001. Fragmentation and loss of primary forest in the Southern Bahian Atlantic Forest of Brazil with Radar imagery. *Conservation Biology* **15** (4), 867-875.

Agosti, D., and Moritz, T. 2002. Ants online – towards the Biodiversity Commons. *Species* **37** (In press).

Grimaldi, D., Agosti, D. & Carpenter, J.M., 1997. New and rediscovered primitive ants (Hymenoptera: Formicidae) in Cretaceous amber from New Jersey, and their phylogenetic relationship. *American Museum Novitates* **3208**, 43pp., 24 figs, 1 table.

Agosti, D., 1994. The phylogeny of the ant tribe Formicini (Hymenoptera: Formicidae) with the description of a new genus. *Systematic Entomology* **19**, 93-117.

d. Synergistic Activities.

1. Social Insects World Wide Web (<http://antbase.org>: A Website to access as much as possible systematics related online information on ants and other social insects, including a complete interactive catalogue of the ants of the world, >800 images, search engines to literature and interactive input/output facilities for standardized ant samples (see also #2); mainly in collaboration with Norm Johnson, Ohio State University. If fully implemented, this project will enormously shorten the time needed to access systematics related information (see *Nature* **416**: 115).
2. Developing standard collecting tools for leaf litter ants (now published as Agosti et al., 2000, see above), including the development of complementary WebPages allowing to deposit and retrieve data in order to enhance the use of ants in biodiversity and conservation research and management. Partially supported by a NSF grant in 1996 (DA Co-PI). This is one of the first attempts to coordinate world wide collecting protocols to enable comparisons of biodiversity data across space and time.
3. Exploring ways to integrate systematics collection data, taxonomic information and full text publications (Virtual Congo Ant project, supported by the Mellon Foundation "Virtual Library Project at the American Museum of Natural History, New York), in collaboration with Tom Moritz and others. If implemented, this would revolutionize the way we work with published data, and will open the way for intensive data mining. > 600 publications online, with funding from Atherton Seidal Foundation for all systematics paper for which permissions are obtained.
4. Exploration of the possibilities to produce predictive distribution models using remotely sensed layers and georeferenced point samples (in collaboration with S. Saatchi, JPL/NASA). To enhance this project we have been awarded a grant from the Japanese Space Agency (NASDA) to retrieve a complete data set of high resolution RADAR scenes covering the entire Atlantic Forest in Brazil, allowing to analyze distribution patterns up to 50m resolution. A grant application to NASA to analyze the entire data set, prepare complementary data sets and finally to integrate those results with georeferenced specimen data is pending. This projects is trying to make the step from "potential" to "actual" near real time distribution patterns, and thus would fill in an important gap in modeling the current distribution of biodiversity, considering logging and other activities.
5. Implementing the "Biodiversity commons" concept through providing prototypes, participating in initiatives (e.g. Global Taxonomic Initiative), and leveraging contracts between publishers and authors of systematics publications.

e. Collaborators & Other Affiliations

i) Collaborators.

Leeanne E. Alonso, Conservation International, Washington DC; Carlos R. F. Brandão, Zoological Museum, University of São Paulo, Brazil; James M. Carpenter, Division of Invertebrate Zoology, American Museum of Natural History, New York; Jacques H. Delabie, Center for Cocoa Research, Itabuna, Brazil; Norm F. Johnson, Biology, Ohio State University, Columbus OH; Brian Fisher, California Academy of Sciences, San Francisco; Jonathan Majer, School of Biological Sciences, CURTIN, Perth, Australia; Ulrich Maschwitz, Johan Wolfgang Goethe Universitaet, Frankfurt, Germany; Tom Moritz, Library, American Museum of Natural History; Bob Morris, Boston University; Mostafa M. Salah, Al Azhar University, Cairo, Egypt; Ted R. Schultz, Smithsonian Institution; Sasan Saatchi, Jet Propulsion Laboratory, NASA, Pasadena; Klemens Boehm, University of Magdeburg, Germany

ii) Graduate and postdoctoral Advisors

Willy Sauter and Georg Benz, Entomology, Federal Institute of Technology, Zurich, Switzerland; Barry Bolton, The Natural History Museum, London; Ruediger Wehner, Zoological Institute, University of Zurich, Switzerland; James M. Carpenter and Rob DeSalle, Division of Invertebrate Zoology, American Museum of Natural History, New York.

iii) Thesis Advisor and Postgraduate-Scholar Sponsor

1 Ph.D. supervision, Sébastien Lacau, Museum d'Histoire Naturelle, Paris

ROBERT A. MORRIS

Professor of Mathematics and Computer Science

The University of Massachusetts at Boston

100 Morrissey Blvd.

Boston, MA 02125-3393

Phone: 617-287-6499; email: ram@cs.umb.edu; fax: 617-287-6499

Education

Ph.D., Mathematics, Cornell University, 1970

M.A., Mathematics, Cornell University, 1967

B.A., Mathematics, Reed College, 1965

Appointments

May 1997-present Director, UMASS Software Engineering Research Laboratory

2001-present Professor of Computer Science (Math and CS departments separated)

1987-2001. Professor of Mathematics and Computer Science

1989 (Fall Quarter) Visiting Scholar, Eye Research Laboratories, University of Chicago

1978-1986 Associate Professor of Mathematics and Computer Science, UMASS Boston

1975-1978 Associate Professor of Mathematics, Univ. of Oklahoma

1973-1975 Member, School of Mathematics, The Institute For Advanced Study, Princeton

1971-1975 Assistant Professor of Mathematics, State University of New York at Albany

1969-1971 Instructor of Mathematics, State University of New York at Albany

Current Grant Support

Attached separately

Publications Especially Relevant to Project

R.D. Stevenson, R.A. Morris, and W. Haber, *Electronic Field Guides and User Communities in the Ecoinformatics Revolution*, submitted.

