

GCMRC

Database Management System
Requirements Analysis

Prepared by:

Grand Canyon Monitoring and Research Center

29 July 2003

Document Change Control

Revision Number	Date of Issue	Authors(s)	Description
1.0	19 November 2001	Raptor Technologies Inc. Vivid Solutions Inc.	
2.0	06 December 2001	GCMRC IT Staff	Re-written per comments of S. Mietz, S. Wyse, M. Liszewski, and C. Flaccus
2.1	12 February 2002	C. Flaccus	Revised Chapters 5-7
2.2	5 March 2002	C. Flaccus	Revised Chapters 5-7
2.3	7 March 2002 21 March 2002 4 April 2002 3 June 2002 5 June 2002	C. Flaccus	Revised Chapters 5-8 Proof Proof Proof Revised Chapters 4.2 – 7
2.4	7 June 2002 10 June 2002 17 October 2002	C. Flaccus C. Flaccus C. Flaccus	Revised per B. Ralston & S. Wyse Revised per L. Coggins Revised per R. Lambert, N. Horewer
2.5	14 July 2003 29 July 2003	C. Flaccus C. Flaccus	Revised data sets per PI interviews Revised data set inventory per PI interviews, and input from S. Mietz, M Breedlove, T. Gushue.

TABLE OF CONTENTS

TABLE OF CONTENTS	3
1. INTRODUCTION.....	8
1.1 BACKGROUND	8
1.2 PROJECT MISSION	9
1.3 DOCUMENT PURPOSE	9
1.4 RELEVANT DOCUMENTS	10
2. SYSTEM OVERVIEW	11
2.1 USER GROUPS	12
3. FUNCTIONAL REQUIREMENTS	14
3.1 DATA SET CONSOLIDATION.....	14
3.2 METHODOLOGY STANDARDIZATION	14
3.3 DATA SET ACCESSABILITY	14
3.3.1 Web Architecture	14
3.3.1.1 Web Based Attribute Query.....	14
3.3.1.2 Web Based Spatial Query.....	14
3.3.2 Intranet/Extranet Architecture	15
3.3.2.1 SQLNet Client Applications	15
3.3.2.2 ODBC Client Applications	15
3.3.2.3 GIS Applications	15
3.4 DATA SET DOWNLOAD CAPABILITY.....	15
3.5 DATA CAPTURE APPLICATIONS	16
3.6 DATA MAINTENANCE / DATA ADMINISTRATION APPLICATIONS	16
3.7 SECTION 508 COMPLIANCE	16
3.8 DOCUMENTATION	17
4. DATA INVENTORY AND REQUIREMENTS.....	18
4.1 DATA SET INVENTORY	18
4.1.1 Sediment Transport – Fine Grained.....	19
4.1.1.1 Gage Information.....	19
4.1.1.2 Suspended Sediment at USGS Gages	21
4.1.1.3 Historical Suspended Sediment at USGS Gages	23
4.1.1.4 Suspended Sediment Trip Based Sampling.....	24
4.1.1.5 Un-gaged Tributary Sediment Inputs.....	25
4.1.1.6 NAU Sandbar Surveys	25
4.1.2 Sediment Transport - Coarse Grained	27
4.1.2.1 Debris Fan Topography.....	27
4.1.2.2 Debris Flow Activity.....	28
4.1.2.3 Debris Flow Monitoring Photographs.....	29
4.1.2.4 Particle Point Counts	30
4.1.2.5 Matrix Samples.....	30
4.1.2.6 Ten Largest Boulders.....	31
4.1.2.7 Cs-137 Samples	32
4.1.2.8 C-14 Samples.....	33

4.1.2.9	<i>Boulder PIT Tags</i>	34
4.1.3	Hydrology	34
4.1.3.1	<i>Water Velocity Measurements</i>	34
4.1.3.2	<i>Dam Release Data (SCADA)</i>	35
4.1.3.3	<i>Dam Release Forecast</i>	36
4.1.3.4	<i>Stream Flow Gage Data</i>	37
4.1.3.5	<i>Stream Flow Gage Logs</i>	41
4.1.3.6	<i>Gage Photographs</i>	41
4.1.3.7	<i>Multi-beam Hydrography</i>	42
4.1.3.8	<i>Side-scan Sonar</i>	44
4.1.3.9	<i>Eyeball</i>	44
4.1.3.10	<i>Cross Sections</i>	45
4.1.3.11	<i>Stage – Discharge Relationships</i>	46
4.1.3.12	<i>Water Quality</i>	47
4.1.4	Climate	49
4.1.4.1	<i>Rainfall</i>	49
4.1.4.2	<i>NWS Weather Stations</i>	52
4.1.4.3	<i>Weather Radar</i>	53
4.1.4.4	<i>Air Temperature</i>	54
4.1.5	Terrestrial Biology	55
4.1.5.1	<i>Tribal Ethno Botany</i>	55
4.1.5.2	<i>Avifauna</i>	56
4.1.5.3	<i>Vegetation</i>	59
4.1.5.4	<i>Mammals</i>	66
4.1.5.5	<i>Invertebrates</i>	66
4.1.5.6	<i>Reptiles and Amphibians</i>	67
4.1.5.7	<i>Taxonomy</i>	68
4.1.6	4.1.6 Aquatic Biology	69
4.1.6.1	<i>Native and Exotic Fish</i>	69
4.1.6.2	<i>Fish Meristics</i>	74
4.1.6.3	<i>Fish Genetics</i>	75
4.1.6.4	<i>Fish Stomach Samples</i>	75
4.1.6.5	<i>Aquatic Food Base (Drift & Benthic)</i>	76
4.1.6.6	<i>Fish Food Base</i>	77
4.1.6.7	<i>Fish Elemental Analysis</i>	77
4.1.7	4.1.7 Cultural Resources	78
4.1.7.1	<i>Archaeology</i>	78
4.1.7.2	<i>Tribal Traditional Resources</i>	79
4.1.7.3	<i>Tribal Sacred Resources</i>	79
4.1.7.4	<i>Mitigation and Monitoring</i>	79
4.1.8	4.1.8 Recreation	79
4.1.8.1	<i>Campsites</i>	80
4.1.8.2	<i>User Attitudes</i>	80
4.1.8.3	<i>Safety</i>	81
4.1.9	Base Data	82
4.1.9.1	<i>Survey Control</i>	82
4.1.9.2	<i>Integrated Long Term Monitoring Reaches</i>	83
4.1.9.3	<i>USGS Quarter-Quad Boundaries</i>	83
4.1.9.4	<i>River Mile Index</i>	83
4.1.9.5	<i>LIDAR</i>	84
4.1.9.6	<i>Orthophotos</i>	84
4.1.9.7	<i>Aerial Photographs</i>	84
4.1.9.8	<i>Aerial Flight Line Index</i>	84
4.1.9.9	<i>High-Gain Aerial Photography</i>	84
4.1.9.10	<i>Cross Section Locations</i>	85
4.1.9.11	<i>USGS Permanent Gage Locations</i>	85
4.1.9.12	<i>GCMRC Stream and Weather Gage Locations</i>	85
4.1.9.13	<i>GCMRC Temperature Gage Locations</i>	85
4.1.9.14	<i>Geomorphic Reaches (Graff and Schmidt)</i>	85
4.1.9.15	<i>Geomorphic Reaches (Melis)</i>	86

4.1.9.16	<i>Sediment Yield Reaches</i>	86
4.1.9.17	<i>Sediment Yield Basins</i>	86
4.1.9.18	<i>HUC Boundaries</i>	86
4.1.9.19	<i>Shoreline Habitat (Korman)</i>	87
4.1.9.20	<i>Shoreline Habitat (Yard, Wyse and Mietz)</i>	87
4.1.9.21	<i>Fish Habitat Transect Locations (USFWS)</i>	87
4.1.9.22	<i>Stage Related Vegetation Transect Locations</i>	87
4.1.9.23	<i>Integrated Structure and Composition Transect Locations</i>	87
4.1.9.24	<i>Bird Patch Boundaries</i>	87
4.1.9.25	<i>Vegetation Boundaries</i>	88
4.1.9.26	<i>Kanab Amber Snail Habitat Boundaries</i>	88
4.1.9.27	<i>Campsites</i>	88
4.1.9.28	<i>Sand Bar Sites</i>	89
4.1.9.29	<i>Surficial Geology</i>	89
4.1.9.30	<i>Archaeological Site Boundaries</i>	89
4.1.10	Modeling	89
4.1.10.1	<i>Hydrologic Modeling - CRFSSGUI</i>	89
4.1.10.2	<i>Grand Canyon Ecosystem Model</i>	91
4.1.11	Imagery	91
4.1.11.1	<i>Imagery - Misc</i>	91
4.1.11.2	<i>Photography – Campsite</i>	92
4.1.11.3	<i>Photography – Stanton Repeat</i>	92
4.1.11.4	<i>Photography – Other Repeat</i>	92
4.1.12	Library	93
4.1.12.1	<i>Follet Card Catalog</i>	93
4.1.12.2	<i>Contacts</i>	94
4.1.13	Administration	94
4.1.13.1	<i>Administrative Agreement</i>	94
4.1.13.2	<i>Contacts</i>	95
4.1.13.3	<i>Reports</i>	96
4.2	GIS DATA STANDARDS	98
4.2.1	Coordinate system	98
4.2.2	Data Formats	98
4.2.3	Software Standards	99
4.3	DATA FILE ARCHIVE	99
5.	SECURITY REQUIREMENTS	99
5.1	GCMRC GUIDELINES FOR RELEASE OF DATA SETS	100
5.1.1	Provisional Data	100
5.1.2	Publicly Available	100
5.1.3	Restricted	100
5.2	MEDIA AND FORMAT FOR RELEASE OF GCMRC DATA	101
5.2.1	Internet	101
5.2.2	Removable Media	101
5.3	USER REGISTRATION	101
5.4	DATA ACCESS CONTROL	101
6.	HARDWARE AND SOFTWARE REQUIREMENTS	102
6.1	EXISTING INFRASTRUCTURE	102

6.1.1	Server Hardware.....	102
6.1.2	Server Software.....	102
6.1.3	Network Configuration	103
6.1.4	Firewall	103
6.1.4.1	VPN (Virtual Private Network).....	103
6.2	CLIENT SOFTWARE REQUIREMENTS.....	103
7.	HIGH LEVEL DESIGN	104
7.1	SYSTEM ARCHITECTURE.....	104
7.2	SERVER DEPLOYMENT.....	104
7.2.1	File And Database Server	105
7.2.1.1	Software Specifications.....	105
7.2.1.2	Hardware Specifications.....	105
7.2.2	Web Application/Internet Map Server	105
7.2.2.1	Software Specifications.....	105
7.2.2.2	Hardware Specifications.....	105
7.3	FUNCTIONAL ARCHITECTURE.....	106
7.3.1	Data Registry	106
7.3.1.1	Data Categorization.....	107
7.3.1.2	Index Map.....	107
7.3.1.3	Data Registry Web Application.....	108
7.3.2	Subject Area Database.....	109
7.3.2.1	Common Attribute Data	109
7.3.2.2	Spatial Data	111
7.3.3	File-Based Data Archive	112
7.3.4	Attribute Data Query Facility.....	112
7.3.4.1	Support offline analysis of data sets.....	112
7.3.5	Spatial Data Query Facility	113
7.3.5.1	Image Servers.....	113
7.3.5.2	Feature Servers.....	113
7.3.5.3	Spatial Query Application.....	113
7.3.6	Custom Web-Based Applications.....	114
7.3.7	Direct Query Facility	114
7.3.7.1	Access via ArcGIS tools.....	114
7.3.7.2	Access to Oracle data via SQLNet/ODBC (Read/Write).....	114
7.3.7.3	Programmatic Access to Oracle Data via a Web Service	114
7.3.8	Data Capture Applications	115
7.3.9	Administration Application	115
7.3.10	Data and System Administration Tools	115
7.4	SECURITY ARCHITECTURE.....	115
7.4.1	Web Applications.....	115
7.4.2	Data Administration Tools	116
7.4.3	Data Capture Applications	117
7.5	DESIGN PRIORITIES.....	118

- 8. APPENDIX I – ENTITY RELATIONSHIP DIAGRAMS (ERD) 119
- 9. APPENDIX II – ARCHIVE DIRECTORY STRUCTURE 133

1. INTRODUCTION

1.1 BACKGROUND

The Grand Canyon Monitoring and Research Center (GCMRC) studies the effects of the operation of Glen Canyon Dam (GCD) on downstream resources within the Glen Canyon National Recreation Area and Grand Canyon National Park. The scientific activities of GCMRC contribute to meeting the statutory requirements placed on the Secretary of the Interior by Congress via the 1992 Grand Canyon Protection Act, the 1995 Glen Canyon Dam Environmental Impact Statement, and the 1996 Record of Decision. Both in-house and external research experts, often in collaborative effort, perform these activities.

