

A Workshop on the Compilation, Maintenance, and Dissemination of Taxonomic Authority Files¹

A. Project Summary

Biological taxonomy plays a fundamental role in our understanding of biological systems at all levels of organization, from molecular biology to biogeography. Biological nomenclature and classification are particularly important to the management of biological information as taxonomic names are among the most commonly used keys for retrieving, sorting and summarizing data. A taxonomic name, typically a species name, is attached to every primary observation or specimen in any database that contains data about more than one species.

Establishing and maintaining nomenclatural integrity within a single database is typically achieved with authority files, sometimes called “look-up tables”. Establishing and maintaining this same type of consistency across a federation of databases requires an analogous authority file, created for and deployed across all participating databases. The authority file establishes a single copy for any piece of taxonomic information, to which all references point, or from which all copies and updates flow. The challenge to the systematics community and biological information managers is to develop mechanisms that facilitate the capture and reconciliation of historical literature, as well as the flow of current taxonomic research, into a coherent authority file, and then the deployment of authority control across all relevant databases. Ultimately dissemination and authority control should extend directly into the processes whereby taxonomic names are applied to observations and specimen records.

The library science community has considerable experience in developing and deploying community-wide authority control for cataloging library contents. Examples relevant to this broad topic include: subject and name authorities, discipline-specific thesauri, and programs to promote cooperative cataloging of shared objects and concepts. The similarities between what the library science community does now and what the systematic community would like to do in the future provide the motivation for this workshop. It will provide the library science and systematics communities with an opportunity to exchange descriptions their respective problem domains, goals, and methodologies.

The proposed two-day workshop will be held June 22-23rd, 1998, and will include experts in library science, representatives from taxonomic database projects, and data managers from disciplines where taxonomic authority control is desired. The format will consist of presentations by invited speakers, with opportunities for questions and general discussion at the end of each presentation and session, respectively. Invited papers and summaries of the discussion will be compiled into a proceedings that will be published electronically (i.e., as a series of electronic documents made available via the Web).

¹ After a proposal to the National Science Foundation; Stanley D. Blum, P.I. (funded as DEB 9726045). Modified from original submission. Last updated June 18, 1998.

B. Table of Contents

A. PROJECT SUMMARY.....	A-1
B. TABLE OF CONTENTS.....	B-1
C. PROJECT DESCRIPTION.....	C-1
1. The Role of Taxonomy in Biological Data Management.....	C-1
2. Taxonomic Database Projects.....	C-2
3. Community infrastructure for managing taxonomic authorities	C-5
4. Comparison with Authority Control in Library Cataloging	C-7
5. A Workshop on Taxonomic Authorities for Describing Biological Data.....	C-8
6. Impact on Scientific Infrastructure.....	C-9
D. REFERENCES CITED.....	D-1

C. Project Description

1. *The Role of Taxonomy in Biological Data Management*

Biological taxonomy plays a fundamental role in our understanding of biological systems at all levels of organization, from molecular biology to biogeography. At the organismal and suborganismal levels, phylogenetic relationships among taxa inform the comparative method and allow us to predict how broadly a newly discovered attribute might hold among species. At the supraorganismal level, species and higher taxa become the units of interest in the systems we study. Our ability to communicate precisely about the biological world, regardless of discipline, depends upon common definitions of taxa and the names we use to designate them. The users of basic taxonomic information extend through all of biology, and include the disciplines of health and medical sciences, biotechnology, agriculture, fisheries, forestry, conservation, land and water management, ecotourism, and all areas of basic biological research (Systematics Agenda 2000, 1994a, 1994b).

In biological data management, the role of taxonomy goes beyond basic communication. Taxonomic names are among the most commonly used keys for retrieving, sorting and summarizing biological information. A taxonomic name, typically a species name, is attached to every primary observation or specimen in any database that contains data about more than one species. Classifications aggregate species hierarchically into more inclusive taxa, and the names of higher taxa create an efficient means to retrieve larger sets of records. Thus, while the importance of systematics as the central organizing system in biology is often under-appreciated, the role of taxonomic nomenclature in databases -- as a logical storage and retrieval system -- is directly evident. Spelling errors and inconsistency in classification cause query results to be incomplete. From a user's perspective, such errors effectively represent the loss of data.

