REQUIREMENTS ANALYSIS

PART I

FUNCTIONAL REQUIREMENTS

COLLECTIONS INFORMATION SYSTEM
RE-ENGINEERING PROJECT

MUSEUM OF VERTEBRATE ZOOLOGY
UNIVERSITY OF CALIFORNIA
BERKELEY

DRAFT – 8/31/95

I. Introduction

A. Collections Computing at MVZ

The Museum of Vertebrate Zoology began to automate its collections information in 1978 using a database management system, called TAXIR. TAXIR was originally developed on the University of Michigan operating system (MTS), and with funding from NSF, was ported to the IBM 3090/CMS platform at UC Berkeley. TAXIR provided nominally interactive capabilities for data entry and querying the database, but supplemental batch data entry procedures were soon developed to run on computers with less expensive connect-time charges than the TAXIR host. The initial batch data entry program was followed by two replacements, with the last, TAXJOBS, being developed in 1985 to run on personal computers. TAXJOBS incorporates at least a minimal degree of data type and content control, and is still the principal mode for entering data into TAXIR.

Retrospective capture of MVZ collections data began with the mammal collection, and continued with the bird collection, from 1982 to 1985. Data capture for the three primary collections was finished by 1988 with automation of the herpetology collection. Because of the size and importance of these collections, complete automation of basic specimen data placed MVZ among the leaders of computerization in natural history museums. Subsequent projects have automated catalog data for the egg and nest collections and a significant portion of the frozen tissue collection.

Slightly different data structures and rules for field content were developed for each of the automated collections. These modifications were made in part to accommodate peculiarities of the different disciplines, and in part to correct problems discovered in earlier projects. Further segmentation of the TAXIR data files was required because memory management limitations prevent TAXIR from operating on data files larger than 40,000 records. MVZ collections data are partitioned taxonomically among 39 files (TAXBANKS) of five different kinds. While any one file can be queried interactively, it is often more efficient to query multiple files with a script (batch job).

Although a suite of auxiliary programs were developed for use with TAXIR in 1989 (to provide, for example, substring search capability), the core TAXIR software has not been upgraded or enhanced since the original installation. The only additional collection management function to be automated was the production of loan forms in 1988, and this was accomplished with a completely separate d-Base III application. TAXIR's basic flat-file structure, batch data entry, and lack of interoperability with modern PC based (DOS/Windows and Mac) software, represent an information management paradigm now more than 15 years old. More importantly, progressive upgrading of the IBM CMS operating system, without commensurate upgrading of the TAXIR software, has caused the MVZ's data management system to become unstable. Only a professional-level backup regimen prevented a significant data "loss" (i.e., the necessity of re-entering a large amount of data). A significant procedural overhead now accompanies any update to the database to guarantee its integrity. Finally, the

Department of Information Systems and Technology (IS&T), which maintains and operates the IBM 3090 platform on which TAXIR runs, has determined that this mainframe computer is no longer consistent with the University's general information technology strategy. IS&T has scheduled the 3090 to be decommissioned (removed from general use) in July of 1997. The MVZ must therefore acquire or develop a new system to maintain even its current information management capabilities.

As an institution, the MVZ has dedicated past information management efforts exclusively to collections cataloging and loan processing. Beyond that, computing technology has been deployed within the Museum only to provide the basic infrastructure expected in a modern university, i.e., desktop computers, network connections, and software for word-processing, statistical analysis, e-mail, Internet access, etc.

B. The Collections Information System Re-Engineering Project: the context for the requirements analysis

The MVZ has initiated a long-term project (3-5 years) to replace TAXIR and to enhance the management of collections information generally. The term "collections information" is used here broadly, and is intended to encompass all collections and research information gathered or produced by the Museum in sufficient volume to warrant being considered an institutional resource. The new system is thus intended to reflect the entire scope and diversity of the Museum's research, education, and service activities, not just those traditionally associated with collections management.

The replacement system is intended to address all of the Museum's information management needs that can be anticipated to arise within the next five years. This is not meant to imply that technology developed now will be obsolete within five years, but only to acknowledge that, in general, needs-assessment is strongly influenced by notions of what is possible. Information technology is changing so rapidly that projecting needs and possibilities beyond a five-year horizon is of questionable value.

Development of the replacement system will be divided into the following phases:

1. Analysis and Design
    a) Requirements Analysis
        i) Functional Requirements
        ii) System Requirements
    b) Information Modeling
    c) Interface Design (function/process modeling)
2. Programming and Testing
3. Data Migration/Capture
4. Deployment and Maintenance

The activities listed above will not be conducted in a single linear progression. The system will be developed as component applications, such that phases 2-4 can be

iterated for each application, allowing the higher priority applications to be brought on-line before the entire system is complete. The initial analysis and design activities, however, must be comprehensive in scope so that applications can be developed serially without requiring earlier applications to be re-designed and re-coded to accommodate later applications.

The requirements analysis is the first of three analysis and design products. It is conducted at the outset as a means to describe and prioritize the Museum's information management needs, and to describe the capabilities and characteristics of the desired system. A requirements analysis is typically conducted when an organization undertakes a system development effort that will extend over several years and will involve several individuals. Conducting a requirements analysis helps to reconcile different perspectives and to coordinate subsequent analysis and design activities.

The requirements analysis is divided into two major sections: functional requirements, which document the needs for information processing technology, and system requirements, which describe the essential characteristics of the hardware and software that will meet those needs.

The functional requirements analysis contains a high-level overview of the organization, in which the common activities, processes, and products are described in relation to how they create, use, and modify information. The functional requirements analysis poses the question:

> "How can information technology be used by the organization to accomplish its 'mission' more efficiently and more effectively?"

The question is answered with a prioritized list of the functions the system must be capable of performing. (It is also appropriate to re-examine whether each information processing activity currently performed in fact serves a valid purpose, and whether any new products or uses of information can be developed from existing information resources.)