The GCMRC works within an Adaptive Management Program (AMP) that allows flexibility in Dam operations in response to increased understanding of ecosystem dynamics. The GCMRC works cooperatively with the Adaptive Management Work Group (AMWG) to define research objectives, and is advised with respect to information needs by the Technical Work Group (TWG). Both groups include representatives from federal and state resource management agencies, Native American tribes, and a diverse set of research experts. Additionally, Independent Review Panels (IRP) provide assessments of program proposals and accomplishments to ensure scientific objectivity and credibility.

In carrying out its mandate the GCMRC has accumulated valuable legacy data sets from a variety of sources. These data exist in a variety of formats and locations, which makes access to information difficult. This project is designed to create a system to catalog this disparate information, organize it and combine it in a systematic fashion, and provide information access to a variety of users in a common way.

In July 2001 Raptor Technologies Inc was awarded a contract to develop this system. In August and October interviews were carried out with GCMRC staff to evaluate existing data, and the information requirements of system users. The term **Data Management Framework** (DMF) was coined for the system during this time. This document will refer to the project as a DMF from this point forward. This document summarizes the results of these user interviews, along with other pertinent information, and provides a foundation upon which database development can proceed.

The GCMRC has elected to work collaboratively with the CDI (Center for Data Insight) at NAU (Northern Arizona University) for database development. Follow-on interviews with GCMRC staff and data providers will be conducted during July and August of 2002, in order to clarify needs identified in this document, and to work on completing a Pilot Phase of database development. Upon completion of the Pilot Project, needs and progress will be re-evaluated before proceeding with full database development. The requirements and goals of the Pilot Project will be discussed in Chapter 7.5 of this document.

1.2 PROJECT MISSION

A Mission Statement and Mission Goals have been formulated in order to guide and focus parties involved in the specification and design of the DMF. The Mission Statement establishes the purpose of this project, and provides focus for the developers of the system. The Mission Goals are used to support the defined mission statement and are used in determining and testing various aspects of the project specification and development.

Mission Statement

Develop an accessible, multi-disciplinary, spatially referenced, relational database to consolidate, organize, document, store, and distribute scientific information related to the Grand Canyon Ecosystem.

The Mission Goals are:

1. Identify and catalogue all possible data sets that may be included in the database.
2. Determine requirements for incorporating each data set.
3. Prioritize data sets for immediate inclusion.
4. Develop and program data structure.
5. Populate the database.
6. Develop distribution technologies for access by internal and external users.
7. Document database structure and maintenance.

1.3 DOCUMENT PURPOSE

The purpose of this document is to outline existing data collection activities of the Center, and to define system requirements of both data providers and data consumers. From these definitions technical design concepts will be evaluated, and a framework for detailed design and pilot project implementation will be produced.

To summarize, the goals of a Requirements Analysis are to:

1. Identify all potential users of the system
2. Determine the main functions that the DMF should provide to these users
3. Specify the necessary system outputs such as reports and analyses
4. Analyze the existing and anticipated data to be loaded into Oracle and/or GIS databases
5. Determine the Security requirements of the data stored in the system
6. Review existing GCMRC information systems infrastructure and standards
7. Prioritize system components for implementation

To accomplish these goals the document is divided into the following sections:

System Overview

Defines major components of the DMF.

Defines the internal and external user groups that will interact with the DMF.

Functional Requirements	Defines high-level functions that a DMF system must provide.
Data Requirements	Describes known data sets that must be included in the DMF. A logical data model is included as an Appendix, that defines these data in terms of entities, attributes and relationships.
Security Requirements	Outlines the access requirements of each user group listed in the system overview
Hardware and Software Requirements	Describes hardware and software constraints under which the DMF must work.
High Level Design	Details existing components and relationships. This design will be tested in an initial pilot project, and it will be subject to change as requirements are refined.

1.4 RELEVANT DOCUMENTS

Other documents that are complementary to this one include:

- GIS Legacy Data Report
- Survey Legacy Data Report
- Tabular Legacy Data Report
- [GCMRC Guidelines for Publication of Data Sets](#)
- [GCMRC Data Standards and Delivery Requirements](#)
- GCMRC Website www.gcmrc.gov
- GCMRC Score Report - [The State Of Natural And Cultural Resources In The Colorado River Ecosystem](#)
- GCMRC FY2002 Annual Plan
- GCMRC ftp site [ftp.gcmrc.gov](ftp://ftp.gcmrc.gov)
- [Northern Arizona University \(NAU\) Sandbar website](#)
- USGS National Water Information System [NWISWeb](#)
- [ESRI ArcIMS Security Whitepaper](#)

2. SYSTEM OVERVIEW

The DMF encompasses the following key components: Data Registry, GIS Base Data, and Subject Area Data sets. The framework is not limited to data warehoused in the Oracle Database, but contains all GCMRC data, including GIS coverages, shapefiles, offline electronic or paper resources, as well as links to other data providers. A definition of these components is as follows:

Data Registry: A Data Registry will be developed in order to provide an index to data sets that have been delivered to the GCMRC as results of scientific investigations. The Data Registry will help organize and archive GCMRC data sets, and will also provide a mechanism to query individual data sets using Web technology.

Subject Area Data sets: Each subject data set will be analyzed and modeled to determine whether or not there is sufficient consistency and value in creating a related tables within the Oracle Database. Once a data set's tables and relationships have been created, they will be populated with new and legacy information, allowing the information to be queried using Web or other client applications. Whenever possible these data sets will be spatially referenced, allowing access to information using a variety of ARC/INFO tools.

GIS Base Data: Other GIS resources such as orthophotos, LIDAR, and Mr. SID images are used commonly as base mapping tools in a variety of subjects. Because of their enormous sizes, and the manner in which they are used, it may not be appropriate to import all of these resources into Oracle tables. Instead, some of these may be left in an archive directory structure that is referenced by the Data Registry. GCMRC staff will investigate the applicability of using Oracle to store these data during the Pilot phase of this project.

2.1 USER GROUPS

The DMF will support diverse groups of users. Some, such as members of the general public, will simply be consumers of information within the system. Other groups will also be providers of information to the system, and will therefore require a greater degree of access to information within it. The Oracle database system allows for a sophisticated security model based on assigning individual users to “roles”. This security model will be discussed in more detail in Chapter 5. Various query and analysis needs, with respect to each data set, will be identified in Chapter 4, Data Inventory and Requirements. The following diagram and descriptions outline the user groups that have so far been identified to use this system, and the expected directions of information flow.

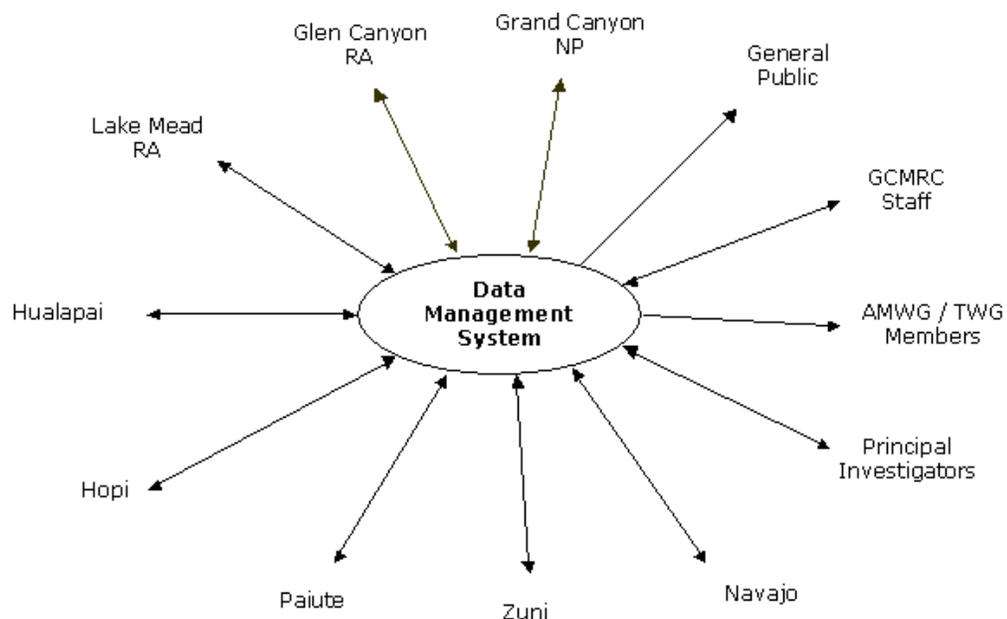


Figure 2-1 - System Users Diagram

GCMRC Internal Staff – Staff working on the GCMRC campus will have access to the DMF via an Intranet connection. They may also connect securely to the Intranet (termed an Extranet connection) from offsite locations using a VPN (Virtual Private Network) connection, or from the Internet using a Web browser. These users have specific analysis needs with respect to certain areas or subjects that may require read/write access to data from selected disciplines.

Principal Investigators (PI) – This group is comprised of approximately 30 individuals who also require both Intranet and Internet access to the DMF. These users also have

query and analysis needs specific to their own area, subject or project, but also require read access to other subject areas.

AMWG / TWG Members – The GCMRC works cooperatively with the Adaptive Management Work Group (AMWG) and the Technical Work Group (TWG) to define research objectives and information needs. Both groups include representatives from federal and state resource management agencies, Native American tribes, and a diverse set of research experts, who will be members of other user groups as well. AMWG/TWG group members will generally be consumers of information, and will therefore access the DMF via the Internet. As members who oversee how moneys are spent, their needs tend to be on Information Products, such as Final Reports, instead of detailed data sets like individual fish sample data. Interested AMWG and TWG members may also perform analyses on scientific data sets, however.

Individual Tribes - Cooperating Native American Tribes include the following:

- Hopi Tribe
- Hualapai Tribe
- Navajo Nation
- Pueblo of Zuni
- San Juan Southern Paiute Tribe
- Southern Utah Paiute Consortium:
 - Kaibab Band of Southern Paiutes
 - Paiute Indian Tribe of Utah
 - Shivwits Band of the Southern Paiute Nation

Each group potentially has information about physical, biological and archaeological resources to contribute to an integrated ecological database maintained by the GCMRC, and much of this data will be of exclusive value to the group providing it. Data “owner” and “visibility” fields must therefore flag any sensitive archaeological data that is contributed by individual groups. Its owners, through the database administrator, must have the ability to grant the level of accessibility to this type of information. A password, by secure VPN link, can be used with Oracle security to administer accessibly. Data owners may in turn permit or deny read only access to their data by other individuals or groups.

National Parks and Recreation Areas – Grand Canyon National Park, Glen Canyon National Recreation Area, and Lake Mead National Recreation Area personnel will each access the DMF via the Internet. Each may also require management access to their data via a more secure connection such as VPN. There are specific ownership and security issues that will surround data provided by each of these management entities, depending on what information is to be provided, and whether or not each group wishes to retain management and access rights to the data. The individual data management and access needs of these user groups are similar in nature to those of individual tribes.

General Public – The general public has a right to certain types of information collected by federally managed programs. Notable exceptions to this are in cases where that information concerns archaeological sites or endangered species, or when it is deemed provisional in nature. The specific policies of the GCMRC with respect to these data are outlined in the document “GCMRC Guidelines for the Publication of Data sets”. The information needs of the public are presumed to be quite varied, and will be accommodated as they are encountered. The general public will have read access to the system via the Internet. Certain cultural and endangered species data, that have been identified and coded as sensitive, will be unavailable to the public.

3. FUNCTIONAL REQUIREMENTS

3.1 DATA SET CONSOLIDATION

Prior to the formation of GCMRC, a loosely organized system existed for dealing with monitoring and research data sets. Each Program and PI had different ways of delivering information products. This resulted in individual data sets, documentation and final reports that were delivered in a variety of formats and stored in a number of locations. The resulting array of data sets prevented the kind of integrated data analysis that can be afforded by a single comprehensive database system.

The GCMRC has taken steps to document, catalogue, standardize, and organize both legacy and newly created data sets. It has done so by developing data standards, archive structures, and consolidated data storage facilities. To further this effort, the following architectural elements are required of the DMF:

- File system application architecture to store and retrieve archived deliverables
- An Oracle database to warehouse data
- Data Registry framework to catalogue and store meta-data for applicable data
- Spatial data storage for GIS data sets

3.2 METHODOLOGY STANDARDIZATION

In order to refine and standardize data collection methods for ecosystem monitoring, Protocol Evaluation Panels have been conducted for most program areas. While these panels have standardized how modern monitoring data sets are collected, legacy data sets pose unique problems to the process of consolidation. Each existing data set must be analyzed for common entities and attributes, reduced to "common denominator" status, and finally combined into a manageable integrated database.

A benefit of this process is that it helps enforce consistent data collection efforts in the future, as is required by a relational database model.

3.3 DATA SET ACCESSABILITY

3.3.1 Web Architecture

Through the use of client browsers and web servers, GCMRC data sets can be published to a broad and diverse group of users, each with their own specific data requirements. The use of Web technology will ensure the broadest audience with the least support cost. This will be the database access mechanism presented to the general public.