As biological data managers have come to appreciate the practical significance of taxonomic nomenclature as a logical data retrieval system, as well as its size and complexity, they have recognized that establishing and maintaining nomenclatural integrity within a single database requires the use of local authority files or look-up tables. Many large biological databases now include one or more authority files to control the assignment of taxonomic names to observations or specimens (i.e., a normalized representation of taxonomic names and relationships).

Taxonomic authority files require special effort to compile and maintain. Most organizations that have implemented authority files have compiled their own files independently of other organizations, or have begun by copying another organization's files and now maintain their own local files independently of the original source. Relatively little cooperation exists among organizations to develop or maintain authority files. This is not to dismiss the cooperation that has occurred, but only to point out that *ad hoc* or informal mechanisms for data sharing are neither sufficient to distribute authoritative nomenclature as widely as it is needed, nor sufficient to eliminate the following problems that arise from independent compilations of taxonomy and classification information:

- **Inefficiency:** Each nomenclatural compilation requires significant human resources. To the extent that any compilation overlaps another (i.e., covers the same taxa) it represents an inefficient expenditure of resources (unless it is done intentionally to create another classification).
- **Inaccuracy:** Accurate taxonomic nomenclature requires verification against the primary literature. Original publications extend nearly 250 years into the historical literature, and are scattered through hundreds of journals and thousands of monographs. Accurate interpretation of the original

literature requires extensive training in systematics. Few organizations have the resources to employ such expertise, and authority files compiled and maintained in its absence are typically inaccurate.

- **Inconsistency:** The existence of multiple and inaccurate compilations guarantees inconsistency among authority files. Different spellings will exist for what is intended to be the same name; different judgments will be made about which names (taxa) are valid and which are synonyms; and finally, different judgments will be made about the placement of taxa within higher taxa (i.e., will lead to different classifications, which make the names of higher taxa less useful as retrieval keys).

The most important impact of inconsistency among authority files becomes apparent from the perspective of a user attempting to integrate information across multiple sources. It has become apparent that the information needs of large and diverse scientific communities cannot be met by all-encompassing, monolithic information resources. Bio-informatics specialists have recognized that the more feasible strategy entails developing more specialized resources, and designing them such that information can be retrieved and combined across sources. Bio-informatics specialists have also recognized that this desired level of interoperability cannot be achieved unless concerted efforts are made to develop and maintain shared (standard) vocabularies and to resolve key semantic conflicts among overlapping data (information) models. For example, participants in the 1993 DOE genome informatics workshop concurred that:

“A truly federated information infrastructure cannot be achieved unless some minimum level of semantic consistency exists among the participating systems. No amount of syntactic connectivity can compensate for semantic mismatches. [...] In short, the development of a federated information infrastructure will require that more effort be made to ensure semantic consistency across participating systems. The use of controlled vocabularies and common-denominator semantics is important. Support for the necessary coordination and communication must be provided by DOE and other agencies.” (Robbins, 1994)

The prevalence of taxonomic nomenclature in biological databases and its potential role as a common logical data retrieval system make biological taxonomy an important area in need of standardization (or if not standardization, then even more extensive work on mapping taxonomic names between classifications). Without such efforts, a user will have to address each database on its own terms.

2. Taxonomic Database Projects

Within the systematics community, many projects are currently underway to compile comprehensive lists of taxonomic names, to sort out the valid names (taxa) from synonyms, and to provide a classification that summarizes the current state of systematic knowledge for that group. A list of examples is given in Table 1. Progress to date in several of these projects demonstrates that the goals are feasible. The number of existing and incipient projects also indicates that individuals and institutions in the community are willing to undertake and sustain such efforts. Awareness is growing that such databases will play an important role in making the primary product of taxonomic research more readily available to the broader scientific community. It is not unreasonable to expect that, within the foreseeable future, basic nomenclatural and classification information for all organisms known to science will exist somewhere in electronic form.