System requirements describe the system from a technical perspective. They should specify the capabilities, capacities and characteristics of the system in both qualitative and quantitative terms. Both functional and system requirements should express specifications (i.e., describe capabilities and characteristics) without making undue assumptions about their implementation. To do this, requirements are typically expressed in terms of how the operations of the organization (normal work activities) place demands on the system. For example, it would be inappropriate to state that the system must use a relational database, unless this were known to be an existing constraint (i.e., the organization already owns or has access to a particular DBMS). It would be more appropriate to state that the system must be able to manipulate data records by field values, update multiple data objects as a single logical transaction, accommodate up to 20 simultaneous users, etc.

II. MVZ Background

A. Mission

The Museum of Vertebrate Zoology is a research and educational center dedicated to the study and conservation of amphibians, reptiles, birds, and mammals. The program is an inclusive one that focuses on the biology of terrestrial vertebrates from the perspectives of their ecology, behavior, morphology, population genetics, and systematics. It involves both field and laboratory investigations, applying both theoretical and empirical approaches to questions in evolutionary biology.

While the Museum's original benefactor, Annie M. Alexander, stipulated that the Museum was not to devote significant resources to public programs and exhibits, the Museum's first director, Joseph Grinnell, soon prevailed in convincing Miss Alexander that graduate education
would be vital to the furtherance of the Museum's research and conservation missions. Accordingly, the Museum has become one of the principal training grounds for graduate and postdoctoral students who then go on to careers in higher education as well as to positions at museums and in government, both in this country and abroad. The common housing of faculty, staff, postdoctoral associates, visiting professionals, and graduate students in the Museum makes for an extraordinarily rich intellectual atmosphere, paralleling the integration of the collections and all accumulated materials into a single, dynamic resource for education, research, and conservation. The Museum is internationally acknowledged as one of the most important and dynamic centers for research and education in vertebrate zoology. The training of foreign graduate students has become a priority, with recognition that only education will permit the conservation of resources globally.

The foundation and core of this program are the ever-growing and incomparable specimen collections and supplemental materials associated with them. All systematics collections provide physical documentation for some portion of the biotic world. The MVZ collections are unique among them, however, because they have been derived almost entirely from the research programs of the MVZ staff and students. Hence, ancillary information associated with specimens is extensive in most cases, and large series of specimens from given localities, often sampled repeatedly over time, exist. These specimens, the special preparations derived from them, and the information describing the biotic and abiotic conditions under which they were obtained, constitute a singular educational and research resource.

B. Philosophy

The Museum of Vertebrate Zoology's first Director wrote:

> "It will be observed, then, that our efforts are not merely to accumulate as great a mass of animal material as possible. On the contrary, we are

> expending even more time than would be required for the collection of specimens alone, in rendering what we do obtain as permanently valuable as we know how, to the ecologist as well as the systematist" (Grinnell, 1910).

Thus, the MVZ was founded with the philosophy that organisms should be studied in relation to their environments and habits. Accordingly, specimens in the Museum are enhanced by an array of supplemental materials, e.g., hundreds of volumes of field notebooks, thousands of topographic maps and photographs of specimens and their habitats, a collection of natural sounds on tape, tissue samples, slides of chromosome and histological preparations. The Museum also houses the correspondence from its date of founding to the present, with the exception of letters regarding its establishment and endowment, which are housed in the University's archives.

These Museum materials are not only important to scientific investigators, but increasingly have attracted historians, sociologists, philosophers and educators interested in a broad spectrum of topics, from the development of vertebrate natural history as a discipline, to creation of the national park system in the United States. It is both the supplemental materials and the fact that collections management has always been addressed on a museum-wide basis rather than according to traditional taxonomic divisions that make such studies possible in this institution.

Our intention is to use the vision that has resulted in this uniquely integrated collection of materials as the guiding theme of this project. A system that accurately represents all aspects of MVZ's "business" is the only one that would be in keeping with the Museum's founding philosophy and thus, is the aim of this project.

III. Functional Requirements

A. Introduction

As noted above, the Museum’s collections exist to support its research, education, and conservation missions. The collections, in turn, have grown so large that a fourth function, collections management, has emerged as a distinct infrastructural entity (Fig. 1).

The primary motivation for this project is clearly the need to improve the Museum’s abilities to manage its collections information. The MVZ, however, is not simply in the business of developing and managing collections. Thus, it is appropriate to examine whether any significant benefits can be obtained by applying information technology to the three primary functions, independently of collections management. The functional requirements analysis therefore begins with a brief assessment of the activities that create, use, or process (transform) information in the research, education, and conservation areas. The intention is not to send “a solution [information technology] in search of a problem”, but rather to ensure that no significant opportunity to improve efficiency and effectiveness is overlooked for having focused the project too narrowly.

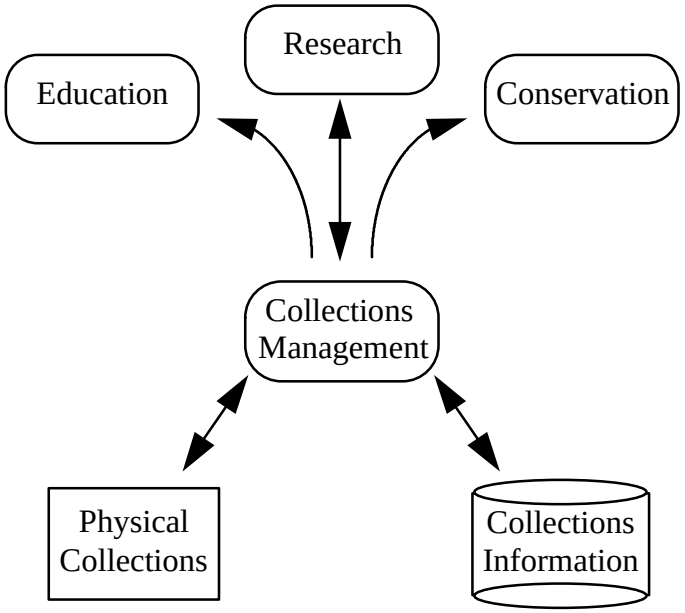

Figure 1. High-level Museum Functions. The four dominant Museum functions are shown as rounded rectangles. The collection management function coordinates the flow of objects and information between the primary functions and the permanent repositories.