3.3.1.1 Web Based Attribute Query

The ability to query textual attributes via the web is required. This can be achieved by using either a Web-based query form, or an ArcIMS-based spatial query. Depending on the requirements of the user, both may be valuable mechanisms to select information.

3.3.1.2 Web Based Spatial Query

Most GCMRC data sets are spatial in nature, and could be queried with a Web-based spatial front end. A Web-based map-publishing tool, such as ESRI's ArcIMS, is required for this functionality.

Use of this technology will allow GCMRC to dynamically serve maps to the Internet, and enable users to pan, zoom, identify, and spatially query certain data sets, thereby providing an intuitive and highly visual environment for exploring information related to the Grand Canyon.

3.3.2 Intranet/Extranet Architecture

Some user groups require higher levels of access to both spatial and tabular information than a browser application can provide. These include certain GCMRC staff, which may be responsible for maintaining portions of the database, or others who may require access to secure information within the database. Oracle Security will control an individual's access to these data on the GCMRC Intranet, depending on the needs of the user or group. VPN (Virtual Private Network) connections to the database are also considered such secure connections to the Intranet.

3.3.2.1 SQLNet Client Applications

SQL Plus and Oracle Discoverer, among others, are client applications that will be allowed to connect to the Oracle database via the Oracle product SQLNet. Knowledgeable Oracle users will be able to perform ad-hoc queries of tabular information in order to answer specific questions, edit data, or select subsets for analysis. Views and stored procedures will be used to simplify the display of complex relational structures.

3.3.2.2 ODBC Client Applications

Custom data entry and reporting forms will allow users at GCMRC to look at data within Oracle in ways that are familiar to them, using software applications that they commonly use. Client software tools such as MS Access, MS Excel and SAS software can connect to Oracle via ODBC (Open Database Connectivity), and provide familiar user-friendly front ends to the database. These will also allow certain GCMRC staff to more easily manage specific parts of the Oracle database for which they are responsible.

3.3.2.3 GIS Applications

GCMRC staff and principal investigators using the GCMRC Intranet will be required to manage spatial components of the database, and to do sophisticated GIS analyses. To accomplish these tasks, connectivity is required using the full suite of ArcGIS tools, including ARC/INFO Workstation, ArcCatalog, ArcMap and ArcView. ArcSDE will be employed, which will allow an ARC/INFO GIS to be built on an Oracle Spatially enabled database, and allow for a tight integration of spatial and non-spatial data sets. Connections to the spatial database will be managed through ArcCatalog.

3.4 DATA SET DOWNLOAD CAPABILITY

It would be impossible to write a single generalized application interface to support all conceivable analyses of data. Simply put, there are an infinite number of ways to present data, depending on the scientific questions being asked. Users are also typically familiar with specific tools for analyzing and reporting on data sets. It is therefore not required to have complex analysis tools built into a Web interface to the DMF. The ability to extract data from a relational structure, either as comma delimited ASCII text files or as ARC/INFO

export files, is considered required functionality for any application interface to the DMF. The user may then import these extracts into the analysis or reporting application of their choice.

The priority of this system development is to deal with new data sets, and to put these data sets into a relational structure in Oracle. Legacy data sets will be analyzed, prioritized and imported into these structures when practical. However, there may be legacy data sets that, because of their unique nature, do not lend themselves to consolidation into a relational database. These may be better left in original form, in an archive, provided that they can be retrieved by some query methodology. Internet users must be able to search the data archive for relevant data sets to download as CSV ASCII text files.

3.5 DATA CAPTURE APPLICATIONS

Some program areas have specialized field data capture needs that could be automated with the use of custom applications, running on small, rugged field computers such as PalmPCs, PocketPC, or laptop computers. These applications would allow electronic data collection in the field without the use of traditional paper data entry forms. Using these automated techniques will allow field data to be uploaded directly to the Oracle database at a later time, thus minimizing data entry errors. Potential data capture applications will be identified in section 7.3.8.

3.6 DATA MAINTENANCE / DATA ADMINISTRATION APPLICATIONS

The DMF system must be built with internal (GMCRC) staff maintenance of the application(s) in mind. New data sets and data types will be added in the future, and this must be accomplished without having to significantly revise programming code.

A Data Maintenance application is required to provide for the daily administration of the DMF. Some of these administrative tasks include:

- Adding new Information Sources to the Data registry.
- Administering Data Registry Security for user groups and individuals.
- Verifying links between information sources and linked resources.
- Automating repetitive management tasks such as database backup and restoration.
- Maintaining domain tables in the Oracle database
- Importing data sets into the Oracle database.
- Running data capture applications.

3.7 SECTION 508 COMPLIANCE

[Section 508](#) requires that when Federal agencies develop, procure, maintain, or use electronic and information technology (EIT), Federal employees with disabilities have comparable access to information and data as Federal employees who have no disabilities, unless an undue burden would be imposed on the agency. Section 508 also requires that individuals with disabilities who are members of the public seeking information or services from a Federal agency have comparable access to and use of information and data as the public without disabilities, unless an undue burden would be imposed on the agency.

Federal agencies have an explicit statutory obligation to make all EIT that they develop, maintain or use compliant with Section 508. This system must address this requirement.

3.8 DOCUMENTATION

The design and implementation of the DMF must be thoroughly documented, and included as data sources. Methodologies used in consolidating data sets, defining and populating database tables, and database structure are all considered meta-data that must be accessible through the DMF.

4. DATA INVENTORY AND REQUIREMENTS

This section identifies existing GCMRC data sets that must be registered and accessible through the DMF system. Included in this summary are potential uses of the data, and requirements of those users of data. As research and monitoring requirements are refined in the future, these requirements may be modified. All custom software components that deal with data must be designed to allow for a level of flexibility within project budget and time constraints. Refer to Appendix I, Entity Relationship Diagrams, as these form integral parts of this section.

4.1 DATA SET INVENTORY

The concept of distinct Physical, Biological and Cultural data sets has been discarded as an organizational structure that inadequately reflects the interdisciplinary approach to studies done at GCMRC. Areas and Subjects are considered attributes that better define each scientific data set collected, for any study performed, in support of monitoring or research activities at the GCMRC. These categories facilitate combining like information into common database tables and spatial feature sets, and also provide ways to search for information based on attributes of a particular study or scientific data set. It is expected that the Area/Subject designation will ultimately translate into "Object and Feature Data set" names within the ArcSDE Geodatabase schema. Nothing in this model precludes adding or removing categories in the future. The Area and Subject categorizations that have been identified at GCMRC to date are as follows:

Area	Subject (Object/Feature Data set)
Sediment Transport	Fine grained Coarse grained
Hydrology	Stream flow Channel geometry Water quality (chemical, physical, biological)
Climate	Rainfall Weather Service Air temperature
Terrestrial Biology	Tribal ethnobotany Avifauna Kanab Amber Snail Vegetation Mammal Invertebrates Reptiles & amphibians
Aquatic Biology	Native and exotic fish Fish genetics Aquatic food base (drift & benthic)
Cultural Resources	Archaeology Tribal traditional resources Tribal sacred resources Mitigation/Monitoring
Recreation	Campsite condition User attitudes and satisfaction

	Angler information Safety
Base Data	Imagery Topography Index maps Geomorphology Survey control
Modeling	Hydrology Ecosystem

In the following sections are lists of data sets that were collected during research and monitoring activities in the Grand Canyon. These data sets are either located at GCMRC, or at participating institutions.

A particular data set must fit into at least one subject. A data set might also have been collected in support of research covering a number of different subject areas.

In many cases, several data sets may be combined into in single entity that can be modeled in a set of related Oracle database tables. Examples include fish sampling, or water quality sampling.

4.1.1 Sediment Transport - Fine Grained

4.1.1.1 Gage Information

Data set Name	Stream Gages																
Data set Area	Sediment Transport, Hydrology																
Data set Subject	Fine Grained, Stream Flow, Water Quality																
GCMRC Contact	Ted Melis																
Principal Investigator(s)	David Topping, Nancy Hornewer																
Designer ERD/Entity Names(s)	Physical/Phys Sediment																
Description of Project	<p>This data set refers to descriptive information about 60+ temporary and permanent USGS and GCMRC gages that collect stream flow and sediment transport data. Information about each gage includes a station name, number, and location. Gages may be activated and abandoned several times throughout their history.</p> <p>Information from gages within the USGS Arizona District is available from the ADAPS system. River flow information collected at these stations constitutes the Stream Flow Gage Data set discussed in section 4.1.3.4.</p> <p>Several hydrologic parameters are measured at eight "permanent" USGS gage locations in the Colorado River and selected tributaries. A list of these gages is as follows:</p> <table border="1"> <thead> <tr> <th>Station #</th> <th>Station Name</th> </tr> </thead> <tbody> <tr> <td>09379910</td> <td>Colorado River below Glen Canyon Dam, AZ</td> </tr> <tr> <td>09402000</td> <td>Little Colorado R. near Cameron, AZ</td> </tr> <tr> <td>09404200</td> <td>Colorado R. above Diamond Creek near Peach Springs, AZ</td> </tr> <tr> <td>09383100</td> <td>Colorado R. above Little Colorado R. near Desert View, AZ</td> </tr> <tr> <td>09380000</td> <td>Colorado R. @Lees Ferry, AZ</td> </tr> <tr> <td>09402500</td> <td>Colorado R. near Grand Canyon, AZ</td> </tr> <tr> <td>09401200</td> <td>Little Colorado R. @Cameron, AZ bridge</td> </tr> </tbody> </table>	Station #	Station Name	09379910	Colorado River below Glen Canyon Dam, AZ	09402000	Little Colorado R. near Cameron, AZ	09404200	Colorado R. above Diamond Creek near Peach Springs, AZ	09383100	Colorado R. above Little Colorado R. near Desert View, AZ	09380000	Colorado R. @Lees Ferry, AZ	09402500	Colorado R. near Grand Canyon, AZ	09401200	Little Colorado R. @Cameron, AZ bridge
Station #	Station Name																
09379910	Colorado River below Glen Canyon Dam, AZ																
09402000	Little Colorado R. near Cameron, AZ																
09404200	Colorado R. above Diamond Creek near Peach Springs, AZ																
09383100	Colorado R. above Little Colorado R. near Desert View, AZ																
09380000	Colorado R. @Lees Ferry, AZ																
09402500	Colorado R. near Grand Canyon, AZ																
09401200	Little Colorado R. @Cameron, AZ bridge																

	<p>Stream flow, sediment concentration and rainfall gages have been located by GCMRC in selected lesser tributaries to the Colorado, as well as in selected locations in the main stem. These gages do not have USGS station numbers, nor do they contribute data to the USGS system. Data from these gages must be downloaded manually to a PC. Information collected at GCMRC gages will only be available through the DMF. In the future, information from GCMRC gages will be downloaded to GCMRC by a telemetry system, thereby eliminating the need for manual intervention.</p>	
Description of Data sets	Data Format	CSV
	Quantity of Data	
	Location of Data	<p>USGS gages: ADAPS system with copies located at GCMRC.</p> <p>GCMRC gages: GCMRC</p>
	Major Tasks	<ul style="list-style-type: none"> • Define Oracle structure • Import CSV files into Oracle • Define data access queries
	Time required to complete tasks	
	List of Entities & Attributes	<p>Gages</p> <ul style="list-style-type: none"> • Station_id • Station_type • USGS_Station_number • Station_name • Operator • X_COORD • Y_COORD • Elevation • Description • Active Date • Abandoned Date • Parameters measured
	Search Keys and tables	
Current Collection Method	<p>Some gages have been intermittently operated, making active and abandoned dates important parameters.</p> <p>USGS gages send water temperature (intermittently), turbidity, conductivity, and stage discharge information via telemetry. In addition, water samples are collected along cableways to yield at depth and sent to a lab for analysis. At certain gages, a time series laser analysis of sediment grain sizes is also performed in situ using a LISST system. Data from the LISST is now manually downloaded to a PC on an infrequent basis. Data from GCMRC gages must be downloaded manually to a PC, though telemetry will be employed in the near future.</p> <p>Resulting lab and laser sample analyses described in the "Suspended Sediment at USGS Gages" data set.</p>	
Future Collection Method	LISST data will be downloaded via telemetry.	
Data Analysis		
Data Query	1) Public, query gage descriptive and location information.	