Table I. Examples of taxonomic compilation projects conducted by systematists.

Taxonomic Group / Project Name	Institution	P.I. or Contact
Recent Taxa:		
Mammals / Mammal Species of the World	Smithsonian	Wilson
Amphibians & Reptiles / Checklist of Amphibians and Reptiles of the World	BRD-Smithsonian, Am Mus Nat Hist, Villanova Univ.	McDiarmid, Frost, Bauer
Fishes / Catalog of the Genera and Species of Fishes	California Academy of Science	Eschmeyer
Arthropods of North America (excluding Crustacea) / BIOTA	USDA-Smithsonian	Thompson
Plants; Australia / Australian Plant Name Index (APNI)	ERIN, Australian National Botanic Garden	Chapman, Croft
Plants; Seed-bearing; New World / Gray Card Index	Harvard University Herbarium	Boufford
Plants; Flowering / Index Nominum Genericorum (genera only)	Smithsonian	Farr
Plants; North America / Biota of North America Program - BONAP	North Carolina Botanical Garden	Kartesz
Plants; North America / Flora of North America (FNA)	Missouri Botanical Garden, St. Louis	
Plants / Index Kewensis	Royal Botanical Garden, Kew	Oxford University Press
Fossil Taxa:		
Mammals / Fossil Mammals of the World	American Museum of Natural History	McKenna
Vertebrates / Bibliography of Fossil Vertebrates	UC Berkeley	Clemmens
Invertebrates / Treatise of Invertebrate Paleontology	KU - Paleo. Inst.	Kaesler
Plants / Plant Fossil Record Database	Univ. East London / Intl Org Palaeobotany	Boulter

At the international level, coordination among taxonomic database projects has been achieved primarily through the activities of three closely-related organizations:

1. Taxonomic Databases Working Group (TDWG)² – In 1985, TDWG was commissioned by the International Union of Biological Sciences (IUBS) to be a forum for developing and promulgating standards relevant to the development of taxonomic databases and the interchange of taxonomic data. TDWG's initial focus was on vascular plant taxa, but was expanded by IUBS in 1994 to cover all organisms.
2. International Organization for Plant Information (IOPI)³ – IOPI was founded in 1991 in response to calls issued at earlier TDWG meetings to create a Global Plant Species Information System (GPSIS) and a proposal from the Royal Botanic Garden, Kew to build a global taxonomic database for plants, the Species Plantarum Project. The main thrust of IOPI is now to develop a Global Plant Checklist
3. Species 2000⁴ – The Species 2000 Program was established in 1994 by IUBS, the Committee on Data for Science and Technology (CODATA) and the International Union of Microbiological Societies (IUMS). Its aims are to create a common access system on the Internet that enables a user to search all participating (federated) taxonomic databases for information on a given species, to publish a global species checklist annually, and to seek funding for taxonomic databases and coordinating infrastructure.

TDWG and IOPI meet annually, often jointly, and promote international participation by holding meetings around the world. TDWG has published a variety of standards, including content standards, data dictionaries, and data transfer standards.⁵ (A conceptual model of taxonomic information⁶ has been under development for several years, but is not yet listed as an accepted standard.) The IOPI Checklist Committee completed an extensive project plan (Wilson *et al.*, 1994), and recently received limited funding from private sources and CODATA to support implementation. A prototype of the Global Plant Checklist⁷ is now operating and will form part of the Species 2000 common access system. Species 2000 has developed a prototype of the common access system, but significant work remains to be done.