Morris, R.A., M. Passell and R.D. Stevenson. *itA Software Engineering Perspective on Developing Electronic Field Guides: Lessons Learned For Bioinformatics* , European Environmental Agency Technical Report Series

Robert A. Morris, Matthew Passell and Elvan Kerim, *The UMB Electronic Field Guide*, software system, 1998-2000. Described at <http://www.cs.umb.edu/efg>

Robert A. Morris, ,, *Fundamentals of Computer Image Processing*^{TMTM}, *Microscopic and Spectroscopic Imaging of the Chemical State*, Michael D. Morris, editor, Marcel Dekker, New York, 1993, pp. 109-130.

R. Morris, E. Blachman, and C. Meyer, *ieA Constraint-based editor for linguistics scholars*^{TMTM}, *Electronic*

Publishing , *Origination, Dissemination and Design*, 6(4), 349-360. (Proceedings of the 1994 Conference on Electronic Publishing, Darmstadt, 1994), John Wiley & Sons. 1994

Other Publications

Robert A. Morris, Charles Meyer and Edward Blachman, ,, *Can you see whose speech is overlapping*^{TMTM},

Visible Language 28(2), 110-133, 1994

Robert A. Morris, *UMB-RSVP*, an object oriented software system for gathering reading rate data from human subjects, 1996-2000.

E. Blachman, C. Meyer and R. Morris, *ioThe UMB Intelligent ICE Markup Assistant*lt, in *Comparing English Worldwide*, Sidney Greenbaum, Ed., Oxford University Press, 1996, pp. 54-64

Robert A. Morris, *ioAn unorthodox approach to the teaching of undergraduate software engineering*le, *Computing Systems*, vol. 1(4), pp. 405-419, Fall 1988.

Kathy Aquilante, Dean Yager and Robert A. Morris, "Repeatability of Reading Rate Measurements with MNREAD, RSVP, and SCROLL text for low vision subjects with central visual field loss and age-matched normals," submitted to *Optometry and Vision Science*

Synergistic Activities

IT advisor to American Museum of Natural History; Extensive participation about XML in the Subgroup on Descriptive Data of the Taxonomic Data Working Group of IUBS; Extensive research in electronic document processing; Managed and contributed to major software projects; teach software engineering;

Collaborators and other affiliations

i. Collaborators

Katherine Aquilante, SUNY College of Optometry
James Beach, Kansas University Museum of Natural History
Alexander Coimbra, Ecole Polytech. Fed. Lausanne
William Haber, Missouri Botanical Gardens
Roger Hersch, Ecole Polytech. Fed. Lausanne
Tom Moritz, American Museum of Natural History
Matthew Passell, UMASS Boston
Hannu Saarenmaa, European Environmental Agency
Robert D. Stevenson, UMASS Boston
Fred St. Ours, UMASS Boston
David Vieglais, Kansas University Museum of Natural History
Jun Wan, UMASS Boston
Harry Wyatt, SUNY College of Optometry
Dean Yager, SUNY College of Optometry

ii. Graduate and Postdoctoral Advisors.

Alex Rosenberg, Professor Emeritus, U. California at Santa Barbara.

iii. Thesis Advisor and Postgraduate-Scholar Sponsor.

No Ph.D. thesis or post-doctoral students. Approximately 20 masterTMs students supervised in major software engineering projects.

iv. Research Assistants:

Matthew Passell, Jun Wan, Henry Mugasha, Niti Upadhyay, Naibing Du, Hui Dong, Hua Tang, Yun Zhang, Elvan Kerim, Yixin Yu, Qian Zhou

Biographical Sketch
Norman F. Johnson

a. Professional Preparation

State University of New York College of Environmental Science & Forestry	Forest Entomology	B.S., 1976
---	-------------------	------------

Cornell University	Entomology	Ph.D., 1981
--------------------	------------	-------------

b. Appointments

1992-present	Director, Ohio State University Insect Collection
1987-present	Associate Professor, Department of Entomology, Ohio State University
1981-1987	Assistant Professor, Department of Entomology, Ohio State University

c. Publications

(i) A list of up to 5 publications most closely related to the proposed project.

- Johnson, N. F. and L. Musetti. 2002. Rediscovery of the genus *Platyscelidris* Szabó (Hymenoptera: Scelionidae) and description of a new species. *Proceedings of the Entomological Society of Washington*, accepted for publication.
- Johnson, N. F., L. Musetti, and J.-W. Janzen. 2001. A new fossil species of the Australian endemic genus *Peradenia* Naumann & Masner (Hymenoptera: Proctotrupeoidea, Peradeniidae) from Baltic amber. *Journal of Insect Systematics and Evolution* **32**:191-194.
- Dotseth, E. J. and N. F. Johnson. 2001. The Neotropical genus *Acanthoscelio* Ashmead (Hymenoptera: Scelionidae). *The Canadian Entomologist* **133**: 487-507.
- Johnson, N. F. and L. Musetti. 2000. Data warehousing architecture and tools for Hymenoptera biodiversity informatics. Pages 313-319, in A. D. Austin and M. Dowton, eds., *Hymenoptera: evolution, biodiversity and biological control*. CSIRO Publishing, Collingwood, Victoria, Australia. 468 pp.
- Johnson, N. F. and L. Musetti. 1999. Revision of the proctotrupoid genus *Pelecinus* (Hymenoptera: Pelecinidae). *Journal of Natural History* **33**:1513-1543.

(ii) A list of up to 5 other significant publications.