B. Research

For purposes of this analysis, information-related activities within the research area are divided grossly into four categories:

1. Data gathering – e.g., planning field work; collecting specimens, samples, and their associated data; performing subsequent laboratory analyses (i.e., genetic typing, morphometric data collection, behavioral experiments, etc.);
2. Data Management and Analysis – the derivation of intermediate and final results;
3. Publication –any presentation of research results, from manuscript preparation, to the development of visual materials for presentations, and electronic publication (making information available via the Internet);
4. Communications – e.g., correspondence, Internet access (e-mail, ftp, Gopher, World-Wide-Web, etc.).

The information processing activities that support research are diverse and change quickly with advances in science. Moreover, the boundaries between institutional and individual responsibilities for information are not always clear, and it is difficult to determine the ways that information management resources should be allocated to support the research activities of individuals. It is therefore more appropriate to express research-related "functional requirements" as a series of principles or

guidelines that should be kept in mind as the system is being developed and as research programs are modified to take best advantage of emerging technologies.

1. Data Gathering

An important principle that can improve the data gathering function is to capture data electronically as close to their sources or origins as possible. It is becoming increasingly practical to record data directly into computer data files, by bringing computers into the field or laboratory, or by setting up real-time data logging. Benefits associated with direct data capture include increased efficiency, elimination of data transcription errors, and the support (enforcement) of a consistent data collection protocol. Drawbacks can include a dependency on equipment that is fragile or operates only in restricted circumstances, and a restricted ability to record supplementary remarks or illustrations. Examples where this principle could be applied in the near future include the use of personal digital assistants (PDAs) or laptops to record observations in the field, or the use of sensors and software in labs to determine nucleotide sequences. Another important example is the use of hand-held Global Positioning System (GPS) receivers to record the latitude-longitude of collecting localities. Some GPS devices have only data display capability, which means that data must still be recorded by hand, but others provide data logging, temporary storage, and the ability to upload data files to larger systems. In either case, the quality of geo-reference data is improved substantially over that obtained from maps. The MVZ already has one GPS device, and will be acquiring additional units in the coming years.

2. Data management and analysis

From an individual researcher's perspective, perhaps the most important requirement for the new system is the capability to view and use collections data directly from his/her desktop computer. The network infrastructure required to fulfill this requirement is already in place. An easy-to-use query and reporting interface, based on a client-server architecture, is the largest remaining component needed to make the requirement a reality. The benefits expected from improved access include the ability to integrate these data with visualization and analysis applications, such as mapping and geographic information systems (GIS), and the improvement in data quality that results from having the data creator use (and review) data soon after they have been entered into the system. Greater use of collections data should also increase researchers' awareness of problems with accuracy, completeness, consistency, etc., that can be eliminated by better data collection protocols.

From a data manager's perspective, the challenge is to ensure that large research data sets are structured and described in ways that maximize their long-term usefulness. This means that institutional resources for data management, including database design and application development be committed in support of research. It may also entail striking a balance between the researcher's concerns, such as expediency and autonomy, and the data manager's concerns for standards that promote long-term

usefulness of the dataset. But with longer-term research programs, and consequently larger datasets, the interests of the researcher, data manager, and institution should tend to converge.

3. Publication

Information technology can facilitate the dissemination of research results at a number of levels. The underlying infrastructure for much of this activity will be the Museum's World Wide Web server. WWW servers are now commonly used by systematics collections to "publish" collections data, both as text and as structured data files, data sets that support traditionally published papers, and "reprints" of published papers, as well as less the images and text of talks given at meetings.

Through cooperation with the Museum Informatics Project, the MVZ has established a presence on "The Web". It is clear, however, that electronic publication is a relatively dynamic activity, and thus the Museum needs to develop its own capability to update and augment the content of its server. In addition to the current basic description of the Museum and its collections, the server should contain: a query interface to all on-line catalogs; digital images of selected collection items (e.g., photographs of type specimens); a directory of all MVZ staff, including bibliographies and electronic reprints; and selected multi-media "documents" describing research programs and activities supported by the Museum.

4. Communications

Communications between MVZ researchers, their colleagues, or the public generally, are facilitated through the campus-run networking and internetworking services. Applications providing the capability for e-mail (Eudora), WWW access (Netscape), gopher, and ftp clients, are installed on virtually all staff desktop computers. The only capability that remains to be established is an ftp server for the Museum (operational 24 hours/day and with flexible security features; useful when an MVZ staff member needs to make a large file available to a colleague, but not the public at large).

C. Education

Education and teaching, through formal course offerings, seminars, or individualized student-faculty interactions, have traditionally been a central part of the Museum program. Curators take primary responsibility for several undergraduate courses and for many aspects of graduate education in the Department of Integrative Biology.

It is likely that the technology installed primarily to make collections information accessible can also be used to facilitate education, or more specifically, the dissemination of information to students. Within this context, "information" will typically consist of documents such as lecture notes, hand-outs, selected readings, etc. A variety of mechanisms could be used to make such information available "on demand", but the most effective way will probably be to establish WWW "home

pages" for classes taught by MVZ faculty. Although some extra effort will be entailed in the initial development of Web-based materials, subsequent revisions and updating will be less labor intensive. The Instructional Technology Program (ITP within the Department of IS&T) is developing tools and expertise to help UCB faculty develop Web-based instructional materials.

The ITP has also recently used an upper-division/graduate course in Evolutionary Biology (IB 160) as a test-bed for the concept of a Virtual Discussion Group. Authors of recent significant papers in evolutionary biology were recruited to form a world-class panel of experts. After reading the selected papers, students submitted their questions to authors. Electronic communication enabled students to discuss current issues with the "best in the field", and eliminated the time lags that attend similar dialogues in journals. The forum was judged to be a success by both students and the experts.