Two projects initiated in 1993 provide strong evidence, in the form of real personnel and monetary commitments, that the importance of taxonomic standardization in data management is recognized outside the systematics community. These are the GenBank taxonomy project at the National Center for Biotechnology Information (NIH-NLM-NCBI),⁸ and the Interagency Taxonomic Information System (ITIS⁹ – which is being hosted by the Biological Resources Division of USGS). At a practical level, the two projects have a similar purpose, to control taxonomic vocabulary and classification in a community database or an incipient federation of databases (the National Biological Information Infrastructure¹⁰). The two projects differ, however, in their taxonomic scope and hence in their respective approaches. Molecular sequence data have been gathered from only a small fraction of known taxa. In addition, GenBank is built

² <http://plants.usda.gov/npdc/18tdwg.html>

³ <http://iopi.csu.edu.au/iopi/iopihome.html>

⁴ <http://www.sp2000.org/>

⁵ <http://plants.usda.gov/npdc/standrds.html>

⁶ <http://www.bgbm.fu-berlin.de/IOPI/IOPIModel73/7301root.htm>

⁷ <http://bgbm3.bgbm.fu-berlin.de/iopi/gpc/>

⁸ <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html>

⁹ Federal agencies participating in ITIS include EPA, Dept. of Commerce/NOAA, Dept. of Interior/USGS, USDA, and the Smithsonian Institution; see <http://plants.usda.gov:80/itis/>

¹⁰ <http://www.nbs.gov/nbii/index.html>

from the submissions of individual scientists and laboratories. The NCBI strategy, therefore, has been to control nomenclature after sequence submission through a curation process.

“In order to be comprehensive, all organisms in GenBank must have an entry in the tree. In this regard, the taxonomy is driven by the organisms being sequenced rather than all organisms that exist. As of August, there were 1,920 family nodes, 5,588 genus nodes, 15,511 species nodes, and 1,903 nodes below the species level represented in GenBank. An average of 10 new organisms are added each day.”¹¹

Environmental data, in contrast, encompass a higher proportion of known taxa. The ITIS approach has therefore been to compile nomenclature prospectively; i.e., to build a comprehensive classification from the literature rather than from taxonomic names already contained in other databases. The ITIS steering committee recognized that to be truly authoritative, compilations of nomenclature and classification must be both comprehensive (historically complete) and curated to the highest standard. In addition, they recognized that these objectives demanded resources beyond what the partner agencies could afford, and would therefore require cooperation from the systematics community. They have sought to engage practicing systematists in both the review of existing files and the development of new ones for groups where compilations are lacking or out of date.

Another taxonomic compilation project is being initiated as an extension to an existing literature abstracting and indexing enterprise. BIOSIS, the not-for-profit organization that publishes Biological Abstracts and the Zoological Record, is developing TRITON¹² (Taxonomic Resource and Index To Organism Names), a system that will provide both a detailed version (subscription-based) and summary index (publicly available) of information on all new names and combinations published since 1978.

A final, but potentially very important influence on the activities of individual systematists, is the software for managing nomenclature, classification, literature references, and distributional information at the level of individual research projects. Desktop level applications that serve this function include *Platypus*,¹³ the ITIS *Taxonomic Work-Bench*,¹⁴ and *BASIS*.¹⁵ While these applications help the systematist to gather, organize, and format information for publication, they do so by capturing data in structured form and are thus capable of generating “batch uploads” to a larger compilation/dissemination server. (Facilitating data capture was in fact the ITIS project’s strategic motivation for developing the *Taxonomic Work-Bench*.)

3. Community infrastructure for managing taxonomic authorities

In the same way that a single database uses a taxonomic authority file to guarantee consistent description of items and predictable data retrieval, a federation of databases needs an analogous authority file, created and deployed across participating databases. The file establishes a single authoritative copy for any piece of taxonomic information, to which all references point, or from which all copies and updates flow. The challenge to the systematics community and biological information managers is to develop mechanisms that facilitate the capture and reconciliation of historical literature, as well as the flow of current taxonomic research, into a coherent authority file, and then the deployment of authority control into all relevant databases. Ultimately dissemination must extend directly into the processes whereby taxonomic names are applied to observations and specimen records.