- Musetti, L. and N. F. Johnson. 2000. First documented record of Monomachidae in New Guinea, and description of two new species (Hymenoptera: Proctotrupeoidea). *Proceedings of the Entomological Society of Washington* **102**: 957-963.
- Johnson, N. F. 2000. A new species and revision of the concept of the genus *Thoronidea* Masner & Huggert (Hymenoptera, Platygastroidea: Scelionidae). *The Canadian Entomologist* **132**:301-305.
- Johnson, N. F., L. Musetti, J. B. Johnson, and K. Katovich. 1999. The larva of *Pelecinus polyturator* (Drury) (Hymenoptera: Pelecinidae). *Proceedings of the Entomological Society of Washington* **101**:64-68.
- Johnson, N. F. and L. Musetti. 1998. Geographic variation of sex ratio in *Pelecinus polyturator* (Drury) (Hymenoptera: Pelecinidae). *Journal of Hymenoptera Research* **7**:48-56.

Johnson, N. F. 1998. The fossil peleciniids *Pelecinopteron* Brues and *Iscopinus* Kozlov (Hymenoptera: Proctotrupoidea, Peleciniidae). *Proceedings of the Entomological Society of Washington* **100**:1-6.

d. Synergistic Activities.

Hymenoptera On-Line: This is an Oracle database storing information on the insect order Hymenoptera (ants, bees, wasps, etc.) derived from both literature and specimens from natural history collections. The Web interface allows retrieval of data in the form of classifications, systematic catalogs, distribution maps, graphs of phenology, as well as links to other on-line databases (e.g., GenBank). URL: <http://iris.biosci.ohio-state.edu/hymenoptera>

Hymenoptera Name Server: Modeled after the numerous geographic name servers, this Web interface provides information on the status, classification, and synonyms for the names used for the insect order Hymenoptera. URL: http://atbi.biosci.ohio-state.edu:210/hymenoptera/nomenclator.home_page

ATBI Information System: A database for collection and observation information derived from the All Taxa Biodiversity Inventory of the Great Smoky Mountains National Park. In development.

Virgin Islands Beetle Fauna: Web gateway to database storing results of a biotic survey of the Coleoptera of the U.S. and British Virgin Islands; in collaboration with M. A. Ivie, Montana State University. URL: http://iris.biosci.ohio-state.edu/vi_beetles

Leaf Litter Ant Database: Database for retrieval of collection data derived from controlled sampling regime for ants from around the world. Data available from Brazil, Africa, Malaysia and datasets compatible for input into EstimateS analysis package; created in collaboration with Dr. D. Agosti, American Museum of Natural History. URL: http://atbi.biosci.ohio-state.edu:898/docs/litter_data

e. Collaborators & Other Affiliations.

(i) Collaborators: Eric J. Dotseth (Ohio State), Flávia Ejchel, Michael A. Ivie (Montana State University), James B. Johnson (University of Idaho), Kerry Katovich (University of Wisconsin), Peter W. Kovarik (Ohio State), Luciana Musetti (Ohio State), V. D. ODDRaju (Florida A&M University), Paul A. Skelley (University of Florida)

(ii) Graduate and Postdoctoral Advisors: William L. Brown, Jr. (Cornell University, deceased).

(iii) Thesis Advisor and Postgraduate-Scholar Sponsor. Students: Alexandre Pires Aguiar (Museu de Zoologia, Universidade de São Paulo), Eric J. Dotseth (current), Flávia Ejchel, Peter W. Kovarik (Ohio State), Chi-Feng Lee (current), T. Keith Philips (Western Kentucky University), David Rosenthal (current), Hojun Song (current). Postdoctoral: Luciana Musetti, James B. Whitfield (University of Illinois).

SUMMARY PROPOSAL BUDGET YEAR 1

ORGANIZATION American Museum Natural History				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Thomas Moritz				AWARD NO.	Proposed	Granted	
				A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)			
	CAL	ACAD	SUMR				
1. Thomas Moritz - Director	0.00	0.00	0.00	\$ 0		\$	
2.							
3.							
4.							
5.							
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00		0		
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00		0		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. (0) POST DOCTORAL ASSOCIATES	0.00	0.00	0.00		0		
2. (2) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	18.00	0.00	0.00		90,600		
3. (0) GRADUATE STUDENTS					0		
4. (0) UNDERGRADUATE STUDENTS					0		
5. (1) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					12,000		
6. (0) OTHER					0		
TOTAL SALARIES AND WAGES (A + B)					102,600		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					27,702		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					130,302		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
SAN storage				\$ 20,000			
TOTAL EQUIPMENT					20,000		
E. TRAVEL							
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					3,150		
2. FOREIGN					9,750		
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____					0		
2. TRAVEL _____					0		
3. SUBSISTENCE _____					0		
4. OTHER _____					0		
TOTAL NUMBER OF PARTICIPANTS (0)				TOTAL PARTICIPANT COSTS		0	
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					0		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					0		
3. CONSULTANT SERVICES					15,000		
4. COMPUTER SERVICES					66,000		
5. SUBAWARDS					55,688		
6. OTHER					0		
TOTAL OTHER DIRECT COSTS					136,688		
H. TOTAL DIRECT COSTS (A THROUGH G)					299,890		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
MTDC (Rate: 59.7400, Base: 224202)							
TOTAL INDIRECT COSTS (F&A)					133,938		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					433,828		
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)					0		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					\$ 433,828	\$	
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME Thomas Moritz				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