Much of the technology required to develop Web-based instructional materials or to host virtual discussion groups is already present on campus and available to MVZ faculty. The primary impediment to actually using these technologies remains the fact that faculty do not have the time to develop the requisite skill sets, or conversely, that some of the technology continues to be difficult to use. Within a year or two, however, tools for developing HTML documents and managing a Web server will require little more skill than it takes to use a full-featured word-processor and perform basic file management on a personal computer. MVZ staff should continue to keep abreast of the activities of the ITP in particular, and the advances in HTML authoring tools and web-server technology, generally.

D. Conservation

As natural habitats and their constituent organisms disappear at an ever increasing rate, the conservation of biodiversity becomes an important issue to both the national and international communities. Conservation strategies must rest on a firm foundation of information concerning natural populations of organisms, their systematics, ecology, population biology, and patterns of evolutionary diversification. Ultimately, all sound conservation policies and decisions, national laws, and international treaties, depend upon information derived, in part, from the collections-based and field research conducted by institutions such as the MVZ.

The size and breadth of the MVZ collections, both geographically and temporally, make them an important potential resource for conservation work. The scope of information required to formulate viable conservation strategies, however, extends far beyond the domain of the Museum's collection information (i.e., must include components such as satellite imagery, spatially referenced parameters of the physical environment, e.g., rainfall and soil-type, and socio-political information, e.g., demographic projections and land-use data). Moreover, the Museum lacks the expertise and technology necessary to transform its collection records into the kinds of information products that effectively influence resource managers and legislators. The

Museum can best support conservation by making its collections data readily available to conservation-oriented analysts, agencies, and organizations.

Although the function "provide access to collections data" is essentially subsumed under collections management (below), the crucial role of geographic data within the conservation function warrants special attention here. To make collections data useful within the context of a geographic information system -- as estimates of species distributions, both past and present --, collecting localities must be machine-readable and represented within either a sufficiently high resolution raster-based framework, or a coordinate-based system. At present, the vast majority of the Museum's collecting localities (98%) are represented only as human readable text. The Museum must undertake a substantial effort to geo-reference these data. The effort will be labor intensive, but significantly greater efficiency could be obtained with a combination of good query and data manipulation capabilities in the database, an electronic gazetteer, and a tool that enables users to digitize coordinate data from digital maps into locality records.

E. Collections Management

The management of collections information has traditionally focused narrowly on specimens. The MVZ collections, however, are much more diverse than specimens and standard preparations. A large component of the collections management work load has consisted of maintaining the associations between specimens, specially prepared materials, and the non-specimen collection items, such as photographs and field notes. Previously, most of this record-keeping has been done on paper because it did not fall within the initial scope of the collections database, and it has not been feasible to expand the scope and capabilities of the system. Much of the supplementary information about specimens and ancillary collections is consequently unavailable electronically. Improving this situation requires that the collection management function be assessed here as encompassing not just the specimen collections, but the complete diversity of collections maintained by the Museum.

The complexity of the collection management function derives not from an intricate series of material and information processing steps, but from the diversity of collections information itself. We therefore begin this section with an overview of collections information, before turning to a more procedural analysis of the collections management function.

1. Collections Information Overview:

The scope of information to be managed by the MVZ's new system is large enough that a structured analysis, i.e., a formal information model, will be required to develop an adequate system design. The analysis of MVZ information will not begin from "scratch", but will draw upon previous efforts at "museum information" analysis (information models, data dictionaries, and other data standards documents), such as:

- Information Models: Association of Systematic Collections (ASC), Specimen Management System for California Herbaria (SMASCH), Australian National Botanic Garden (ANBG), Society of Vertebrate Paleontology (SVP), Citation Database (Genome Database, Human Genome Project, Johns Hopkins University)
- Data Standards/Dictionaries: American Society of Mammalogy (ASM), Society of Vertebrate Paleontology (SVP), Canadian Heritage Information Network (CHIN)
- Thesauri: Art & Architecture Thesaurus, for fine art and cultural collections.

a) The diversity of MVZ collections

The Museum's research, education, and conservation programs are centered around its incomparable collections of MAMMALS, BIRDS, REPTILES and AMPHIBIANS. All three of these collections are synoptic both taxonomically and geographically. There is, however, a particularly strong representation of taxa from the western coast of North America due to the focused nature of the Museum's research programs during its first few decades. The Museum is also known for its large series of specimens from given localities, which were often sampled repeatedly to assess changes in population structure, species distributions, and environmental perturbation.

The MAMMAL COLLECTION contains almost 185,000 specimens, of which approximately 70% are represented by traditional skin and skull preparations. The remaining 30% of the specimens are skeletal preparations and fluid-preserved specimens. In addition to standard preparations and specimen information, ca. 4,000 chromosome preparations of rodents, referred to as the KARYOTYPE COLLECTION, are cross-referenced to original material in the mammal database. Both microscope slides and black and white photographs of chromosome spreads are included in this collection.

The BIRD COLLECTION has 177,000 specimens, the majority of which are preserved as study skins. The remainder are skeletons, skins with associated skeletal material, and fluid-preserved specimens. A separate catalog and database is maintained for the 14,000 SETS OF EGGS AND NESTS in the collection, most of which do not have voucher specimens in the bird collection. The importance of this collection lies partly in its antiquity. It is no longer legal to collect this kind of material, and many specimens were acquired prior to the introduction of environmentally destructive chemicals such as DDT. The COLLECTION OF AMPHIBIANS AND REPTILES contains almost 223,000 specimens, most of which are fluid-preserved. The skeletal collection is of special value because of the inclusion of numerous rare taxa. Their are also sets of amphibian eggs and larvae and several thousand specimens that have been cleared and stained. STOMACH CONTENTS that have been

removed from specimens in the collection are cross-referenced to their vouchers in a card catalog but not in the database. A HISTOLOGICAL SLIDE COLLECTION derived from several hundred Museum specimens has yet to be cataloged and fully cross-referenced in the database.