¹¹ <http://www.ncbi.nlm.nih.gov/Web/Newsltr/sept95.html#Taxonomy>

¹² <http://www.york.biosis.org/triton/>

¹³ <http://www.ento.csiro.au/platypus/platypus.html>

¹⁴ <http://www.itis.usda.gov/itis/twb.html>

¹⁵ <http://www.sel.barc.usda.gov/basis.htm>

At least three major shortcomings impede the development and deployment of taxonomic authority files:

- 1) heterogeneity in the scope and underlying semantics of the data captured;
- 2) the lack of an integrated, systems-oriented, description of the functions and processes that either must be included or can effectively contribute to the long-term goals; and
- 3) the lack of effective mechanisms for distributing the compiled data so that they can be used in other disciplines to control storage and retrieval in data repositories.

Heterogeneous scope and semantics: Despite the homogenizing influences of meetings, publications, and standards developed by TDWG, most compilation projects have been initiated and conducted more or less independently. Each has been driven by different priorities, implemented with a different strategy, and more importantly, based on different conceptual models of taxonomic information. The semantics embodied in the data structures are substantially different in some cases, and most could be viewed as logically incomplete in some way. From the perspective of a data manager attempting to build and maintain a local taxonomic authority file from multiple sources, each different source represents a separate data migration effort, and some sources leave the data manager with an incomplete solution; e.g., historically complete lists of names without designations of which names are valid; or checklists of valid names without historically complete, or worse yet, structurally inaccurate representations of synonyms for those taxa. (For example, one recently completed compilation included only the specific epithets of synonyms, rather than the complete genus-species combination used in the literature and legacy data.)

An integrated description of functions and processes: As noted above, the compilation projects that are being conducted now have all been conceived with relatively narrow perspectives on the process of compiling and disseminating authority files. Each perspective grants different priorities to information that might be gathered and designates different user-communities as the primary audience. Note that the information sought by a systematist from a checklist or taxonomic bibliography differs significantly from the information sought by a molecular biologist. Concentrating on priorities is of course how things get done quickly, but we do not yet have a good understanding of how compilations with narrow geographic and broad taxonomic focus, for example, can be integrated most effectively into compilations with different dimensions. What are the minimally significant pieces of information? How should their accuracy be documented? How can they be transferred to another compilation project, or more generally, made available to all scientific databases? How do smaller, well-researched, pieces of information (like author names, or journal abbreviations) get assembled into more specific, larger pieces of information (like a complete citation to a particular reference work), and how can an assemblage of well-standardized data be distinguished from a non-standard representation? These questions do not just call for project- and item-level metadata, but for a better understanding of how compilation projects, individual systematists, and database managers in other disciplines can cooperate.

Distribution mechanisms: Compilations of valid taxonomic names, synonyms, source literature, and classification are a traditional form of taxonomic work, and have a long history in standard print media. Modern compilation authors continue to use print media because they have reasonable expectations of receiving credit for their work through the standard methods of professional evaluation, and in some cases because some portion of sales revenue might be programmed back into project support. Some authors have added compact discs (CDs) as a distribution medium because of the obvious analogy to print media, but the trend emerging now is for taxonomic databases to be made accessible via the Web.

Web access to taxonomic information is typically provided through one or more search screens, which enable end-users to issue simple queries, such as “list all species in genus X”, or “show the bibliographic details and status of name Y”, and then to read the results on screen. This kind of access enables authors to

make their information freely and quickly available, while at the same time allows them to feel confident that they will retain sole possession and control of the entire work.

Several authors or projects have made their compiled data completely or partially available as structured files that can be downloaded and imported into other systems, but do not “advertise” this kind of availability. Little attention has been given to how distributed copies of authority files can be updated. In general, dissemination of taxonomic authority files as structured data is most appropriately characterized as *ad hoc*.