1 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET YEAR 2

ORGANIZATION American Museum Natural History				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Thomas Moritz				AWARD NO.	Proposed	Granted
					NSF Funded Person-mos.	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. Thomas Moritz - Director				0.00	0.00	0.00
2.						
3.						
4.						
5.						
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	0.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. (0) POST DOCTORAL ASSOCIATES				0.00	0.00	0.00
2. (1) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				3.00	0.00	0.00
3. (0) GRADUATE STUDENTS						0
4. (0) UNDERGRADUATE STUDENTS						0
5. (1) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						12,600
6. (0) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						28,665
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						7,740
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						36,405
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						3,150
2. FOREIGN						9,750
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						15,000
4. COMPUTER SERVICES						0
5. SUBAWARDS						55,892
6. OTHER						0
TOTAL OTHER DIRECT COSTS						70,892
H. TOTAL DIRECT COSTS (A THROUGH G)						120,197
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 59.7400, Base: 64305)						
TOTAL INDIRECT COSTS (F&A)						38,416
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						158,613
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$ 158,613 \$
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PD NAME Thomas Moritz				FOR NSF USE ONLY		
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION		
		Date Checked	Date Of Rate Sheet	Initials - ORG		

2 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET YEAR 3

ORGANIZATION American Museum Natural History				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Thomas Moritz				AWARD NO.	Proposed	Granted
					NSF Funded Person-mos.	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. Thomas Moritz - Director				0.00	0.00	0.00
2.						
3.						
4.						
5.						
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	0.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. (0) POST DOCTORAL ASSOCIATES				0.00	0.00	0.00
2. (1) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				3.00	0.00	0.00
3. (0) GRADUATE STUDENTS						0
4. (0) UNDERGRADUATE STUDENTS						0
5. (1) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						13,230
6. (0) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						30,098
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						8,127
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						38,225
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						3,150
2. FOREIGN						9,750
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						15,000
4. COMPUTER SERVICES						0
5. SUBAWARDS						58,207
6. OTHER						0
TOTAL OTHER DIRECT COSTS						73,207
H. TOTAL DIRECT COSTS (A THROUGH G)						124,332
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 59.7400, Base: 66125)						
TOTAL INDIRECT COSTS (F&A)						39,503
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						163,835
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$ 163,835 \$
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PD NAME Thomas Moritz				FOR NSF USE ONLY		
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION		
		Date Checked	Date Of Rate Sheet	Initials - ORG		

3 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET Cumulative

ORGANIZATION American Museum Natural History				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Thomas Moritz				AWARD NO.	Proposed	Granted
					NSF Funded Person-mos.	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. Thomas Moritz - Director				0.00	0.00	0.00
2.						
3.						
4.						
5.						
6. () OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	0.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. (0) POST DOCTORAL ASSOCIATES				0.00	0.00	0.00
2. (4) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				24.00	0.00	0.00
3. (0) GRADUATE STUDENTS						0
4. (0) UNDERGRADUATE STUDENTS						0
5. (3) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						37,830
6. (0) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						161,363
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						43,569
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						204,932
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
\$ 20,000						
TOTAL EQUIPMENT						20,000
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						9,450
2. FOREIGN						29,250
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						45,000
4. COMPUTER SERVICES						66,000
5. SUBAWARDS						169,787
6. OTHER						0
TOTAL OTHER DIRECT COSTS						280,787
H. TOTAL DIRECT COSTS (A THROUGH G)						544,419
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS (F&A)						211,857
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						756,276
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$ 756,276 \$
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PD NAME Thomas Moritz				FOR NSF USE ONLY		
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION		
				Date Checked	Date Of Rate Sheet	Initials - ORG

C *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

Budget Justification

Personnel

The project needs a PERL programmer for 1 year to take the text and iteratively refine the hypothesized *Taxon-X* schema as guide, and then develop and test to produce an optimized, automated mark-up process. The project needs 50% of the Database Manager's time in the first year to define the XML mark-up standards and manage the out-sourced digitization of the material. In the second and third years, the project needs 25% of the Database Manager's time to work with the programmers to develop the applications which use the XML.

The project needs a half-time clerical position to perform licensing requests, data entry, data modification, data scrubbing, and data quality control.

Equipment

SAN: The project will require an estimated one terabyte disk drive space which will be added onto our museum infrastructure.

Travel

The project estimates one meeting per year at each collaborating institution. This means one trip to Germany for three people and one trip to Ohio State University for three people. The trip to Germany will be at the middle of the first year to evaluate work done to date and solidify the future course of the collaboration.

Consultants

The project needs the consulting services of Donat Agosti in the capacity as expert entomologist with information management expertise. In addition, he will function as Project Coordinator between Germany and the United States. He is Research Associate at the AMNH and is based in Switzerland. Given his history of involvement with both institutions, he is ideally suited for this position.

Computer Services

Digitization: The initial digitization of all requisite material will be outsourced to a vendor at approximately \$3/page (based on extensive AMNH experience with vendors) for 22,000 pages. (This work is best outsourced to a vendor who has the equipment and staff with the expertise to massively produce this work.)

SUMMARY PROPOSAL BUDGET YEAR 1

ORGANIZATION University of Massachusetts Boston				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Robert Morris				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. Robert Morris	0.00	0.00	0.00	\$ 0		\$	
2.							
3.							
4.							
5.							
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00		0		
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00		0		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. (0) POST DOCTORAL ASSOCIATES	0.00	0.00	0.00		0		
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00		0		
3. (1) GRADUATE STUDENTS					25,000		
4. (0) UNDERGRADUATE STUDENTS					0		
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					0		
6. (0) OTHER					0		
TOTAL SALARIES AND WAGES (A + B)					25,000		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					363		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					25,363		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
Workstation				\$ 2,000			
TOTAL EQUIPMENT					2,000		
E. TRAVEL					5,600		
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							
2. FOREIGN					3,500		
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____			0				
2. TRAVEL _____			0				
3. SUBSISTENCE _____			0				
4. OTHER _____			0				
TOTAL NUMBER OF PARTICIPANTS (0)				TOTAL PARTICIPANT COSTS	0		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					0		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					0		
3. CONSULTANT SERVICES					0		
4. COMPUTER SERVICES					0		
5. SUBAWARDS					0		
6. OTHER					500		
TOTAL OTHER DIRECT COSTS					500		
H. TOTAL DIRECT COSTS (A THROUGH G)					36,963		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
Wage&salary (Rate: 74.9000, Base: 25000)							
TOTAL INDIRECT COSTS (F&A)					18,725		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					55,688		
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)					0		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					\$ 55,688	\$	
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME Robert Morris				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