One of the first of its kind in the country, the Museum's TISSUE COLLECTION contains approximately 12,000 mammal, 3,100 bird and 4,400 herpetological samples. Most are frozen, but there is a growing collection of DNA extracts and of tissues preserved in 95% ethanol or in buffer. Approximately 70% of these samples have been removed from cataloged specimens in the collections. Although a database exists for some portions of this collection, a complete database and the capability to track tissue usage are still wanting and represent a clear priority for this project.

Because the Museum was founded with a formalized program for research, collections growth has been extraordinarily focused. Special emphasis has been placed on the accumulation of ancillary information which documents biotic and abiotic factors of an organism's environment that might have bearing on speciation events and on our understanding of the evolution of vertebrate communities. To that end, the Museum maintains an array of NON-SPECIMEN COLLECTIONS that enhance the materials described above and support its primary mission and functions.

Foremost among these is the COLLECTION OF FIELD NOTEBOOKS. This is composed of ca. 670 volumes from over 250 investigators who have placed specimens in the Museum. Information in these notes is the single most important source of data about the specimens, outside of what is written on the specimen tags themselves, and it is the primary source of data for curating the collections. These notebooks, as well as other of our non-specimen collections, are of increasing interest and importance to historians, philosophers, and sociologists of science as well as to environmentalists and private consultants wishing to reconstruct habitats or assess degrees of perturbation.

Starting at the turn of the century, and in conjunction with recording extensive field observations in notebooks, many researchers also took photographs of the specimens they collected and the habitats in which they were caught. The Museum's PHOTOGRAPH COLLECTION contains ca. 8,100 cataloged black and white prints and negatives, and perhaps and equal number of glass lantern slides, uncataloged black and white and color prints, and 35mm color slides. Cataloged material is cross-referenced on the photographs to vouchered specimens or collecting expeditions but has not yet been captured in the specimen database. As with the field notebooks, the value of cataloguing and cross-referencing all of this material is the historical record such data provide on changes in species composition and in habitats over time.

With the advent of tape recording devices, Museum ornithologists began recording bird vocalizations in the field. The Museum houses ca. 200 master reels and several hundred species reels of such recordings in its NATURAL SOUNDS TAPE COLLECTION. Only recently though have recordings been made of birds which were subsequently collected as specimens for the Museum. Although a card catalog and ledger exist for this collection, these records are not currently linked to our specimen database. Vertebrate sounds other than those of birds are represented to a far lesser extent in the collection. Interest in these tapes by producers and directors of movies, documentaries, and radio programs constitutes a non-trivial proportion of their use.

CORRESPONDENCE, which dates from the Museum's founding in 1908 to the present, contains detailed information about all Museum business and is a valued resource for curators and historical researchers alike. Filed alphabetically by correspondent and then by date, no finding aid or item-level catalog exists for this material. While used constantly by the curatorial staff in its work, the correspondence is perhaps most valuable because of the tremendous insight it gives into the social and political issues in which Museum personnel have been involved over time, as well as into the development of seminal ideas in evolutionary biology. Examples of the former would include issues of pest control at the national level and development of Yosemite and the National Park Service in this country. Examples of the latter include exchanges of letters between Museum staff and their colleagues concerning research in progress or manuscript reviews.

The Museum contains additional specimen and non-specimen collections that enhance its multiplicity of functions but, unlike the those listed above, generally are unrelated to the three primary specimen collections. A TEACHING COLLECTION of ca. 4,000 specimens supports that function in an array of undergraduate and graduate-level courses offered by Museum faculty. No catalog exists for this collection which is composed primarily of California vertebrates.

The MILTON HILDEBRAND COLLECTION contains ca. 500 cataloged specimens representing all vertebrate classes. It is distinguished by the unusual nature of most of the specimen preparations which make them invaluable for teaching. Many specimens are cross-referenced to functional laboratory exercises designed by Hildebrand. A separate component of this collection contains more than 100 reels of 16mm film, containing primarily short segments of gaits for a broad array of mammal taxa. These were used by Hildebrand as a basis for numerous research publications. Half of these reels have been converted to video format for ease of use and for archival purposes. A card catalog exists for all of these items.

The Museum's MAP COLLECTION contains an extensive series of standard published topographic maps, state and country maps, original species distribution maps, and maps that are historical records of collecting trips. These are further enhanced by a series of atlases and gazetteers of the world. No catalog exists for any of these resources but they serve a multiplicity of functions. Not only are they useful for planning research but older versions of maps are often the only source of geographic names found in our specimen records. Distribution maps, on the other hand, may provide insight into the development of ideas in evolutionary biology and were often the basis for scientific publications.

Another collection, which is actually a derivative of the Museum's research efforts, is its COLLECTION OF TECHNICAL AND FINE ART. Many of these items are original drawings from Museum publications but there are also paintings by well-known natural history artists that were commissioned to illustrate technical volumes. An inventory and crudely scanned images of the paintings are the only catalog that currently exists for any of the works in this collection, objects whose value lies more as art rather than as science.

The Museum also contains a small, in-house LIBRARY of reprints, books, journals, and standard reference works, of use in curating the collections and to researchers. The collection is distributed throughout the Museum. Those works used primarily for curation are organized by subject, while those for research and the reprint collections are arranged alphabetically by author. Many of the books are out-of-print volumes. An equally out-of-date card catalog of the books is the only catalog that exists for this material.

In general then we have chosen to view the multiplicity of MVZ's collections as constituting three groups, reflecting three levels of priority for application development and data capture. Unquestionably, the most important group encompasses the specimens themselves and their immediate derivatives. These are the vouchers upon which all analyses and conclusions are based, and which will continue to support new research well into the future. They include mammals, birds, reptiles, amphibians, egg sets and nests, tissue samples, stomach contents, and karyological and histological slides. Data for the overwhelming majority of these collections already exist in TAXIR. Retrospective capture of the remaining information-- on tissue samples, stomach contents, and slide preparations, will represent a relatively small effort.