It is likely to be a some time before a significant proportion of biological databases, where authority control ultimately needs to be implemented, have the ability to integrate external data sources, such as an authority file at a remote and separately administered site, in real time. The heterogeneity among database technologies, their expected persistence as legacy systems, and the performance and reliability problems of internetworking all indicate that the more robust way to deploy authority control will be through the distribution (replication) of authority records. This means that taxonomic compilers need to work with data managers in other domains to devise protocols for distributing authority files and updating individual authority records.

4. Comparison with Authority Control in Library Cataloging

The roles of taxonomic nomenclature and classification in biological databases have several important similarities to the use of authorities (documented controlled vocabulary) in library catalogs.

- In biological databases, taxonomic names are applied to specimens or observations to indicate membership in taxa (e.g., species). In library cataloging, one or more subject terms (names or phrases) are assigned to cataloged items to represent the subject of their contents. In both cases, descriptive terms are controlled to enable consistent (predictable) information retrieval and to maximize the power of discrimination among items.
- In both environments, authority records create a system of mutually exclusive concepts (and in some cases hierarchical classifications of those concepts) and then provide a preferred term for each concept as well as any synonyms that might have been used in a particular community or at an earlier time.
- In library cataloging, authority control represents a series of processes that: coordinate the gathering of terms from multiple sources; document the origins and meanings of terms, as well as the reconciliation of competing terms (synonyms); and distribute authoritative decisions across the community to coordinate information retrieval systems. This same type of community-wide implementation and coordination is an important goal in compiling taxonomic authority files.
- In both environments, the processes are on-going to accommodate the discovery and creation of new knowledge and concepts.
- In both environments, users would like information retrieval systems to translate query terms into standard terms automatically, if such a translation is unambiguous, and to facilitate disambiguation whenever necessary.

In addition to these similarities, the Library of Congress has initiated a program to explore the use of new technologies to further “Cooperative Cataloging” in the library community.¹⁶ The program includes the

¹⁶ Program for Cooperative Cataloging (PCC – <http://lcweb.loc.gov/catdir/pcc/>)

replication of name authority records (the names and identifying details of personal and organizational entities, etc.) from distributed cataloging operations, to a central repository at the Library of Congress, and then back out to copy-holding participants.

These basic points of comparison and the library community's experience in creating and operating this type of community infrastructure provide the impetus for this workshop.

5. A Workshop on Taxonomic Authorities for Describing Biological Data

Over the past several years, members of the library science community and informatics specialists in the systematics community have been discussing the similarities between the two disciplines. As a result of these discussions, the American Library Association, Association for Library Collections and Technical Services (ALCTS) has extended an offer to the PI to coordinate the participation of library science experts in the workshop proposed here. Because of their vested interest in taxonomic authority files (the ITIS project), the PI discussed the premise of the workshop with members of USGS-BRD, who then offered a similar expression of interest.

The purpose of the workshop is to identify the missing or inadequate components of a system for assembling and distributing taxonomic information to data managers, both within and outside systematics. The workshop will provide an opportunity:

- 1) for members of the library science community to describe the elements of their authority development and maintenance systems to the taxonomic community;
- 2) for the systematics community to describe the intricacies of taxonomic nomenclature and classification to the library community; and
- 3) for managers of applied biological databases (i.e., outside systematics) to articulate their needs for taxonomic information.

The workshop format will consist of presented papers and opportunities for discussion interspersed throughout the two-day duration. Invitations will be extended to both subject experts in library science, stake-holders in the systematics community, software developers, and representatives from the international organizations (TDWG, IOPI, Species 2000, and BIOSIS), up to a total of approximately 30 supported participants. Invited papers and summaries of discussion will be compiled into a proceedings that will be published electronically (i.e., as electronic documents made available via the Web). This medium is deemed most appropriate because virtually all members of the target audience have access to the Web, and because the timely dissemination of workshop results is viewed as critical.

Potential topics for the workshop are outlined below.