1 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET YEAR 2

ORGANIZATION University of Massachusetts Boston				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Robert Morris				AWARD NO.			
				Proposed	Granted		
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer	Funds granted by NSF (if different)
				CAL	ACAD	SUMR	
1. Robert Morris				0.00	0.00	0.00	\$ 0 \$
2.							
3.							
4.							
5.							
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	0
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	0.00	0
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. (0) POST DOCTORAL ASSOCIATES				0.00	0.00	0.00	0
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00	0
3. (1) GRADUATE STUDENTS							26,250
4. (0) UNDERGRADUATE STUDENTS							0
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							0
6. (0) OTHER							0
TOTAL SALARIES AND WAGES (A + B)							26,250
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							381
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							26,631
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							0
E. TRAVEL							
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							5,600
2. FOREIGN							3,500
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____				0			
2. TRAVEL _____				0			
3. SUBSISTENCE _____				0			
4. OTHER _____				0			
TOTAL NUMBER OF PARTICIPANTS (0)							
TOTAL PARTICIPANT COSTS							0
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							0
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							0
3. CONSULTANT SERVICES							0
4. COMPUTER SERVICES							0
5. SUBAWARDS							0
6. OTHER							500
TOTAL OTHER DIRECT COSTS							500
H. TOTAL DIRECT COSTS (A THROUGH G)							36,231
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
wage&salary (Rate: 74.9000, Base: 26250)							
TOTAL INDIRECT COSTS (F&A)							19,661
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							55,892
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)							0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							\$ 55,892 \$
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME Robert Morris				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
		Date Checked		Date Of Rate Sheet		Initials - ORG	

2 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET YEAR 3

ORGANIZATION University of Massachusetts Boston				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Robert Morris				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. Robert Morris	0.00	0.00	0.00	\$ 0		\$	
2.							
3.							
4.							
5.							
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00		0		
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00		0		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. (0) POST DOCTORAL ASSOCIATES	0.00	0.00	0.00		0		
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00		0		
3. (1) GRADUATE STUDENTS					27,563		
4. (0) UNDERGRADUATE STUDENTS					0		
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					0		
6. (0) OTHER					0		
TOTAL SALARIES AND WAGES (A + B)					27,563		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					400		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					27,963		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT					0		
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					5,600		
2. FOREIGN					3,500		
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____			0				
2. TRAVEL _____			0				
3. SUBSISTENCE _____			0				
4. OTHER _____			0				
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS					0		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					0		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					0		
3. CONSULTANT SERVICES					0		
4. COMPUTER SERVICES					0		
5. SUBAWARDS					0		
6. OTHER					500		
TOTAL OTHER DIRECT COSTS					500		
H. TOTAL DIRECT COSTS (A THROUGH G)					37,563		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) wage&salary (Rate: 74.9000, Base: 27563)							
TOTAL INDIRECT COSTS (F&A)					20,645		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					58,208		
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)					0		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				\$	58,208	\$	
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME Robert Morris				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
		Date Checked	Date Of Rate Sheet	Initials - ORG			

3 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET Cumulative

ORGANIZATION University of Massachusetts Boston				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Robert Morris				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. Robert Morris	0.00	0.00	0.00	\$	0	\$	
2.							
3.							
4.							
5.							
6. () OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00		0		
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00		0		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. (0) POST DOCTORAL ASSOCIATES	0.00	0.00	0.00		0		
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00		0		
3. (3) GRADUATE STUDENTS					78,813		
4. (0) UNDERGRADUATE STUDENTS					0		
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					0		
6. (0) OTHER					0		
TOTAL SALARIES AND WAGES (A + B)					78,813		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					1,144		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					79,957		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
			\$ 2,000				
TOTAL EQUIPMENT					2,000		
E. TRAVEL					16,800		
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					16,800		
2. FOREIGN					10,500		
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____			0				
2. TRAVEL _____			0				
3. SUBSISTENCE _____			0				
4. OTHER _____			0				
TOTAL NUMBER OF PARTICIPANTS (0)				TOTAL PARTICIPANT COSTS	0		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					0		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					0		
3. CONSULTANT SERVICES					0		
4. COMPUTER SERVICES					0		
5. SUBAWARDS					0		
6. OTHER					1,500		
TOTAL OTHER DIRECT COSTS					1,500		
H. TOTAL DIRECT COSTS (A THROUGH G)					110,757		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
TOTAL INDIRECT COSTS (F&A)					59,031		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					169,788		
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)					0		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					\$ 169,788	\$	
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME Robert Morris				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
				Date Checked	Date Of Rate Sheet	Initials - ORG	