The second group of collections are those non-specimen entities which contain valuable ancillary information about the specimens, e.g., supplemental directions to field sites, photographs of specimens and collecting localities, habitat descriptions, vocalizations, and loan histories. These include the field notebook collection, the photographs, correspondence, and vocalizations on tape. While we do not intend to capture all information contained in these

resources initially, on-line access to catalog listings of these materials will precede our eventual goal of capturing digital surrogates for many of the items, i.e., digital images or digital audio files.

Computerizing the last group of collections, i.e., teaching, Hildebrand, maps, library, and art work, may not occur during the initial phase of this project. Although each of these collections supports some piece of the Museum's overall mission, their relationship to the specimen collections and mainstream biological research is only peripheral. These collections are designated as lowest priority because they either serve a relatively small community, as is the case with the technical and fine art collections, or their utility would be only slightly augmented by an automated catalog, as would be the case with the teaching and Hildebrand collections.

The order of implementation outlined above is based first on the use which each collection receives, both for research by Museum personnel and by the scientific community at large, and in matters relating to conservation. Our second consideration has been the cost or level of effort required to bring these collection on-line in a manner that will clearly augment the Museum's ability to meet its overall mission.

Complete automation of each collection in a serial fashion is not, however, the only or most appropriate strategy for implementation. Several of the ancillary collections contain a significant body of material from California and western North America. It may therefore be more appropriate to develop the capabilities to integrate several kinds of collection catalogs, digitized images, etc., and then make a concerted effort to capture the materials from California or western North America. This might yield a useful information resource more quickly than the serial approach. Portions of collections that support current MVZ research may also be given a higher priority for data capture than materials of primarily historical value, which would be left to later projects.

b) Information objects (subject areas)

The following is a list of subjects (information categories/classes/objects/entities) that we intend to incorporate into the new system. After each subject (bold) is a list of sub-topics or kinds of information that will be used to describe the subject. The lists are not comprehensive, nor do they represent the final classification that will be used. They are intended to provide a sense of the breadth and diversity of information to be managed by the system.

1. **Collection Objects** – specimens, parts, preparations, non-specimen items, etc.
2. **Taxa** – name, rank, parent taxon, position in non-alphabetic sequence; original source-reference, version (in classifications)

3. **Collecting event** – date/time; references to specimens, collectors, and locality
4. **Locality** – political and natural geographic names; specific locality; elevation; lat-long; Twshp-Rge-Sec
5. **Supporting documentation** – field notes, correspondence, photographs, etc.
6. **Publication/Reference/Citation** – publication type (monograph, journal article, manuscript); authors, editors, title, pages; relationship to specimen or other object
7. **Person** – name, address, affiliations, MVZ status
8. **Organization** – name, address, acronym
9. **Agent Role** – collector, preparator, identifier, donor, borrower, author, illustrator, photographer, etc.
10. **Transactions** (e.g., loan, accession) – transaction type, transactor, dates, material description
11. **Transaction items** (information about specimens and other objects as they participate in transactions) – return date, permissions/restrictions, condition
12. **Permits** – issuing agency, permit number, purpose, duration, issued to, etc.

2. Collections Management Functions

Collections management includes the following basic functions: creation and maintenance of catalog records, transaction management, and information retrieval and reporting.

a) Creation and Maintenance of Catalog Records

(1) Cataloging

The information-based objectives of the cataloging function are: to assign a persistent unique identifier to a collection object; to record and maintain basic descriptive information about the object, in both electronic and paper media; and to record its relationships to all other objects or concepts within the scope of the system.

(a) Data Entry & Proofing

1. The system must provide the capabilities to enter records for both newly acquired material and material that is already cataloged on paper-based media.
2. The "boundaries" between catalogs should not artificially limit the users ability to associate a new specimen to other specimens or concepts existing in the system. For example, a user should be able to associate a kangaroo rat with a rattlesnake, as an item eaten by the snake.

3. The system should minimize redundant data entry by facilitating the flow of information from the accession process (see transaction management, below) to the cataloging process, and by allowing the user to link many specimens to a locality record or person record, etc.
4. The system should be capable of importing field data that has been captured on a laptop, GPS, or other field or laboratory equipment.
5. The system must support a proofing and correction process for new records (i.e., print reports of new records, allow errors to be corrected, and proofed records to be "posted" to the database).

(b) Generate Cards, Labels, and Tags

The system must support the production of catalog cards, labels and tags. It should provide the capability to print these "output items" both one at a time and in batch-mode, i.e., allow items to be queued for printing later. (See System Requirements section for specific printer, paper, ink, and font requirements.)

(2) Maintain the Catalogs

Both the discovery of errors in existing data and the usage of collection materials create the need to modify and update existing catalog records. In addition to providing the capability to efficiently retrieve and update existing records, the system must provide the capability to print additional or replacement cards/labels/tags as necessary.

Updates will not be limited to specimen or other object records, as the use of authority files (see below) will require that each data concept and it's relationship(s) to other concepts be fully modifiable by authorized users. Through these updates, the Museum will keep it's authority files current with changes in taxonomy, geographic nomenclature, kinds of specimen preparations, etc.

(3) Data quality management

The existing TAXIR data structure is relatively un-normalized, and contains many duplicate representations of the same concept. To a large degree, the responsibility for maintaining consistency among the duplicates rested with data entry personnel, rather than with the system, and a significant amount of inconsistency has crept into the database. Resolving inconsistencies remains a significant work-load on the collections management staff.

(a) Normalized data design: the use of "authority files"

To the greatest extent practical, data consistency should be promoted within the system through the use of normalized data structures, in which each real-world concept, or combination of concepts, is represented by a single master record. Consistency among any intentionally replicated data (to speed performance) should be enforced by system code. Users will perceive this strategy as the extensive use of authority files.