GCMRC Archives? Yes
 Metadata? Yes
 GCMRC Pilot? Yes

4.1.1.2 Suspended Sediment at USGS and GCMRC Gages

Data set Name	Suspended Sediment at Gages	
Data set Area	Sediment Transport	
Data set Subject	Fine Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	David Topping	
Designer ERD/Entity Names(s)	Physical/Phys Sediment	
Description of Project	<p>The concentration and size distribution of suspended sediment particles is measured at both USGS and GCMRC gage locations in the Colorado River and selected tributaries. At USGS gage cable crossings; measurements are made on water samples that are manually collected at various points in the river. Samples are transported to a lab where analyses are made.</p> <p>A time series laser analysis of sediment grain sizes is also performed in situ using a LISST system at the Grand Canyon gage site.</p>	
Description of Data sets	Data Format	Ingress, MS Excel
	Dates of Collection	CRabvDC 11/4/1997 – 4/2/2003 CRabvLCR 8/28/1997 – 3/29/2003 CRatLF 11/3/1997 – 3/8/2001 CRnearRM30 9/4/2000 – 3/28/2003 CRnearRM60 3/28/2002 – 3/30/2003 CRnearGC 11/3/1997 – 4/1/2003 LCRatCAM 9/2/1999 – 4/17/2003 LCRnearCAM 8/3/1999 – 3/26/2003 PARIAatLF 8/10/1997 – 4/17/2003 PRnearKAN 9/8/2002 – 2/26/2003 LESSER_TRIB 3/10/2001 – 8/8/2001
	Quantity of Data	11 Files
	Location of Data	Some at USGS, Some at GCMRC
	Major Tasks	1) Define Oracle structure 2) Import GCMRC files to Oracle 3) Download remaining data from USGS ADAPS. 4) Import remaining data to Oracle.
	Time required to complete tasks	To be completed in Pilot, August 2003

	<p>List of Entities & Attributes</p>	<p>Sediment Samples</p> <ul style="list-style-type: none"> • Station_id • Station_type • Date/Time • Trip_id • Sample_method • Sample_number • Depth • Sampling time • Silt&Clay concentration (mg/l) • Sand concentration (mg/l) • D16 (mm) • D50 (mm) • D84 (mm) <p>Sand (<0.074 mm)</p> <p>Sand (<0.088 mm)</p> <p>Sand (<0.105 mm)</p> <p>Sand (<0.125 mm)</p> <p>Sand (<0.149 mm)</p> <p>Sand (<0.177 mm)</p> <p>Sand (<0.210 mm)</p> <p>Sand (<0.250 mm)</p> <p>Sand (<0.297 mm)</p> <p>Sand (< 0.354 mm)</p> <p>Sand (<0.420 mm)</p> <p>Sand (<0.500 mm)</p> <p>Sand (<0.595 mm)</p> <p>Sand (<0.707 mm)</p> <p>Sand (<0.841 mm)</p> <p>Sand (<1.0 mm)</p> <p>Sample Wt (g)</p> <p>Weight passing through 63 (g)</p> <p>Weight retained on 63 (g)</p> <p>Concentration<63 (mg/l)</p> <p>Concentration>63 (mg/l)</p> <p>Concentration analysis lab</p> <p>Grain Size analysis lab</p>
	<p>Search Keys and tables</p>	
<p>Current Collection Method</p>	<p>Surface sampling conditions such as water and air temperature are collected at some sites, as well as river discharge, and sent by telemetry to a USGS computer system. Water samples are taken manually at depth at various points across the river, from a cable car crossing.</p> <p>Water samples are sent to a lab for grain size analysis.</p> <p>A time series laser analysis of sediment grain sizes is also performed in-situ using a LISST system. These data require calibration against D and P measurements described below. Data from the LISST is now manually downloaded to a PC on an infrequent basis.</p> <p>Other methods used to collect suspended sediment information include:</p> <p>D77 – Depth integrating (1977)</p> <p>D-96 – Depth integrating (1996)</p> <p>P-61 – Point sample taken at depth</p>	

	ISCO – Point pump sampler NORTEK-EASYQ – Acoustic backscatter sampler Resulting lab and laser sample analyses are input to a Fortran program authored by D. Topping for further processing. Results are entered into an MS Excel spreadsheet. This process may change at some point in the future.
Future Collection Method	LISST data will be downloaded via telemetry. At some point in the future in-situ measurements may replace water samples collected manually.
Data Analysis	
Data Query	1) Public, query of sediment tables by gage location and date range. 2) Public, extract and download of CSV files of sediment tables by gage location and date range.

GCMRC Archives? Yes
Metadata? No
GCMRC Pilot? Yes

4.1.1.3 Historical Suspended Sediment at USGS Gages

Data set Name	Historical Suspended Sediment at USGS Gages	
Data set Area	Sediment Transport	
Data set Subject	Fine Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	David Topping	
Designer ERD / Entity Names(s)	Physical/Phys Sediment	
Description of Project	Historically, the concentration and size distribution of suspended sediment particles was measured at permanent USGS gage locations in the Colorado River and selected tributaries. These gage locations are listed above. Structure of data set is similar to the data currently collected at gage locations.	
Description of Data sets	Data Format	TSV, CSV
	Quantity of Data	
	Dates of Collection	Intermittent. Prior to 1984.
	Location of Data	USGS ADAPS
	Major Tasks	<ul style="list-style-type: none"> • Download data from ADAPS • Examine data structure • Re-design Oracle structure if necessary • Import legacy data into Oracle
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query	1) Public, query of historic sediment tables by gage location and date range. 2) Public, extract and download as CSV files historic sediment tables by gage location and date range.	

GCMRC Archives? No
Metadata? No
GCMRC Pilot? No

4.1.1.4 Suspended Sediment Trip Based Sampling

Data set Name	Suspended Sediment Trip Based Sampling	
Data set Area	Sediment Transport	
Data set Subject	Fine Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Nancy Hornewer	
Designer ERD / Entity Names(s)	Physical/Phys Sediment	
Description of Project	<p>The concentration and size distribution of suspended sediment particles is measured at trip based sample locations in the Colorado River. Trip based sampling is done from watercraft, not at fixed gage locations.</p> <p>Trip based sampling methods include the following:</p> <ul style="list-style-type: none"> • Depth-integrated samples from the center of the River • Surface dip samples from the center of the river • Point samples from the center of the river • Surface dip samples from local eddies <p>Some sample sites are repeated year after year. Some sample sites are selected based on other factors, and are generally not repeated year after year. In either case, the methods of processing the water sample are the same.</p>	
Description of Data set	Data Format	Ingress, CSV ASCII text
	Quantity of Data	
	Dates of Collection	September 1999 trip July 2000 trip May 2000 trip September 2000 trip
	Location of Data	GCMRC Archive
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	<ul style="list-style-type: none"> • RiverMile • Notes • Method • Date • Time • Grain size distribution.
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query	<ol style="list-style-type: none"> 1) Public, query of sediment tables by gage location and date range. 2) Public, extract and download of CSV files of sediment tables by gage location and date range. 3) Public, Report, Sediment load as a function of discharge. 4) Public, Report, Average sediment load by month by gage location. 5) PI, Application, Import sediment distribution from Excel spreadsheet to Oracle Sediment and Sampling Conditions tables. 	

GCMRC Archives? Yes

Metadata? No
GCMRC Pilot? Yes

4.1.1.5 Un-gaged Tributary Sediment Inputs

Data set Name	Un-gaged Tributary Sediment Inputs	
Data set Area	Sediment Transport	
Data set Subject	Fine Grained, Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)	Physical/Phys Sediment Inputs	
Description of Project	Estimates of sediment inputs from 700 un-gaged tributaries of the Colorado River. These are based on various field and modeling activities.	
Description of Data set	Data Format	CSV ASCII text
	Quantity of Data	
	Location of Data	USGS Tucson
	Major Tasks	1) Define Oracle structure 2) Input data
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis	Use estimates of sediment input to predict sediment budget of Colorado River.	
Data Query	<ol style="list-style-type: none"> 1) Public, query of sediment input tables by tributary and date range. 2) Public, extract and download of CSV files of sediment input tables by tributary and date range. 3) Public, Report, Sediment load as a function of rainfall event. 4) Public, Report, Average sediment load by month by tributary. 5) PI, Application, Import estimated sediment loads to predictive model. 	

4.1.1.6 NAU Sandbar Surveys

Data set Name	NAU Sandbar Surveys
Data set Area	Sediment Transport
Data set Subject	Fine Grained
GCMRC Contact	Ted Melis
Principal Investigator(s)	Rod Parnell, Matt Kaplinski, Joe Hazel
Designer ERD / Entity Names(s)	PHYSICAL / PHYS SANDBAR
Description of Project	<p>Since 1990, topographic and hydrographic surveys have been repeated at 37 long-term study sites, to monitor the accumulation and movement of fine-grained sediment throughout the Colorado River ecosystem.</p> <p>Topographic surveys are used to determine the total volume and spatial distribution of sediment stored in sand bars. Hydrographic surveys are used to determine these parameters within the channel margins.</p>

	Sand deposition provides habitat for both aquatic and terrestrial species along the river corridor. Sand bars deposited within eddies are also the primary campsites for rafting and hiking groups. Understanding how these habitats change with river flow regimes is critical to predicting the effects of GCD operations.	
Description of Data sets	Data Format	CSV ASCII text, Excel spreadsheet,
	Dates of Collection	1990 to present
	Quantity of Data	
	Location of Data	NAU Geology Dept.
	Major Tasks	Get CSV data from NAU Define Oracle structure Import CSV data files
	Time required to complete tasks	1 weekl
	List of Entities & Attributes	<ul style="list-style-type: none"> • RM • Date • Point# • StateX • StateY • Elev • PointType
	Search Keys and tables	
Current Collection Method	<p>Both topographic and hydrographic surveys use digital data collectors in fieldwork. Outputs are used to generate point surface elevation files.</p> <ul style="list-style-type: none"> • Point files contain point number, x, y, z, and point code values. The code defines whether the point is a slope, break, boundary or waters edge record. • Slope break, perimeter, and waters edge lines provide boundary conditions for TIN creation • Most topographic surveys are done in local coordinate systems. Data sets are later projected to State Plan coordinates. • Data sets are stored in Terramodel format at NAU. Deliveries to GCMRC are as CSV ASCII text files <p>A detailed description of NAU's Sandbar Survey study methods can be found at http://vishnu.glg.nau.edu/gcmrc/methods.html</p>	
Future Collection Method	<p>A new long-term monitoring program will study debris fan/eddy complexes, which consisting of multiple eddies over 1 – 3 mile stretches of the river. Because of the length of each stretch, complete hydrographic surveys of all such complexes would be impractical. This study effort is newly proposed, and a complete data model is undefined at this time.</p>	
Data Import Tool	<p>A mechanism will be needed to import TerraModel format topographic and hydrographic survey information into the DMF.</p>	
Data Analysis	<p>Topographic and hydrographic survey points are combined and used to create a TIN (Triangulated Irregular Network) surface model of the channel, eddy, and sand bar topography. This TIN surface model is then used to generate cross sectional profiles, and comparison maps. Area and volume calculations are performed for sand contained in beaches, eddies, and channel margins, based on fixed eddy and channel boundaries. NAU records these volume calculations in Excel spreadsheets.</p> <p>Stage/Discharge relation - Stage-discharge relationships are determined for each site and used to predict the water's edge at specific flows.</p> <p>Survey data analyses are done in conjunction with reach-scale aerial photogrammetric mapping, flow modeling, and suspended sediment sampling. Data analyses also require access to USGS gage station discharge values, typically daily max, min, and mean. The source and quality of stream flow data used by NAU is unclear, and access to a consistent data set maintained at GCMRC would be a good thing.</p>	

Data Query	<ol style="list-style-type: none"> 1) Extract, as ASCII CSV, raw surface data for a given site, on a given date. 2) Download a TIN export file of a site surface, for a given date. 3) Look at cross sectional profiles of river channel and beaches by year. 4) Select volume results, by site, by year.
Priority	

GCMRC Archives? No

Metadata? No

GCMRC Pilot? No

4.1.2 Sediment Transport - Coarse Grained

Coarse Sediments are those with intermediate particle diameters greater than 64 mm. The deposition of these sediments at tributary junctions builds large debris bars that constrict the river to form rapids. Debris fans, and their resulting debris bars, provide stable substrate for aquatic organisms, and are primary components of the Colorado River Ecosystem.

A USGS Fact Sheet entitled "Monitoring of Coarse Sediment Inputs to the Colorado River in Grand Canyon", FS019-01 provides an overview of aspects of the Coarse Sediment monitoring program.

4.1.2.1 Debris Fan Topography

Data set Name	Debris Fan Topography	
Data set Area	Sediment Transport, Base Data	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD / Entity Names(s)		
Description of Project	<p>Topographic maps of debris fan geometry, prepared using one or more of the following methods:</p> <ul style="list-style-type: none"> • Conventional total station surveys • Modern orthophotographic methods. • Historic aerial photography (not rectified). • Integration of channel bathymetry from single and multi beam sonar. 	
Description of Data sets	Data Format	Unknown
	Dates of Collection	Unknown
	Quantity of Data	Unknown
	Location of Data	USGS, Tucson
	Major Tasks	Receive data from PI
	Time required to complete tasks	Unknown until data delivered
	List of Entities & Attributes	Unknown
	Search Keys and tables	Unknown
Current Collection Method	<ol style="list-style-type: none"> 1) Local control points are established for conventional total station surveys and ortho-rectification of aerial photography. 2) Large boulders and features identifiable from aerial photography are tied into control using conventional total station survey methodology. 3) Complete topographic survey of debris fans are conducted using total station survey methodology. 	