Potential Workshop Topics:

- Examples of authority control and tools for cooperative cataloging in library science:
 - ◊ Subject headings, name authorities
 - ◊ Thesauri for specific subject areas (e.g., NLM's Unified Medical Language System¹⁷)
 - ◊ Integrating thesauri and classification systems into information retrieval systems
 - ◊ Review and description of resources in the library community that could be applied to developing taxonomic authorities
 - Taxonomic authority files
 - ◊ Review of basic concepts (data model: taxa/names, references, authors, type specimens, classification, etc.)
 - ◊ Strategies for compiling authoritative data (emphasis on work flow, level of effort, results - quality & quantity); lessons learned.
 - ◊ Aggregating the results of smaller projects into larger assemblages
 - Technologies for implementing authority control; i.e., providing access to and distributing (replicating) authority files
 - Intellectual property rights and economic models for ongoing maintenance of authority files
-

6. Impact on Scientific Infrastructure

The National Science Foundation is preparing a research funding initiative called “Knowledge and Distributed Intelligence” (KDI),¹⁸ which is being developed further through workshops involving broad representation from the scientific information management community.¹⁹ Briefly, the initiative seeks to augment the flow of information between scientists, within and across disciplines, in ways that create new knowledge, and then make that knowledge available, intelligible, and useful to the widest array of audiences, including other scientists, decision makers, and the public. Taxonomy provides an interesting example of how scientific knowledge flows from one scientific discipline to another. Basic systematics research is responsible for discovering and describing the attributes of organisms, and then synthesizing that information into predictive classifications; i.e., knowledge. The taxonomic names and classifications representing this knowledge are then incorporated as simple descriptive labels on other data objects to manage those data across nearly the complete breadth of biological disciplines (e.g., as labels on gene sequences, or observations in ecological studies). The hard-won conclusions of systematists – the circumscription, name, and classification of a taxon – become mere pieces of infrastructure to scientists in other disciplines. If that infrastructure is not consistent across databases, however, the synthesis of information into additional knowledge will be impeded. The information integration challenge posed by taxonomy is to develop a data management and distribution network across a highly distributed scientific activity, involving thousands of individuals around the globe, that enables the products of the systematics research to be integrated into data management in all relevant areas of biology.

¹⁷ <http://www.nlm.nih.gov/pubs/factsheets/umls.html>

¹⁸ <http://www.nsf.gov/bfa/bud/fy1998/98ov021.htm>

¹⁹ <http://www.scd.ucar.edu/info/KDI/>

The resulting workshop report should provide taxonomic database managers with a broader perspective on: 1) the processes used to gather and manage basic taxonomic data, 2) the existing resources in library science that can be co-opted in the development of taxonomic databases, and 3) a better awareness of end-user requirements in the implementation of authority control across biological data.

D. References Cited

- Association of Systematics Collections [ASC]. 1993. An information model for biological collections. Report of the Biological Collections Data Standards Workshop, August 18-24, 1992. URL <gopher://www.biodiversity.uno.edu:70/11/standards/asc>
- Robbins, R .J. 1994. Report of the invitational DOE workshop on genome informatics, 26-27 April 1993. Genome informatics I. Community databases. *Journal of Computational Biology* 1(3):173-190. Available from http://www.bis.med.jhmi.edu/Dan/DOE/whitepaper/inf_rep2.html
- Systematics Agenda 2000 [SA2K]. 1994a. Charting the biosphere. Systematics Agenda 2000.
- Systematics Agenda 2000 [SA2K]. 1994b. Charting the biosphere – technical report. Systematics Agenda 2000.
- Wilson, K., W. Berendsohn, F. Bisby, H. Burdet, J. Burnett, A. D. Salinas, D. Geltman, A George, C. Heyn, G. Lucas, N. Morin, S Peterson, M. Roos, D. Sutton, C. Zellweger. 1994. Global plant checklist project plan (Ver. 1.2, January 1994). Unpubl.