(b) Examine data content and improve consistency

Cleaning up the TAXIR data will be a labor intensive process and existing inconsistencies are likely to persist well into the future. The system should support the continued search for inconsistencies, and their correction, with powerful query, reporting, and update capabilities.

b) Transaction Processing and Management of Transaction Information

Documenting collection transactions constitutes a substantial portion of collections management activity. Generally, the system should the facilitate these tasks by allowing users to generate the normal transaction documentation (e.g., loan forms and accession cards), while at the same time capturing transaction information in the collections database. We do not intend to build into the system the extensive communications, electronic signaturing capabilities that are required for a "paperless office". Hard-copy documents will still constitute the legal records of Museum business, and the work load is not large enough to warrant an expensive solution.

Specific requirements of the transaction management function are as follows:

(1) Acquisitions (Accessions)

Important objectives of acquisition/accession documentation include: describing the means by which title to an object (ownership) is transferred to, or established by the Museum; establishing a permanent information-base that can be queried to determine how an object was acquired and the disposition of any and all material encompassed by a given acquisition/accession.

1. The system should facilitate the flow of information from acquisition documentation to the collection catalogs.

2. The system should minimize the redundant data entry and should support actual physical processing by producing cards, labels, tags, etc.
3. To facilitate the management of acquisition processing, the system should support monitoring and reporting functions that enable a manager to determine, for example, whether material from a given acquisition is being skeletonized, how much material is awaiting what stage of processing, etc.
4. The system should link acquisitions to collecting, salvage, CITES, and shipping permits. It should allow users to report how many specimens were collected under a given collecting permit or all permits within a given period.
5. The system should link accessions to specimens (i.e., accession numbers to specimen numbers).

(2) Loans

The most important purposes of the "loan function" are to facilitate the off-site use of collection materials, to enable MVZ to know the current location or disposition of any collection object, and to allow MVZ staff to report on the usage of collections and collection objects.

1. The system must support the logging of requests for material, as this information is typically included in annual reports and grant proposals as a measure of workload and collection usage.
2. The system must support querying and reporting against the collection catalogs (see below), and should allow query results to be transferred to the item list of the loan in preparation.
3. The system must support the production of loan invoices and shipping records (see shipping below). System should "enforce" or promote the policy that loans cannot be made to students, but rather to the student's sponsor *for* the student.
4. The system must support the description of loan material (on a loan invoice) that has not been cataloged, as well as material that is not a biological specimen (e.g., a video).
5. Any material sent out on loan should be "marked" as "on loan" in the collection catalog so that subsequent retrieval of the record will clearly show that the specimen is on loan, and the system should allow the user to display more detailed information as necessary, including the borrower, initial loan date, and due date.
6. The system should keep a permanent record of the loan history for each specimen.

7. The system should allow loans to be monitored for receipt (i.e., link loans to shipments).
8. The system should monitor active loans, (i.e., monthly reports on active loans), and should allow the staff to produce dunning letters and modify due-dates in response to requests for extensions. The system should also allow long-term loans to be marked as "not to be dunned".
9. The system should facilitate reconciling the return of a loan, either partial or in full against the "original invoice"; and should allow the user to update the description, condition, and status (returned, deaccessioned [with link to deaccession number], transferred, etc.) of returned material.
10. The system should also allow item(s) on loan to be transferred to de-accession transaction (recipient of loan (or a third party) becomes recipient of deaccession).
11. The system should support the "loan transfer" function, in which all or a portion of lent material is transferred from the borrower directly to another borrow (with permission from the MVZ).
12. The system should be capable of producing annual statistics on loan activity for the current and any past year (since implementation).

(3) Borrows (Incoming Loans)

The Museum needs to know who has responsibility for and title to any object that is physically located within the Museum and could be construed as a member of an MVZ collection. In the past, relatively little institutional record-keeping has been done with borrowed material and this has caused problems for the collection management staff. In the future, the MVZ intends to record basic transaction-level information in the new system, but will not itemize borrowed material. Photocopies of the lending institution's loan invoice suffice for item-level descriptions, and marked up copies will be used to invoice a "return of borrowed material". The recorded transaction-level data will be used to monitor the status of borrowed material and to generate labels for outgoing shipments.

1. The system must support the tracking of borrows at transaction level (not at the item level);
2. The system must be able to record:
    - the borrower (MVZ faculty/research staff and student if applicable),
    - the lending institution/person, their invoice number, due date, and a summary description of the material borrowed.

3. The system should be able to record whether copies of the invoice and relevant correspondence have been deposited in the department files.
4. The system must be able to record the storage location of borrowed material.
5. The system should print a transaction-level "invoice" that the lender can use to acknowledge the receipt of returned material (must be able to do this for both partial and final returns of borrowed material).
6. The system should enable collection management staff to track return of borrow shipments.
7. The system should enable borrow transactions and their shipments to be associated with all relevant permits.
8. The system should be expandable to accommodate item-level descriptions in the event that the systematic collections community adopts a form of electronic invoicing. This would enable MVZ staff to monitor borrowed material at the item level without having to key in the bulk of those data.

(4) De-accessions

The purposes of documenting de-accession transactions are to ensure accountability for all MVZ collection material and to ensure that staff do not waste time searching for something that has not been misplaced, but rather has been removed from the collections by intention. The original catalog records are not deleted because information about the item may be useful even though the item is not physically present in the Museum. The system must support the following de-accession-related functions:

1. Update catalog record that item has been de-accessioned, indicating to whom (person and/or institution), date, type of de-accession (e.g., gift or exchange), and correspondent's name if applicable.
2. Link de-accession to shipping records, permits, and correspondence
3. Produce de-accession summary in the Annual Report.

(5) Shipping

The MVZ's need to document the shipping function is limited to outgoing shipments. Incoming shipments relevant to the transaction management function (accession, return of loan, and return of borrow) are treated directly as the represented transaction. The MVZ needs to document outgoing shipments because when a common carrier is used, the carrier has custody of material until it is transferred to the recipient.