	4) Complete topographic maps of debris fans are prepared using survey data, or ortho-rectified aerial photographic data.
Future Collection Method	Topo maps may be created using ERDAS imaging software on high-resolution aerial photography. This will reduce the use of more expensive ground survey techniques.
Data Analysis	1) Debris fan area, volume, composition and distribution as a function of year.
Data Query	
Priority	

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.2.2 Debris Flow Activity

Data set Name	Debris Flow Activity	
Data set Area	Sediment Transport	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD / Entity Names(s)		
Description of Project	A descriptive inventory of debris flow events in the Grand Canyon that could be represented in a summary table that gives general information about each discrete debris flow event. Events are identified either by observation, radio carbon dating vegetative units, or other stratigraphic techniques.	
Description of Data sets	Data Format	Unknown
	Dates of Collection	Unknown
	Quantity of Data	Unknown
	Location of Data	USGS, Tucson
	Major Tasks	Receive data from PI
	Time required to complete tasks	Unknown
	List of Entities & Attributes	Debris_flow_activity <ul style="list-style-type: none"> • Station_id • Station location description • Date of Debris Flow • Reach river (Y/N) • Affect rapid (Y/N) • Event source description • Identification method
Search Keys and tables		
Current Collection Method	Observation, C14, CS137, Stratigraphic	
Future Collection Method	Same	
Data Analysis		
Data Query	1) Query debris flow events for a given tributary canyon. 2) Query debris flow activity by year.	
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.2.3 Debris Flow Monitoring Photographs

Data set Name	Monitoring Photographs	
Data set Area	Sediment Transport, Terrestrial Biology	
Data set Subject	Coarse Grained, Vegetation	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD / Entity Names(s)		
Description of Project	<p>Monitoring photographs were established at 23 different sites during 2001, by matching historic photographs obtained during the past 100 years. These photos will be repeated annually in order to monitor changes in debris fan activity.</p> <p>These photographic sequences are also be used to monitor changes in terrestrial vegetation.</p>	
Description of Data sets	Data Format	TIFF
	Dates of Collection	1890 to present, randomly collected
	Quantity of Data	<ul style="list-style-type: none"> 25 historic photographs at 23 different sites. Yearly repeat photography of these 25 historic photographs.
	Location of Data	USGS, Tucson
	Major Tasks	<ul style="list-style-type: none"> Acquire digital photography from PI Design Oracle structure Import photography and attributes into Oracle.
	Time required to complete tasks	Unkown
	List of Entities & Attributes	<p>Monitoring Photographs</p> <ul style="list-style-type: none"> Event_id Station_id Trip_id Date Description Camera_Y Camera_Y Camera_Azimuth Photograph Comments
	Search Keys and tables	
Current Collection Method	Repeat photography, using exact camera location and azimuth of exposure.	
Future Collection Method	Same	
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No

GCMRC Pilot? No

4.1.2.4 Particle Point Counts

Data set Name	Particle Point Counts	
Data set Area	Sediment Transport	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)		
Description of Project	<p>Particle point counts, using the Wolman Method, are used to quantify matrix composition of sediments on debris fans in the Grand Canyon. Point counts were taken at 20 newly deposited debris fans during 2001, at water's edge, and at 1m from water's edge.</p> <p>Point counts under-represent both large boulders and smallest sediments. The 10 Largest boulders and Matrix samples techniques represent the distributions of these sizes, respectively.</p>	
Description of Data sets	Data Format	Unknown
	Dates of Collection	2001
	Quantity of Data	Unknown
	Location of Data	USGS, Tucson
	Major Tasks	Receive data from PI
	Time required to complete tasks	Unknown
	List of Entities & Attributes	<p>Particle Size Distributions</p> <ul style="list-style-type: none"> • Station_id • Trip_id • Date • X_COORD • Y_COORD • Number_Size1 • Number_Size2 • Number_Size3 • Etc • Description
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.2.5 Matrix Samples

Data set Name	Matrix Samples
----------------------	----------------

Data set Area	Sediment Transport	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)		
Description of Project	<p>Sediment point counts do not accurately measure particle size distributions at the fine end of the size spectrum.</p> <p>Sediment matrix samples were collected from selected debris fans during 2001, and used to determine particle size distributions and reconstituted water content of fine sediments in laboratory analysis. Matrix samples are collect at water's edge, and at 1m from water's edge.</p>	
Description of Data sets	Data Format	TIF and JPEG
	Dates of Collection	2001
	Quantity of Data	Unknown
	Location of Data	USGS, Tucson
	Major Tasks	Acquire data from PI
	Time required to complete tasks	Unknown
	List of Entities & Attributes	Unknown
	Search Keys and tables	Unknown
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.2.6 Ten Largest Boulders

Data set Name	10 Largest Boulders	
Data set Area	Sediment Transport	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)		
Description of Project	<p>Sediment point counts do not accurately measure particle size distributions of boulder-sized sediments.</p> <p>During 2001, the diameters of the 10 largest boulders on each of 55 debris fans were measured in order to improve the estimates of distributions of large sized particles.</p>	
Description of Data sets	Data Format	Unknown
	Dates of Collection	2001
	Quantity of Data	Unknown
	Location of Data	USGS, Tucson
	Major Tasks	Acquire data from PI
	Time required to complete tasks	Unknown
	List of Entities & Attributes	Unknown

	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.2.7 Cs-137 Samples

Data set Name	Cs-137 Samples	
Data set Area	Sediment Transport	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)		
Description of Project	<p>The concentration of Cs-137 in tributary sediment deposits should correlate with the quantity of sediment provided by a given tributary watershed.</p> <p>Sediment samples of > 9g were collected at 57 tributaries during 2001, and analyzed for Cs-137 concentration.</p>	
Description of Data sets	Data Format	
	Quantity of Data	Unknown
	Location of Data	USGS, Tucson
	Major Tasks	Acquire data from PI
	Time required to complete tasks	Unknown
	List of Entities & Attributes	<p>Cs-137</p> <ul style="list-style-type: none"> • Station_id • Station_type • Trip_id • Date • Sample_number • X_COORD • Y_COORD • Sample_wt (g) • CS-137_concentration • Comments
	Search Keys and tables	
Current Collection Method	<p>Sample size >= 9g.</p> <p>Samples are taken in slack water sediment deposits located at tributary mouths.</p> <p>Samples are analyzed by contract lab.</p>	
Future Collection Method	Same	
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.2.8 C-14 Samples

Data set Name	C-14 Samples	
Data set Area	Sediment Transport	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)		
Description of Project	<p>The concentration of C-14 in debris flow sediments can be used to date stratigraphic events. Events dated in this manner could be added to the DF event table.</p> <p>15 radiocarbon samples were collected at 6 different debris flow sites during 2001.</p>	
Description of Data sets	Data Format	Unknown
	Dates of Collection	2001
	Quantity of Data	Unknown
	Location of Data	Unknown
	Major Tasks	Acquire data from PI
	Time required to complete tasks	Unknown
	List of Entities & Attributes	<p>C-14</p> <ul style="list-style-type: none"> • Station_id • Trip_id • Date • Sample_number • X_COORD • Y_COORD • Sample_depth • Sample_wt(g) • C-14_concentration • C-12_concentration • Sediment_age • Comments
	Search Keys and tables	
Current Collection Method	<p>Samples taken in debris fan soil horizons to date significant depositional events. Samples are analyzed by contract lab.</p>	
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.2.9 Boulder PIT Tags

Data set Name	Boulder PIT Tags	
Data set Area	Sediment Transport	
Data set Subject	Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)		
Description of Project	<p>The purpose of this experiment is to monitor the movement of boulders and cobbles in response to hydrographic events.</p> <p>During 2001, 48 PIT (Passive Integrated Transponder) tags were inserted into boulders and cobbles, situated below the 20,000 CFS stage, at four selected debris fan sites. The locations of these marked boulders and cobbles were then surveyed using total station instruments relative to control points established at these sites.</p> <p>The locations of these marked boulders will be re-surveyed yearly to determine movement in response to hydrographic changes in the Colorado River or tributaries.</p>	
Description of Data sets	Data Format	CSV
	Dates of collection	2001
	Quantity of Data	48 records, monitored yearly, starting in 2001. 100 additional tags in the future
	Location of Data	USGS Tucson
	Major Tasks	Unknown
	Time required to complete tasks	Unknown
	List of Entities & Attributes	<p>Boulder PIT Tags</p> <ul style="list-style-type: none"> • Station_id • Station_type • Trip_id • Date • Tagnum • X_Coordinate • Y_Coordinate • Particle_size • Description
	Search Keys and tables	
Current Collection Method	As described above	
Future Collection Method	100 additional boulders/cobbles to be tagged in the future.	
Data Analysis	Determine movement of boulders per year, or per stage event.	
Data Query	1) Select * from BOULDER_PIT_TAGS group by TAGNUM order by DATE	
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.3 Hydrology

4.1.3.1 Water Velocity Measurements

Data set Name	Water Velocity	
Data set Area	Sediment Transport, Hydrology	
Data set Subject	Fine Grained, Coarse Grained, Stream Flow, Modeling	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb	
Designer ERD/Entity Names(s)		
Description of Project	<p>Order of magnitude measurements of average water surface velocity have been made by timing the travel of both tether-balls and oar powered boats through rapids.</p> <p>Instantaneous measurements of subsurface velocity were made during 2001 using an Acoustic Doppler Velocimeter (ADV). This device allows for measuring 3-dimensional water velocities at any given point, and/or velocity depth profiles. Velocity measurements are mapped as a function of discharge.</p>	
Description of Data sets	Data Format	
	Dates of Collection	2001
	Quantity of Data	
	Location of Data	USGS Tucson
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	<p>Surface Velocity</p> <ul style="list-style-type: none"> • Station_id • Station_type • Trip_id • Date/Time • Discharge • Method • X_COORD • Y_COORD • Depth • Velocity • Comments
Search Keys and tables		
Current Collection Method	Average surface velocity through rapids, by timing tracer travel times.	
Future Collection Method	Instantaneous velocities measured using ADV instrumentation suspended from a boom assembly attached to a motorized boat.	
Data Analysis	1) 2-D velocity profiles of river at selected stations.	
Data Query	Public, Query average surface velocity, by rapid, by stage.	
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.3.2 Dam Release Data (SCADA)

Data set Name	SCADA
Data set Area	Hydrology
Data set Subject	Stream Flow

GCMRC Contact	Ted Melis	
Principal Investigator(s)	Bill Vernieu	
Designer ERD/Entity Names(s)	PHYS STREAM FLOW / PHYS SCADA	
Description of Project	The Bureau of Reclamation monitors actual dam releases every 15 minutes. Summary total discharge data reported hourly. Dam releases are also available for each draft tube, though total discharge is the most important information to report.	
Description of Data sets	Data Format	CSV ASCII text, SAS tables, Oracle table
	Dates of Collection	1963 - Present
	Quantity of Data	350,000 records
	Location of Data	USBOR. A copy resides in SAS at GCMRC (Vernieu)
	Major Tasks	1) Update Oracle table structure. 2) Update flow table. 3) Import data from SAS table.
	Time required to complete tasks	
	List of Entities & Attributes	SCADA_Hydrograph <ul style="list-style-type: none"> • ID • DATE_TIME • Q (discharge)
	Search Keys and tables	
Current Collection Method	Bill Vernieu of GCMRC receives downloads daily from USBOR. He then imports data into SAS tables, and spurious values are corrected.	
Future Collection Method	GCMRC needs to receive this data in an automated way, systematically updating its hydrograph table in Oracle.	
Data Analysis	Dam hydrograph data are used as input to the FlowGUI, to predict stage and discharge at points downstream on a given date.	
Data Query	Public: Web access to dam discharge as a function of date. GCMRC: FlowGUI query of hydrograph table for given date range.	
Priority		

GCMRC Archives? No. A copy of these data reside in SAS directories at GCMRC.
 Metadata? Some
 GCMRC Pilot? Yes

4.1.3.3 Dam Release Forecast

Data set Name	Dam Release Forecast	
Data set Area	Hydrology	
Data set Subject	Stream Flow	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Clayton Palmer (WAPA)	
Designer ERD/Entity Names(s)	PHYS STREAM FLOW / PHYS FORECAST	
Description of Project	The Western Area Power Authority (WAPA) predicts GCD releases six months in advance, based on predicted inputs to Lake Powell and current lake levels. WAPA also produces a 3 day forecast of releases that is fairly accurate.	
Description of Data sets	Data Format	CSV ASCII text
	Quantity of Data	6 month forecast = 4,320 records
	Location of Data	USBOR