C *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET YEAR 1

ORGANIZATION Ohio State University Research Foundation				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Norman F Johnson				Proposed	Granted		
				AWARD NO.			
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR				
1. Norman F Johnson - Co-PI	0.00	0.00	1.00	\$ 7,779		\$	
2.							
3.							
4.							
5.							
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	0			
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	1.00	7,779			
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. (0) POST DOCTORAL ASSOCIATES	0.00	0.00	0.00	0			
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00	0			
3. (0) GRADUATE STUDENTS				0			
4. (0) UNDERGRADUATE STUDENTS				0			
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0			
6. (0) OTHER				0			
TOTAL SALARIES AND WAGES (A + B)					7,779		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					1,284		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					9,063		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT					0		
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)					0		
2. FOREIGN					0		
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____				0			
2. TRAVEL _____				0			
3. SUBSISTENCE _____				0			
4. OTHER _____				0			
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS					0		
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES					2,000		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					0		
3. CONSULTANT SERVICES					0		
4. COMPUTER SERVICES					0		
5. SUBAWARDS					0		
6. OTHER					0		
TOTAL OTHER DIRECT COSTS					2,000		
H. TOTAL DIRECT COSTS (A THROUGH G)					11,063		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 47.5000, Base: 11063)							
TOTAL INDIRECT COSTS (F&A)					5,255		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					16,318		
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)					0		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)					\$ 16,318	\$	
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME Norman F Johnson				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
		Date Checked		Date Of Rate Sheet		Initials - ORG	

1 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET YEAR 2

ORGANIZATION Ohio State University Research Foundation				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Norman F Johnson				AWARD NO.			
				Proposed	Granted		
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer	Funds granted by NSF (if different)
				CAL	ACAD	SUMR	
1. Norman F Johnson - none				0.00	0.00	1.00	\$ 8,090
2.							
3.							
4.							
5.							
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	0
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00	8,090
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. (0) POST DOCTORAL ASSOCIATES				0.00	0.00	0.00	0
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00	0
3. (0) GRADUATE STUDENTS							0
4. (0) UNDERGRADUATE STUDENTS							0
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							0
6. (0) OTHER							0
TOTAL SALARIES AND WAGES (A + B)							8,090
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							1,335
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							9,425
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							0
E. TRAVEL							0
1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							0
2. FOREIGN							0
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____				0			
2. TRAVEL _____				0			
3. SUBSISTENCE _____				0			
4. OTHER _____				0			
TOTAL NUMBER OF PARTICIPANTS (0)							
TOTAL PARTICIPANT COSTS							0
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							2,000
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							0
3. CONSULTANT SERVICES							0
4. COMPUTER SERVICES							0
5. SUBAWARDS							0
6. OTHER							0
TOTAL OTHER DIRECT COSTS							2,000
H. TOTAL DIRECT COSTS (A THROUGH G)							11,425
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
MTDC (Rate: 47.5000, Base: 11425)							
TOTAL INDIRECT COSTS (F&A)							5,427
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							16,852
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)							0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							\$ 16,852
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PD NAME Norman F Johnson				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
				Date Checked	Date Of Rate Sheet	Initials - ORG	

2 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET YEAR 3

ORGANIZATION Ohio State University Research Foundation				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Norman F Johnson				AWARD NO.	Proposed	Granted
					NSF Funded Person-mos.	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. Norman F Johnson - none				0.00	0.00	1.00
2.						
3.						
4.						
5.						
6. (0) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	1.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. (0) POST DOCTORAL ASSOCIATES				0.00	0.00	0.00
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00
3. (0) GRADUATE STUDENTS						0
4. (0) UNDERGRADUATE STUDENTS						0
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						0
6. (0) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						8,414
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						1,388
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						9,802
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						0
2. FOREIGN						0
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						2,000
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						0
4. COMPUTER SERVICES						0
5. SUBAWARDS						0
6. OTHER						0
TOTAL OTHER DIRECT COSTS						2,000
H. TOTAL DIRECT COSTS (A THROUGH G)						11,802
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 47.5000, Base: 11802)						
TOTAL INDIRECT COSTS (F&A)						5,606
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						17,408
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$ 17,408 \$
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PD NAME Norman F Johnson				FOR NSF USE ONLY		
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION		
				Date Checked	Date Of Rate Sheet	Initials - ORG

3 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET Cumulative

ORGANIZATION Ohio State University Research Foundation				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Norman F Johnson				AWARD NO.	Proposed	Granted
					NSF Funded Person-mos.	
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. Norman F Johnson - none				0.00	0.00	3.00
2.						
3.						
4.						
5.						
6. () OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				0.00	0.00	3.00
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. (0) POST DOCTORAL ASSOCIATES				0.00	0.00	0.00
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.00	0.00	0.00
3. (0) GRADUATE STUDENTS						0
4. (0) UNDERGRADUATE STUDENTS						0
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						0
6. (0) OTHER						0
TOTAL SALARIES AND WAGES (A + B)						24,283
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						4,007
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						28,290
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						0
2. FOREIGN						0
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____				0		
2. TRAVEL _____				0		
3. SUBSISTENCE _____				0		
4. OTHER _____				0		
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS						0
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						6,000
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						0
3. CONSULTANT SERVICES						0
4. COMPUTER SERVICES						0
5. SUBAWARDS						0
6. OTHER						0
TOTAL OTHER DIRECT COSTS						6,000
H. TOTAL DIRECT COSTS (A THROUGH G)						34,290
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS (F&A)						16,288
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						50,578
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.C.6.j.)						0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$ 50,578 \$
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$		
PI/PD NAME Norman F Johnson				FOR NSF USE ONLY		
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION		
				Date Checked	Date Of Rate Sheet	Initials - ORG

C *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

Budget Justification Page

One month summer salary is requested for the Co-PI for each year; fringe benefit rates for off-quarter salary calculated at 16.5%. Materials and supplies costs are for license fees for Oracle database software.

Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.	
Investigator: Thomas Moritz	Other agencies (including NSF) to which this proposal has been/will be submitted.
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Creating a Model for Web-Based Sytematic Catalogs: Using the Heteroptera (Insecta) as Proof of Concept	
Source of Support: NSF Total Award Amount: \$ 843,956 Total Award Period Covered: 01/01/03 - 12/31/06 Location of Project: AMNH Person-Months Per Year Committed to the Project. Cal: 0.50 Acad: 0.00 Sumr: 0.00	
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Digitizing Unique Holdings within the AMNH Library	
Source of Support: Andrew A. Mellon Foundation Total Award Amount: \$ 2,000,000 Total Award Period Covered: 07/01/02 - 12/31/03 Location of Project: AMNH Person-Months Per Year Committed to the Project. Cal: 4.00 Acad: 0.00 Sumr: 0.00	
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Collaborative Research: Development of new digital library applications in the context of a basic ontology for biosystematics information using the literature of	
Source of Support: NSF Total Award Amount: \$ 758,679 Total Award Period Covered: 05/01/03 - 04/30/06 Location of Project: AMNH Person-Months Per Year Committed to the Project. Cal: 2.40 Acad: 0.00 Sumr: 0.00	
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title:	
Source of Support: Total Award Amount: \$ Total Award Period Covered: Location of Project: Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:	
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title:	
Source of Support: Total Award Amount: \$ Total Award Period Covered: Location of Project: Person-Months Per Year Committed to the Project. Cal: Acad: Summ:	

*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.

Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.	
Investigator: Norman Johnson	Other agencies (including NSF) to which this proposal has been/will be submitted.
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Parnerships in Enhancing Expertise in Taxonomy: Parasitic Wasps of the Family Scelionidae (Hymenoptera) - Supplement.	
Source of Support: National Science Foundation Total Award Amount: \$ 33,106 Total Award Period Covered: 09/15/95 - 08/31/02 Location of Project: Columbus, OH Person-Months Per Year Committed to the Project. Cal: 0.00 Acad: 0.90 Sumr: 1.00	
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: An Information System for the Great Smoky Mountains National Park ATBI	
Source of Support: National Science Foundation Total Award Amount: \$ 274,774 Total Award Period Covered: 10/01/99 - 09/30/02 Location of Project: Columbus, OH Person-Months Per Year Committed to the Project. Cal: 0.00 Acad: 0.90 Sumr: 1.00	
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: A Specimen-Level Database of the Ohio State University Insect Collection: 1. Hymenoptera and Cicadellidae.	
Source of Support: National Science Foundation Total Award Amount: \$ 216,855 Total Award Period Covered: 01/01/03 - 12/31/04 Location of Project: Columbus, OH Person-Months Per Year Committed to the Project. Cal: 0.00 Acad: 0.90 Sumr: 1.00	
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Evolution of Host Relationships in the Parasitoid Family Scelionidae (Hymenoptera)	
Source of Support: National Science Foundation Total Award Amount: \$ 307,423 Total Award Period Covered: 01/01/03 - 12/31/05 Location of Project: Columbus, OH; Ottawa, Canada; Adelaide, Australia Person-Months Per Year Committed to the Project. Cal: 0.00 Acad: 0.90 Sumr: 1.00	
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: This proposal	
Source of Support: National Science Foundation Total Award Amount: \$ 47,628 Total Award Period Covered: 01/01/03 - 12/31/05 Location of Project: Columbus, OH; New York, NY; Magdeburg, Germany Person-Months Per Year Committed to the Project. Cal: 0.00 Acad: 0.45 Summ: 1.00	
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.	

FACILITIES, EQUIPMENT, AND OTHER RESOURCES

Facilities crucial to the implementation of this project are available at the American Museum of Natural History. These include:

- Internet I and II connected to a 100 megabit museum-wide network backbone.
- Two Unix-based servers (Sun E250, Sun Netra T1) with sufficient capacity to handle XML markup, PERL scripting, and database development.
- SAN (Storage Area Network) Infrastructure in place although more drive space will be required to accommodate this new application.
- Library Services NT based Local area network available for data entry and user testing.
- Library Services resources capable of supplying 90% of material to be digitized.
- Digital Lab housed in library with the following camera equipment:

FACILITIES, EQUIPMENT & OTHER RESOURCES

FACILITIES: Identify the facilities to be used at each performance site listed and, as appropriate, indicate their capacities, pertinent capabilities, relative proximity, and extent of availability to the project. Use "Other" to describe the facilities at any other performance sites listed and at sites for field studies. USE additional pages as necessary.

Laboratory:

Clinical:

Animal:

Computer: Sun 450 server with Oracle 8.1.7 and WebDB installed; 5 PCs for data input.

Office: 600 sq. ft. office/lab; 100 sq. ft. office/lab

Other: _____

MAJOR EQUIPMENT: List the most important items available for this project and, as appropriate identifying the location and pertinent capabilities of each.

OTHER RESOURCES: Provide any information describing the other resources available for the project. Identify support services such as consultant, secretarial, machine shop, and electronics shop, and the extent to which they will be available for the project. Include an explanation of any consortium/contractual arrangements with other organizations.

X-Sender: nht@mailserver.nhm.ac.uk
X-Mailer: QUALCOMM Windows Eudora Version 5.1
Date: Tue, 30 Jul 2002 11:58:42 +0100
To: Tom Moritz <tmoritz@amnh.org>
From: Neil Thomson <n.thomson@nhm.ac.uk>
Subject: Re: Entomology Proposals...?
Cc: nht@nhm.ac.uk

The Natural History Museum in London is very pleased to support the joint US-German proposal made by the American Museum of Natural History in New York, Ohio State University and Otto-von-Guericke-Universitat to develop a much-needed framework for managing digitised biosystematics literature.