The system should support this function by facilitating the following processes:

1. Print address labels, shipping invoice (different than loan invoice), separate cover letter and invoice.
2. Monitor shipments (produce reminder/query postcards if receipt of shipment is not acknowledged).
3. Link to permits; not just shipping permits but also "possession" permits.
4. Produce summary reports of shipping activity.
5. The system may be used to maintain on-line descriptions of carrier's practices, restrictions, and capabilities with regard to cost, weight, permits, and insurance.

c) Information Retrieval and Reporting

Under the current system, MVZ data are directly available only to the two Curatorial Associates. All requests for information, therefore, must be processed by one of these two individuals. The situation does not reflect the Museum's information access policy, but rather the expertise required to interact and extract information from the system. The replacement system should make as much of the MVZ's data as possible directly available to users, both local and remote, and should restrict access only in ways that reflect actual policy. Assistance by MVZ staff should be required only when a user lacks access to the Internet, needs assistance with a difficult problem, needs information that is not publicly available, or needs information that is not yet on-line.

We include under the information retrieval and reporting functions a variety of tasks and capabilities, ranging from simple canned reports, to query-by-example, a powerful query language, and the ability to link the "live" database to other applications such as GIS and statistical analysis.

(1) Ad-hoc query and reporting

It is probable that the most important use of the system will be for ad hoc query and reporting. This function will serve more users, more often than any other. It is therefore appropriate to give this function more attention than any other except the maintenance of data integrity and security.

1. The system must support simple form-based queries that allow the user to enter simple selection criteria and to choose from several report styles and output devices, including printers and files. The forms-based interface should:

- should allow the user to retrieve records as either formatted text or structured data, appropriate for importing into another application
- allow the user to browse/see the controlled vocabulary in appropriate fields
- should prevent "runaway" queries, but should also provide actual or estimated result statistics to the user, and in appropriate cases should allow the user to proceed (e.g., "Your query would have returned approximately *N* records in a file of *X* megabytes. Do you really want to do this?")

2. The system must provide an ad-hoc query and reporting facility that allows users to customize queries and reports, and presents familiar "information objects to the user while hiding the underlying complexity of the data structures. Intermediate results should be available for re-use within a session.
3. The system must support a standard query language (e.g., SQL) that enables technically sophisticated users to combine data in ways that were unanticipated by the programmers.

(2) Standard Reports

In addition to specimen cards, tags, and labels used in cataloging, and the invoices, mailing labels, form letters and reports described under Transaction Management, the system should contain a series of standard reports that support day to day collections management, as well as summary statistics used in annual reports and grant proposals. Information contained on these reports includes the following:

1. a summary of transaction activity, particularly accessions, loans, and deaccessions, with breakdowns by country and state (within the US);
2. a summary of cataloging activity within a user-specified period; e.g., number of specimens or objects cataloged, by collection;
3. the number of visitors and tours within a user-specified period
4. the number of service activities performed for the USFWS and California Department of Fish and Game;
5. a summary of requests for information received, with counts of specimen records delivered, including breakdowns by country and state (within the US);
6. a summary of specimens prepared, by collection;
7. a report on the growth and usage of the tissue collection;

8. a summary of MVZ publications within a user-specified period -- counts by staff-author, plus complete bibliography.

IV. Implementation Priorities

The functional requirements described above specify an information system that is, by museum standards, comprehensive and ambitious, and several years of sustained effort will be needed to bring it on line. The need for some capabilities, however, is relatively acute. It is essential, therefore, that development be conducted in a phased manner so that critical functions can be brought on line as quickly as possible. The following priorities will guide the development effort.

1. Main Catalog Applications. The highest priority functions are those now performed by TAXIR: basic cataloging and catalog maintenance (for all collections with catalogs now automated in TAXIR), and a suite of information retrieval and reporting capabilities for use with those catalogs. In addition to the mammal, bird, and herp collections, this effort must include at least minimal support for the tissue collection, egg and nest collection, and the Hildebrand collection, because these data must be removed from the IBM/CMS platform before it is decommissioned. The catalog applications should support the flow of material through the preparation processes (including the production of tags, labels, etc.), should provide mechanisms for maintaining authority files, and should enable staff to query the catalogs.
2. Transaction Management Application. The second priority will be to develop a suite of transaction management applications that allow the collection management staff to process the complete set of Museum transactions. The accession application may actually be developed in conjunction with the cataloging applications because this function is so intimately related to the initial capture of basic catalog data. The loan processing application will be developed next as this function represents the second most labor-intensive responsibility of the collections management staff.
3. Electronic Publication of Collections Information. The third implementation priority will be to make collection information directly available, via the Internet (World Wide Web), to external researchers and the public at large. Some of the infrastructure to support this function has already been put in place by the UC Berkeley, Museum Informatics Project. If the level of effort required to establish a simple query interface is sufficiently small, this capability may be established ahead of transaction management applications. A full-featured Web server for MVZ collections information will, nevertheless, come after the transaction management applications have been deployed.
4. Ancillary Collections Catalogs. The fourth implementation priority will be the development of automated catalogs for the ancillary collections, such as

the photographic and sound tape collections. Bringing these catalogs on line will involve both application development and data capture, as none of these data have been automated.

5. Support for Research and Education. Applications and capabilities developed to support research and education represent, in some sense, a "miscellaneous" category. These projects, however, are not necessarily of a lower priority than the ancillary collection catalogs. As the primary source of collections material and information, research applications and system capabilities that enhance the quality, and thus the long-term utility of collections information, should receive the highest priority after the basic collection management functions. Applications that can be developed rapidly and can help researchers to capture data efficiently and in a well structured form should be inserted in the development schedule as appropriate. Applications that support the Museum's education function most probably will not be considered until the Museum has acquired access to or developed its own Web-server capability. Once the Museum has this capability, however, the Museum's technical staff should be able to assist curators in developing such applications without expending a lot of time and postponing other development efforts.