	Major Tasks	1) Create Oracle table structure. 2) Import forecast data. 3) Decide on frequency to import forecast. 4) Automate the import process.
	Time required to complete tasks	
	List of Entities & Attributes	Hydrograph <ul style="list-style-type: none"> • ID • DATE_TIME • PREDICTED_Q • FORECAST_TYPE (6mo, 3day)
	Search Keys and tables	
Current Collection Method		
Future Collection Method	GCMRC needs to receive this data in an automated way, systematically updating its predicted hydrograph table on a daily basis. It is assumed that both six month and 3 day forecasts could be accommodated in the same table.	
Data Analysis	Forecasted hydrograph data could be used as input to a predictive model of stage and discharge at points downstream on a given date.	
Data Query	Public: Web access to forecast dam discharge as a function of date. GCMRC: FlowGUI query of forecast hydrograph table, for given range of dates.	
Priority		

4.1.3.4 Stream Flow Gage Data

Data set Name	Stream Flow Gage Data	
Data set Area	Hydrology, Sediment Transport	
Data set Subject	Stream Flow, Water Quality, Fine Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	David Topping, Nancy Hornewer	
Designer ERD/Entity Names(s)	PHYS STREAM FLOW / PHYS STREAM FLOW UNIT VALUE AND PHYS STREAM FLOW DAILY AND PHYS WATER TEMP COND TURB	
Description of Project	Both permanent and temporary gages are maintained by both the USGS and GCMRC along the Colorado River and its tributaries. These are used to measure a variety of hydrologic and water quality parameters.	
Description of	Data Format	TSV, CSV, MS Excel

Collection Method	<p>in the Grand Canyon. Current data collection from these stations involves a satellite transfer of data every 4 – 5 hours, to the USGS ADAPS (Automated Data Processing System) system. Stream flow data can be extracted from this system in the following ways:</p> <p>Grand Canyon Basic Data USGS Website</p> <p>Historical snapshots of USGS Streamflow gages specific to the Grand Canyon are available at http://az.water.usgs.gov/gcbd/index.html as follows:</p> <table border="1" data-bbox="362 394 1451 1031"> <thead> <tr> <th>Gage</th> <th>Gage Description</th> <th>Measured</th> <th>Year Range</th> </tr> </thead> <tbody> <tr> <td>09379910</td> <td>Colorado River below Glen Canyon Dam</td> <td>Gage Height (ft) and Discharge (m³/sec)</td> <td>1988 to present</td> </tr> <tr> <td>09380000</td> <td>Colorado River at Lee's Ferry, AZ</td> <td>Gage Height (ft) and Discharge (ft³/sec) Water Temp, Conductivity, Turbidity</td> <td>1921 to present 1988 to present</td> </tr> <tr> <td>09382000</td> <td>Paria River at Lee's Ferry</td> <td>Gage Height and Discharge</td> <td>1988 to present</td> </tr> <tr> <td>09383100</td> <td>Colorado River above Little Colorado River, near Desert View</td> <td>Gage Height and Discharge, Water Temp, Conductivity, Turbidity</td> <td>1988 to present</td> </tr> <tr> <td>09402000</td> <td>Little Colorado River near Cameron</td> <td>Gage Height and Discharge</td> <td>1988 to present</td> </tr> <tr> <td>09402500</td> <td>Colorado River near Grand Canyon</td> <td>Gage Height and Discharge, Water Temp, Conductivity, Turbidity</td> <td>1988 to present</td> </tr> <tr> <td>09404200</td> <td>Colorado River above Diamond Creek, near Peach Springs</td> <td>Gage Height and Discharge</td> <td>1988 to present</td> </tr> </tbody> </table> <p>Notes: An interval over which data are collected is variable. Older data was collected every six hours, newer data every 15 minutes. Time intervals are inconsistent and sometimes missing. Units are inconsistent. Older stages and discharges are reported in ft and ft³/sec; newer data are in m and m³/sec.</p> <p>These data are available as downloadable data sets, and as such are not available to query.</p> <p>National Water Information System (NWISWeb)</p> <p>The NWIS Website (http://water.usgs.gov/az/nwis/current/?type=flow) gives access to Stream Flow Unit Values for the above listed gages through a sophisticated web query interface. Recent version of the NWIS website have only made Stream Flow Unit values available for the last 31 days. Beyond 31 days, Daily Means are available. Data are available to view, or to download, either in graphic or tabular form.</p> <p>Direct ADAPS Access</p> <p>The underlying database system for the time-series stream flow data served by NWISWeb is the ADAPS system. Direct access to data in the ADAPS system can be accomplished either through an ADAPS client application, or through a Unix script called "nwts2rdb". This script provides a number of advantages:</p> <p>Nwts2rdb could be run on a scheduled basis to extract a specified data set from the ADAPS system. This data set could then be imported into the GCMRC Oracle system</p> <p>Extracts done using the nwts2rdb include additional Quality Assurance and Remark flags that are not available thru the web interface. The QA/QC flags provided some information as to changes that may have been made to the data over time, as well as the expected level of accuracy of the data.</p> <p>GCMRC Temporary Stream Gages</p> <p>GCMRC maintains seven "temporary" stream gages in formerly un-gaged tributaries of the Colorado River. The distribution of these sites is a subset of the Rain gage sites (See Physical / Climate/Rainfall for site details – gaged sites are suffixed with a "G"). Data collected at these sites has the same stream flow unit value attributes as the main USGS gages (with the exception of ADAPS Q/A flags). It differs in that the locations of sites may change, data are not recorded in real-time, and processing of raw data does not follow that of data from the USGS gages. This data will be</p>	Gage	Gage Description	Measured	Year Range	09379910	Colorado River below Glen Canyon Dam	Gage Height (ft) and Discharge (m ³ /sec)	1988 to present	09380000	Colorado River at Lee's Ferry, AZ	Gage Height (ft) and Discharge (ft ³ /sec) Water Temp, Conductivity, Turbidity	1921 to present 1988 to present	09382000	Paria River at Lee's Ferry	Gage Height and Discharge	1988 to present	09383100	Colorado River above Little Colorado River, near Desert View	Gage Height and Discharge, Water Temp, Conductivity, Turbidity	1988 to present	09402000	Little Colorado River near Cameron	Gage Height and Discharge	1988 to present	09402500	Colorado River near Grand Canyon	Gage Height and Discharge, Water Temp, Conductivity, Turbidity	1988 to present	09404200	Colorado River above Diamond Creek, near Peach Springs	Gage Height and Discharge	1988 to present
Gage	Gage Description	Measured	Year Range																														
09379910	Colorado River below Glen Canyon Dam	Gage Height (ft) and Discharge (m ³ /sec)	1988 to present																														
09380000	Colorado River at Lee's Ferry, AZ	Gage Height (ft) and Discharge (ft ³ /sec) Water Temp, Conductivity, Turbidity	1921 to present 1988 to present																														
09382000	Paria River at Lee's Ferry	Gage Height and Discharge	1988 to present																														
09383100	Colorado River above Little Colorado River, near Desert View	Gage Height and Discharge, Water Temp, Conductivity, Turbidity	1988 to present																														
09402000	Little Colorado River near Cameron	Gage Height and Discharge	1988 to present																														
09402500	Colorado River near Grand Canyon	Gage Height and Discharge, Water Temp, Conductivity, Turbidity	1988 to present																														
09404200	Colorado River above Diamond Creek, near Peach Springs	Gage Height and Discharge	1988 to present																														

	<p>owned by GCMRC, not the USGS, and there will be no distinction between Provisional and Published information as with USGS data.</p> <p>USGS Temporary Stream Gages</p> <p>The U.S. Geological Survey collected river-stage data at 53 temporary gauging stations along the Colorado River downstream from Glen Canyon Dam to upper Lake Mead during 1990 to 1994. These data were collected to support research scientists in their investigations as part of the Bureau of Reclamation's Glen Canyon Environmental Studies. The data-collection period included about 14 months of dam releases designed for research purposes from June 1990 to July 1991. This data are available in electronic form from Nancy Hornewer of USGS. Data files are Column Aligned Text, CAT format, and contain only River Stage values, not flow rates. Refer to the readme.txt on the CD or in RiverStage90-94.zip for Station Details, as well as column format and description,</p> <p>Reconstructed Values</p> <p>Stream Flow Unit values have been reconstructed for Colorado River @ Lee's Ferry, and the Colorado River at Grand Canyon, for the period 1921 – 1986. These data resulted from a study by D.J. Topping, J.C. Schmidt, and L.E. Vierra, Jr. Please refer to their USGS Professional Paper for information on how this instantaneous stage and discharge time series was computed, and for flow-duration, flood-frequency and discharge-variability analyses conducted on this time series. These reconstructed values currently exist as CSV (Comma-Separated-Value) format files at ftp://ftp.gcmrc.gov/data/physical/streamflow_unit_values/.</p>
Future Collection Method	<p>A MS Access application is needed to import Stream Flow Unit Values and Stream Flow Daily Mean values into the underlying DMF Stream flow tables. This import application must be able to deal with input data streams of the types listed above.</p> <p>The USGS NWISWeb website goes to great lengths to point out the Provisional nature (http://water.usgs.gov/az/nwis/help/?provisional) of its real-time data. Any process employed for using these data must accommodate updating real-time data, should any changes to provisional status take place. Daily Mean values are finalized and published by the USGS in a series of water-data reports. The requirement to update historical values is not as rigorous.</p>
Data Analysis	<p>Stream Flow is a long-term monitoring data set. As such, it functions as a resource to research and modeling efforts in a number of other areas and subjects. Data Analyses will be specific to these areas, and will most likely be done in conjunction with other data sets.</p> <p>A common query need is a representation of discharge versus time. This frequency distribution, the amount of time that discharge exceeds a certain value, can be used to represent the flow characteristics for a given time period such as a year.</p> <p>The time interval between sample values recorded in gage data sets is quite variable. In the days before gages became automated, instruments might have only been read once or twice a day. Since gages became automated, their sample interval is typically in the 5 – 60 minute range. Sampling interval can change with time, and it is common to see intervals in which samples have been missed. These conditions should be considered by anyone using gage data for analysis.</p>
Data Query	<p>Public, query access to the Admin Stream Flow tables.</p> <p>The units of measure commonly used by the USGS and GCMRC are cubic feet per second (cfs) for discharge, and feet for gage height. This does not conform to GCMRC Data Standards, which specify the use of SI (Metric) units. A mechanism must be provided to present Stream Flow values in both English and SI units.</p>
Priority	

GCMRC Archives? Yes
 Metadata? Some
 GCMRC Pilot? Yes

4.1.3.5 Stream Flow Gage Logs

Data set Name	Gage Logs	
Data set Area	Hydrology	
Data set Subject	Stream Flow	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	David Topping	
Designer ERD/Entity Names(s)	PHYS STREAM FLOW / PHYS GAGE LOG	
Description of Project	The Gage log is based on a hardcopy log of gage activity, and was maintained up until the 1960's during the period when the gage stations were manned. The log contains gage related information, such as maintenance logs and notes of sandbar formation, as well as anecdotal information from the sampling period that may not be related to the gage operation itself.	
Description of Data sets	Data Format	Hard Copy
	Quantity of Data	
	Location of Data	National Archives
	Major Tasks	1) Define Oracle table structure. 2) Manually enter comments from gage logs.
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Historical gage logs up till the 60's will be transcribed into electronic form. Most likely this will require a manual data entry task.	
Future Collection Method	No longer collected	
Data Analysis	Because this information is qualitative in nature, it is unlikely that it would support any kind of analysis. Its value to someone looking at historical stream flow information might be to search for conditions that affected stream flow or gauging accuracy. Anecdotal information might also provide clues to general weather trends, special circumstances on the river, or information about specific river trips.	
Data Query	Public: Web access to comments in the gage log table.	
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.3.6 Gage Photographs

Data set Name	Gage Photographs	
Data set Area	Hydrology	
Data set Subject	Stream Flow	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Nancy Hornewer, David Topping	
Designer ERD/Entity Names(s)	PHYS STREAM FLOW / PHYS GAGE PHOTOS	
Description of Project	Photographs have been taken of the gauging stations over the years. These photos are related to anecdotal information contained in the gage logs. For initial modeling purpose, Gage Photos have been identified with their own entity. However, treatment of the gauging station photographs will done by the	

	generalized image storage and access capabilities provided by the DMF. These capabilities are detailed in the section Photography.	
Description of Data sets	Data Format	Hardcopy, TIFF
	Quantity of Data	
	Location of Data	National Archives
	Major Tasks	1) Scan hardcopy photographs. 2) Define Oracle table structure. 3) Import photographs to Oracle.
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query	Public: Web access to the gage log photographs	
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.3.7 Multi-beam Hydrography

Data set Name	Multi-beam Hydrography	
Data set Area	Sediment Transport, Hydrology	
Data set Subject	Fine Grained, Coarse Grained, Stream Flow, Channel Geometry, Modelling	
GCMRC Contact	Ted Melis, Mark Gonzales	
Principal Investigator	Ted Melis	
Designer ERD / Entity Names(s)		
Description of Project	Multi-beam sonar techniques are used to map channel geometry to a high level of detail. Channel mapping is used to determine accurate stage-discharge relationships, to estimate volumes of fine grained sediments stored in the channel bottom and margins, to estimate volumes of debris flow sediments, to analyze river bottom substrate, etc.	
Description of Data sets	Data Format	CSV point files, TIFF images, ARC/INFO TIN
	Quantity of Data	
	Location of Data	GCMRC survey department
	Major Tasks	
	Time required to complete tasks	