The printed literature of systematics is vital for the understanding of the natural world and underpins all biological disciplines and biodiversity research - yet it is often difficult to access. The techniques that will be developed through this proposal will not only make this literature freely available to all, but will be a huge step towards making the research task considerably more efficient, speeding up the efforts to conserve existing biodiversity.

Neil Thomson
Head of Systems & Central Services
Natural History Museum, Cromwell Road, London SW7 5BD
Tel: 0207 942 5294, Fax: 0207 942 5559, Email: n.thomson@nhm.ac.uk
<http://www.nhm.ac.uk>

Tom Moritz
Boeschstein Director, Library
American Museum of Natural History
79th St. @ Central Park West
New York, New York 10024
USA

Dear Dr. Moritz:

We are pleased to support the joint US-German proposal by AMNH Ohio State University and Otto-von-Guericke-Universität to develop a comprehensive digital framework for the management of the literature of biosystematics.

The fundamental goal of providing extended worldwide access to this literature is, in itself, of real value. Moreover, the research components of the proposal promise to develop and test new techniques for the extraction and management of data latent in the extended legacy of systematics literature. This effort can have value for all biological disciplines, beyond the subject scope of the proposal itself.

We believe this work can make a key contribution to fundamental biosystematic research and can be instrumental in supporting the worldwide effort to conserve biological diversity.

We are especially pleased with the project's potential to enhance the capability of the Integrated Taxonomic Information System to link scientific names to bibliographic sources.

Sincerely,

Michael A. Ruggiero, Ph. D.
Director, Integrated Taxonomic Information System
Smithsonian Institution
P.O. Box 37012
NMNH, Room CE-120, MRC - 0180
Washington, DC 20013-7012

phone: 202-786-3117
fax: 202-786-2934

-----Ursprüngliche Nachricht-----

Von: Edwards, James [mailto:JEdwards@zmuc.ku.dk]

Gesendet: Friday, July 26, 2002 6:53 PM

An: 'agosti'

Betreff: RE: digitizing systematics literature, support from GBIF

I am pleased to provide the following statement:

Dear Colleagues

On behalf of the Global Biodiversity Information Facility (GBIF), I am pleased to state that the proposal being submitted by the American Museum of Natural History, Ohio State University and Otto-von-Guerick-Universität appears to be an exciting development that will be of great value to biodiversity research at large and to GBIF in particular. The proposed project will develop a comprehensive digital framework for the management of the literature of biosystematics. The fundamental goal of providing extended worldwide access to this literature is, in itself, of real value.

This effort will also have value for all biological disciplines, beyond the subject scope of the proposal itself. Techniques for developing digital libraries of biological diversity are an area of great need, as are robust models for extracting information in order to develop such additional products as gazetteers and thesauri. In addition, an extremely positive component of the project is that all data developed through it will be in the public domain.

Although GBIF is committed in the long term to producing a digital library of biodiversity information, in the short term it is focusing on other activities, specifically on completing the authority file of the approximately 1.7 million scientifically valid species (including synonyms) and on digitizing specimen data from natural history collections. Thus, the project being proposed by US-German consortium is in no way duplicative with GBIF's current work program. Instead, it will provide a sound base for GBIF's (and other organizations') future developments.

In short, I am convinced that the proposed work will make a key contribution to fundamental biosystematic research and will also be instrumental in supporting the world wide effort to conserve biological diversity,

Sincerely,

Dr. James L. Edwards
Executive Secretary
Global Biodiversity Information Facility
Zoological Museum
University of Copenhagen
Universitetsparken 15
DK-2100 Copenhagen Ø
DENMARK

Phone: +45 3532 1470
Fax: +45 3532 1480
E-mail: jedwards@zmuc.ku.dk

Dear colleagues:

I am very glad about the proposed project to put the whole body of ant systematic literature on the web and to make it even more valuable with an intelligent search engine. This aspect is fully in line with the German GBIF program, in which I will participate to bring all ant types stored in German collections to the web with a citation and the full original description. Thus the ant literature can act as an example for further projects on the world taxonomic literature of other animal and plant groups, as the technique you develop will be independent of the taxon.

I therefore fully support your idea and will be glad if I could bring in my expertise on ants in your project where ever it will be needed. Second, the State Museum of Natural History in Karlsruhe will be interested to host a mirror site of the proposed literature database and the search engines, once they are set up, and when our own IT-system with Internet server is fully functioning.

Manfred Verhaagh
Curator and current Head of Entomology
Staatl. Museum für Naturkunde Karlsruhe
Erbprinzenstr. 13
76133 Karlsruhe
Tel.: 0721-1752877
manfred.verhaagh@smnk.de

Date: Mon, 29 Jul 2002 11:20:13 -0400
From: "Robert A. Morris" <Robert.Morris@cs.umb.edu>
Reply-To: ram@cs.umb.edu
X-Mailer: Mozilla 4.73 [en]C-CCK-MCD {Sony} (Win98; U)
X-Accept-Language: en
To: tmoritz@amnh.org
Subject: Cooperation

Tom Moritz
Boeschstein Director, Library Services
American Museum of Natural History
79th St. @ Central Park West
New York, New York 10024

Dear Tom:

I've read your proposal to provide important ant literature in digital form. My lab is pleased to join this effort by building tools that will produce identification software automatically from such a corpus. We are especially interested to exercise our field guide generation software in this connection, because all our efforts to date involve creating XML from data drawn from relational databases across an ODBC bridge. This leaves us with more control than working from native XML not of our own devising. We see this as an opportunity to push our own work into areas that might prove useful both against native XML stores and against Web Services that offer XML.

Yours truly

Robert A. Morris
Professor of Computer Science
University of Massachusetts at Boston
100 Morrissey Blvd.
Boston, MA 02125-3393