	List of Entities & Attributes	<p>Hydro Surveys (Index to source files)</p> <ul style="list-style-type: none"> • Station_id • Station_type • Date • Trip_id • Start RM • End RM • Link to thinned point file • TIFF image of setup <p>Hydro Points (Thinned files)</p> <ul style="list-style-type: none"> • Station_id • Station_type • Date • X_COORD (State Plane) • Y_COORD (State Plane) • Elevation
	Search Keys and tables	<ul style="list-style-type: none"> • Station_id, Station_type • Date • Image • Point file
Current Collection Method	<ul style="list-style-type: none"> • Multi-beam sonar efforts began in calendar year 2000. • Approximately 50 miles of river are mapped per year. Each set-up covers approximately ½ river mile, and results in a file containing 10-15 million sample points, stored in Hy-pack proprietary format. • Set-up files are edited and thinned using Hy-pack software. • The final product is a thinned point file, in ASCII CSV, for each set-up. These thinned files contain approximately 500,000 to 1.5 million points at 1 point /0.5 m². • Thinned point files are made available to PI's for use in surface modeling software such as ARC/INFO TIN. They are also used to produce TIFF images of the river bottom, which can be made available to the public. 	
Future Collection Method	Same	
Data Analysis	<ul style="list-style-type: none"> • Surface and volume modeling using thinned point files as input (PI). • Substrate analysis using raw or thinned point files (PI). 	
Data Query	<p>Public:</p> <ul style="list-style-type: none"> • Download thinned CSV point file of selected set-ups. • View TIFF images of selected set-ups. <p>PI:</p> <ul style="list-style-type: none"> • Same as public. • May also want selected raw point files for analysis (by special request to Survey Department). 	
Priority		

GCMRC Archive? No
 Metadata? No
 GCMRC Pilot? No

4.1.3.8 Side-scan Sonar

Data set Name	Side-scan Sonar	
Data set Area	Hydrology, Sediment Transport	
Data set Subject	Channel Geometry, Substrate , Fine Grained, Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Anima	
Designer ERD / Entity Names(s)		
Description of Project	Side scan sonar data are used to produce a substrate map of the Colorado River channel.	
Description of Data sets	Data Format	TIFF images, ARC/INFO polygon coverage
	Quantity of Data	
	Location of Data	USGS Menlo Park
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	Substrate <ul style="list-style-type: none"> • Polygon_id • Substrate_code • Interpretation Method
	Search Keys and tables	
Current Collection Method	River bottom is imaged, to within 0.5 m resolution, along a 100m wide swath	
Future Collection Method		
Data Analysis	Changes in substrate at a given location by year.	
Data Query	Polygon coverage base map of substrate.	
Priority		

GCMRC Archive? No
 Metadata? No
 GCMRC Pilot? No

4.1.3.9 Eyeball

Data set Name	Eyeball	
Data set Area	Hydrology, Sediment Transport	
Data set Subject	Channel Geometry, Substrate, Fine Grained, Coarse Grained	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	David Rubin	
Designer ERD / Entity Names(s)		
Description of Project	Underwater photography used to map sediment grain sizes on the Colorado River bottom.	
Description of Data sets	Data Format	TIFF images
	Quantity of Data	
	Location of Data	USGS Menlo Park
	Major Tasks	
	Time required to complete tasks	

	List of Entities & Attributes	Substrate <ul style="list-style-type: none"> • Polygon_id • Substrate_code • Interpretation Method
	Search Keys and tables	
Current Collection Method	Conventional photographs of the river bottom using a camera lowered into the water.	
Future Collection Method		
Data Analysis	Changes in substrate at a given location by year.	
Data Query	Polygon coverage base map of substrate.	
Priority		

GCMRC Archive? No
 Metadata? No
 GCMRC Pilot? No

4.1.3.10 Cross Sections

Data set Name	Cross Sections	
Data set Area	Hydrology, Modeling	
Data set Subject	Stream Flow, Channel Geometry, Flow Modeling	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Steve Weile	
Designer ERD/Entity Names(s)		
Description of Project	Surveyed cross sections of the river channel serve as input to stage discharge models, and the Flow GUI model developed by Ecometric. These include the Wilson, Stars, GCMRC and NAU cross-sections.	
Description of Data sets	Data Format	CSV, MS Access
	Quantity of Data	Wilson: 213 surveyed cross-sections, each with approximately 150 points per section. Stars Interpolated: 32,000 spot elevations in 750 cross sections.
	Location of Data	USGS, GCMRC
	Major Tasks	1) Define Oracle tables 2) Convert units to SI 3) Import data.
	Time required to complete tasks	

	List of Entities & Attributes	Cross Section Details <ul style="list-style-type: none"> • Site_id • Site_type (stars or nau) • Old_id (FlowGUI identifier) • Description • Origin X_COORD • Origin Y_COORD • End X_COORD • End Y_COORD • Stars_Interpolated (Y/N) • Date surveyed • Who surveyed Cross Section Elevation Points <ul style="list-style-type: none"> • ID • Site_id • Site_type • Trip_id • Distance from origin • Elevation (MSL)
	Search Keys and tables	
Current Collection Method	Along each profile, ground surface elevations are measured using conventional survey techniques to water's edge. River bottom elevations are measured using either single or multi-beam sonar. The Stars Interpolated cross-sections were not surveyed directly, but were interpolated between measured sections. These profiles have the same form as Wilson cross-sections.	
Future Collection Method		
Data Analysis	GCMRC, PI: Input to Flow Model	
Data Query	Public: Select profile for given site	
Priority		

GCMRC Archive? No
 Metadata? No
 GCMRC Pilot? No

4.1.3.11 Stage - Discharge Relationships

Data set Name	Stage-Discharge	
Data set Area	Hydrology, Modeling	
Data set Subject	Stream Flow, Channel Geometry, Flow Modeling	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Steve Weile, Matt Kaplinski, Nancy Hornewer	
Designer ERD/Entity Names(s)		
Description of Project	<p>All values of stage and discharge, whether measured or predicted, can be combined into a single data set by station.</p> <p>Empirical stage-discharge relationships have been determined at a discrete number of stations. Many of these were taken in conjunction with NAU sandbar surveys.</p> <p>Stars model output of stage and discharge relationships have been produced at Wilson cross-section locations, as well as Stars Interpolated cross-section locations.</p>	
Description of Data sets	Data Format	CSV, MS Access

	Dates of Collection	NAU: 1990 to present, at 31 sites. USGS gages: See collection dates in Streamflow.
	Quantity of Data	10,000 records of S-Q data.
	Location of Data	USGS Stars: a version resides within the current FlowGUI at GCMRC. NAU: Remains to be delivered to GCMRC. USGS gages: In GCMRC Archive. USGS temporary sites: In GCMRC Archive.
	Major Tasks	1) Define Oracle table 2) Convert units to SI 3) Import data
	Time required to complete tasks	
	List of Entities & Attributes	SQ_Data <ul style="list-style-type: none"> • ID • Site_id • Site_type • Water surface elevation (stage) • Discharge • Cross-sectional area • Velocity • Top width
	Search Keys and tables	
	Current Collection Method	
	Future Collection Method	
	Data Analysis	
	Data Query	
	Priority	

USGS: GCMRC Archives? Yes
 Metadata? Some
 GCMRC Pilot? Yes

NAU: GCMRC Archives? Some
 Metadata? Minimal
 GCMRC Pilot? No

4.1.3.12 Water Quality

4.1.3.12.1 Light Attenuation

Data set Name	Light Attenuation
Data set Area	Water Quality
Data set Subject	Physical
GCMRC Contact	Mike Yard

Principal Investigator(s)	Mike Yard	
Designer ERD / Entity Names(s)	PHYS LIGHT ATTENUATION / PHYS LA SEDIMENT, PHYS LIGHT ATTENUATION	
Description of Project	<p>Vertical aquatic light attenuation profiles have been measured at 24 sites between GCD and Diamond Creek since 1991. Simultaneously, sediment loads were measured at each site in order to determine the relationship between extinction coefficients and sediment load.</p> <p>Ultimately these data will be inputs for a predictive model of primary production potential.</p>	
Description of Data sets	Data Format	Lotus123
	Quantity of Data	120,000 records in 100 different spreadsheets.
	Location of Data	GCMRC
	Major Tasks	<ol style="list-style-type: none"> 1) Perform data cleanup in Lotus. 2) Convert 100 spreadsheets to csv. 3) Define Oracle table structure. 4) Import csv files into table structure.
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis	<ol style="list-style-type: none"> 1) What is the relationship between turbidity derived from light attenuation measurements (extinction coefficients), and direct measurements of turbidity using Hydrolab equipment? 2) What is the relationship between light attenuation parameters and sediment load by river mile, month, year, and discharge pattern? 3) What is the relationship between primary production, sediment load and light attenuation? 4) 	
Data Query	Same as data analysis?	
Priority		

GCMRC Archives? No (currently stored in M. Yard directories)

Metadata? No

GCMRC Pilot? No

4.1.3.12.2 Chemical / Physical / Biological

Data set Name	Water Quality, LP
Data set Area	Hydrology
Data set Subject	Water Quality
GCMRC Contact	Susan Hueftle, William Vernieu
Principal Investigator(s)	Same
Designer ERD/Entity Names(s)	WATER QUALITY / WQ
Description of Project	<p>Water quality information has been collected since the mid-1970s from both Lake Powell and downstream on the Colorado River. Lake water profile samples consist of water temperature, turbidity, conductivity and pH measured by Hydrolab instrumentation. Water samples are collected at depth and analyzed by a contract laboratory for trace elements, chlorophyll, phyto and zooplankton concentrations.</p>

Description of Data sets	Data Format	CSV ASCII text, MS Excel, MS Access, SAS.
	Quantity of Data	
	Location of Data	
	Major Tasks	<ol style="list-style-type: none"> 1) Review structure of surface data in MS Access database. 2) Review structure of MS Access database for water temperature, major ion, chlorophyll, phyto and zooplankton concentrations. 3) Import those data from MS Excel and CSV to MS Access database 4) Convert MS Access database to Oracle. 5) Write data applications to import CSV files from lab deliverables directly to Oracle.
	Time required to complete tasks	
	List of Entities & Attributes	See ER Diagrams
	Search Keys and tables	
Current Collection Method	<ol style="list-style-type: none"> 1) Monthly water sample profiles are taken at a single forebay station in Lake Powell. 2) Monthly tail water samples taken in GCD draft tubes, below GCD, and at Lees Ferry. 3) Continuous monitoring in draft tubes, below GCD, and at LF. 4) Quarterly water sample profiles taken at a number of repeat stations in Lake Powell, including inflow sites. 5) Periodic downstream water quality monitoring is done by GCMRC using Hydrolab equipment. 6) Downstream tributary water temperature monitoring is done using continuous temperature monitors. 7) USGS QWData – Water Quality Samples 	
Future Collection Method	A field data capture application could be employed to record surface information during sampling trips.	
Data Analysis		
Data Query		
Priority	High	

GCMRC Archive? No. LP data Currently stored in SAS Data directory. Downstream data stored in a variety of personal shares at GCMRC.

Metadata? Some

GCMRC Pilot? LP Water Quality – Yes. Downstream Water Quality – No.

4.1.4 Climate

4.1.4.1 Rainfall

Data set Name	Rainfall
Data set Area	Climate
Data set Subject	Rainfall
GCMRC Contact	Ted Melis
Principal Investigator(s)	David Topping
Designer ERD/Entity Names(s)	PHYSICAL / RAIN GAGE

<p>Description of Project</p>	<p>GCMRC maintains a number of rain gages throughout the Grand Canyon Ecosystem. These rain gages are identified as follows:</p>	
	<p>Gage Circuit</p>	<p>Gage Name</p>
	<p>Shinumo Wash</p>	<p>SHIN-1 SHIN-2 SHIN-G CEDAR</p>
	<p>Highway 89</p>	<p>TAN-1 LIME ECHO WATER-1 WATER-2 WATER-G</p>
	<p>Glen Canyon boat</p>	<p>TAN-G</p>
	<p>Highway 89A / Rider Canyon</p>	<p>BADGR-G EMMETT HOUSE-G RIDER-G NORTH-G</p>
	<p>Buffalo Ranch Road / East Side Game Trail</p>	<p>HONEYMN HOUSE-2 NORTH-2 TATER-2 KANE-2 ROCK PASTURE</p>
	<p>Paria Plateau</p>	<p>HOUSE-1 ONETOE BADGR-1</p>
<p>North Rim</p>	<p>NORTH-1 TATER-1 KANE-1 BURRO</p>	
<p>Note: Detailed in ftp://ftp.gcmrc.gov/data/physical/Rain gage data/Road_log.rtf. Gage names that end if a "G" indicate there is stream flow gage in addition to Rain gage.</p>		
<p>Description of Data sets</p>	<p>Data Format</p>	<p>MS Excel</p>

	Dates of Collection	BADGR-1 6/19/2001 – 7/18/2002 BADGR-G 1/8/2001 – 5/30/2002 BURRO6/18/2001 – 5/30/2002 CEDAR3/23/2001 – 6/1/2002 ECHO 1/20/2001 – 5/2/2001 EMMETT 7/26/2000 – 5/29/2002 HONEYMN 6/19/2001 – 8/16/2001 HOUSE-1 7/26/2000 – 7/18/2002 HOUSE-2 1/14/2001 - 5/29/2002 HOUSE-G 1/19/2001 - 5/29/2002 KANE-1 6/18/2001 – 5/30/2002 KANE-2 7/3/2001 – 5/30/2002 LIME 1/20/2001 – 5/29/2002 NORTH-1 6/18/2001 – 5/30/2002 NORTH-2 1/14/2001 – 5/29/2002 NORTH-G 1/6/2001 – 5/29/2002 ONETOE 6/19/2001 – 7/18/2002 PASTURE 7/26/2000 – 5/30/2002 RIDER-G7/26/2000 – 5/29/2002 ROCK 7/3/2001 – 5/30/2002 SHIN-1 1/20/2001 – 5/28/2002 SHIN-2 1/20/2001 – 5/29/2002 SHIN-G 1/12/2001 – 5/28/2002 TAN-1 3/23/2001 – 7/2/2002 TAN-G 1/10/2001 – 5/31/2002 TATER-1 6/18/2001 – 9/13/2001 TATER-2 6/20/2001 – 5/29/2002 WATER-1 1/12/2001 – 5/29/2002 WATER-2 1/15/2001 – 5/29/2002 WATER-G 1/9/2001 – 3/7/2002
	Quantity of Data	30 files
	Location of Data	GCMRC
	Major Tasks	
	Time required to complete tasks	

	List of Entities & Attributes	<p>Rainfall</p> <ul style="list-style-type: none"> • Station_id • Station_type • Date_Time • Inches of Rain • Comments <p>Rain Gage Station Details</p> <ul style="list-style-type: none"> • Station_id • Station_type • Gage_name • X_COORD • Y_COORD • Elevation • Description <p>Station Deploiments</p> <ul style="list-style-type: none"> • ID • Station_id • Station_type • Date/Time • Service Action
	Search Keys and tables	
Current Collection Method	A data logger is used to collect rain gage data at each site. Data are downloaded from each logger to a PC on a monthly basis to CSV files, and subsequently imported to MS Excel. Examples of these Excel data files can be found at ftp://ftp.gcmrc.gov/data/physical/Rain_gage_data/ .	
Future Collection Method	An application will be required to routinely import the Rain gage log files into the appropriate table in the Oracle database, skipping the MS Excel step.	
Data Analysis	<p>Rainfall is a long-term monitoring data set. As such, it functions as a resource to research and modeling efforts in a number of other areas and subjects. Analyses are specific to these areas, and will most likely be done in conjunction with other data sets. Typically, one might want to display the average monthly, or yearly rainfall for a given watershed.</p> <p>The time interval between samples in the rain gage data sets is highly variable. The logger basically creates a new record when it has seen an increase in Inches of Rain. If no rain has fallen, no record will exist for that period. As well, at some point in time, a rollover of Inches of Rain will occur. This rollover value will need to be determined.</p> <p>The above conditions should be considered by anyone using the Rainfall data for analysis.</p>	
Data Query	Query average rainfall by day, month, year, or decade, for a given watershed.	
Priority		

GCMRC Archives? Yes
 Metadata? Some
 GCMRC Pilot? Yes

4.1.4.2 NWS Weather Stations

Data set Name	Weather Stations
Data set Area	Climate
Data set Subject	Weather Service

GCMRC Contact	Ted Melis	
Principal Investigator(s)	Ted Melis	
Designer ERD/Entity Names(s)		
Description of Project	10 stations from the National Weather Service provide pertinent weather data, containing air temperature, rainfall, and wind speed and direction, dew point and atmospheric pressure. Climatic data are also available in the Colorado River Corridor at Phantom Ranch, and on the North and South Rims at Ash fork, Cameron, Desert View RS, Flagstaff, Grand Canyon National Park, Lee's Ferry, Mt. Trumball, Page, Seligman, Tuweep, and Williams. Other resources (USGS Water Resources Investigations Report WRIR00-4055) indicate weather stations at Bright Angel RS, Peach Springs, Supai, and Tuba City as well. This data is available from NCDC (National Climatological Data Center) website for a fee, or a compact disk of all Western weather station data may be purchased from EARTHINFO on some regular basis.	
Description of Data sets	Data Format	ASCII text
	Dates of Collection	Depending on station, 1890 to present
	Quantity of Data	
	Location of Data	National Weather Service (http://www.ncdc.noaa.gov) Some data is stored at GCMRC in SAS directory.
	Major Tasks	1) Purchase data 2) Decide on pertinent stations 3) Define Oracle table structure 4) Import data to Oracle 5) Decide on update frequency
	Time required to complete tasks	1 week, depending on the number of stations, and the history required
	List of Entities & Attributes	<ul style="list-style-type: none"> • Station name • Date • Daily Maximum temperature (.1 Fahrenheit) • Daily Minimum temperature (.1 Fahrenheit) • Daily Precipitation (.01 inch)
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No (In SAS directory)
 Metadata? Yes
 GCMRC Pilot? No

4.1.4.3 Weather Radar

Data set Name	Weather Radar
Data set Area	Climate

Data set Subject	Weather Service	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Ted Melis, David Topping	
Designer ERD/Entity Names(s)		
Description of Project	A subscription to weather radar data is available for 3 stations on the rim. This weather radar information is used to forecast and model sediment conditions in the tributaries of the Grand Canyon. This is a time-series image data set.	
Description of Data sets	Data Format	TIFF
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archive? No
 Metadata? No
 GCMRC Pilot? No

4.1.4.4 Air Temperature

Data set Name	Air Temperature	
Data set Area	Climate	
Data set Subject	Air Temperature	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Ted Melis, David Topping, others	
Designer ERD/Entity Names(s)		
Description of Project	Air temperatures are collected during most sampling activities as a part of observations of climatic conditions while sampling.	
Description of Data sets	Data Format	CSV, dBase, MS Access
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	Air Temperature <ul style="list-style-type: none"> • ID • Station_id • Station_type • Trip_id • Date/Time • Air Temperature (Centigrade)
	Search Keys and tables	
Current Collection Method	Air temperatures are collected at certain GCMRC gage locations, USGS gages,	

	and from other weather stations around the Grand Canyon. Generally these are from time series data collection, though daily max and min temperatures are recorded at various sites. Air temperatures are also collected in conjunction with other scientific efforts, such as fish sampling. These usually take the form of discrete measurements delivered within another data set, often in conjunction with descriptive comments about the weather or sampling conditions.
Future Collection Method	
Data Analysis	
Data Query	
Priority	

GCMRC Archive? No
 Metadata? No
 GCMRC Pilot? No

4.1.5 Terrestrial Biology

4.1.5.1 Tribal Ethno Botany

Data set Name	Tribal Ethno Botany	
Data set Area	Terrestrial Biology	
Data set Subject	Tribal Ethno Botany	
GCMRC Contact	Barbara Ralston, Ruth Lambert	
Principal Investigator(s)	Michael Kearsley, Mike Yeatts, Brenda Drye, Kerry Christiansen, Lorretta Jackson	
Designer ERD/Entity Names(s)		
Description of Project	<p>Tribal ethno botanical monitoring closely relates to other vegetation monitoring in terms of data set structure. The main differences are:</p> <ul style="list-style-type: none"> • A Tribal member is present while data collection work is done. • Species of tribal significance are recorded. • Certain qualitative information is recorded at the same time. <p>The Tribes have been involved in ethno botanical studies (i.e., research, assessments, field interviews with elders, and monitoring) to different degrees. The monitoring design included plant study transects, and provides information on the effects of high releases and low flows on ethno botanical cultural resources at sites that have significance as Traditional Cultural Properties (TCPs) to the cooperating Tribes.</p> <p>These locations include traditional resources such as plants, springs, rock art, and archaeological remains.</p>	
Description of Data sets	Data Format	
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Varies by tribe	
Future Collection Method		
Data Analysis		

Data Query	
Priority	

GCMRC Archive? No
 Metadata? No
 GCMRC Pilot? No

4.1.5.2 Avifauna

Typically four trips per year are carried out in order to perform riparian bird surveys, using either walking transect or point count methodologies. As well, opportunistic observation of waterfowl may be performed on any trip.

Historically, different avifauna data sets represent slightly different sampling methodologies, and these must be generalized and standardized into a single data model that supports the sampling protocols currently in use by the Terrestrial Biology program at GCMRC.

A Vegetation Survey is usually performed in conjunction with Bird Surveys. Refer to the Vegetation subject for a discussion of this data set. Bird surveys are generally conducted within the long term "Bird Patch" vegetation plots located along the River corridor. These "Bird Patches" are the integrated sites of bird, vegetation, vertebrate and invertebrate transect counts, and are now conducted during three sampling trips per year.

4.1.5.2.1 Bird Point Counts

Data set Name	Bird Point Count	
Data set Area	Terrestrial Biology	
Data set Subject	Avifauna	
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Helen Yard	
Designer ERD/Entity Names(s)	BIO BIRDS/ BIRD WALKING TRANSECT, BIRD POINT COUNT, BIRD WALK PT HY, BIRD HERP COUNT BIRD WATERFOWL	
Description of Project		
Description of Data sets	Data Format	MS Excel
	Dates of Observation	May 2001 – June 2003
	Quantity of Data	
	Location of Data	H.Yard Consulting/GCMRC
	Major Tasks	
	Time required to complete tasks	

	List of Entities & Attributes	Point Counts <ul style="list-style-type: none"> • Date • Site • Side • Pt# • Species • Num • Dist • Type • Zone • Substrate • Sex • Age • Activity • Breed • Notes
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archive? Yes
 Metadata? No
 GCMRC Pilot? No

4.1.5.2.2 Bird Walking Transects

Data set Name	Bird Walking Transect	
Data set Area	Terrestrial Biology	
Data set Subject	Avifauna	
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Helen Yard	
Designer ERD/Entity Names(s)	BIO BIRDS/ BIRD WALKING TRANSECT, BIRD POINT COUNT, BIRD WALK PT HY, BIRD HERP COUNT BIRD WATERFOWL	
Description of Project		
Description of Data sets	Data Format	MS Excel
	Dates of Collection	May 2001 – June 2003
	Quantity of Data	64 sites/yr, 3 times/yr
	Location of Data	Hyard Consulting/GCMRC
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	Walking Surveys Note: Walking Surveys have same attributes as point counts, but also have start time and end time recorded.
	Search Keys and tables	

Current Collection Method	Walking surveys and Point counts. Conducted in OHWZ and NHWZ. Sites are randomized within each Bird Patches.
Future Collection Method	Same
Data Analysis	
Data Query	
Priority	

GCMRC Archive? Yes
 Metadata? No
 GCMRC Pilot? No

4.1.5.2.3 Waterfowl Couts

Data set Name	Bird Waterfowl Count	
Data set Area	Terrestrial Biology	
Data set Subject	Avifauna	
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Helen Yard, Larry Stevens	
Designer ERD/Entity Names(s)	BIO BIRDS/ BIRD WALKING TRANSECT, BIRD POINT COUNT, BIRD WALK PT HY, BIRD HERP COUNT BIRD WATERFOWL	
Description of Project		
Description of Data sets	Data Format	MS Excel
	Dates of Collection	May 2001 – June 2003
	Quantity of Data	
	Location of Data	H.Yard Consulting, GCMRC archive
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	Observations <ul style="list-style-type: none"> • Site • Side • Zone • Species • Number • Date • Reach • Species • Number • Class
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archive? Yes
 Metadata? No
 GCMRC Pilot? No

4.1.5.2.4 Bird Population Density

Data set Name	Bird Population Density	
Data set Area	Terrestrial Biology	
Data set Subject	Avifauna	
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Helen Yard	
Designer ERD/Entity Names(s)	BIO BIRDS/ BIRD WALKING TRANSECT, BIRD POINT COUNT, BIRD WALK PT HY, BIRD HERP COUNT, BIRD WATERFOWL	
Description of Project		
Description of Data sets	Data Format	MS Excel
	Dates of Collection	
	Quantity of Data	
	Location of Data	NAU, SWCA
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Data are now recorded on paper field sheets. Manual data entry to MS Excel is done in the office.	
Future Collection Method	<p>A Field Data Recorder (FDR), comprised of a specialized Data Capture application running on field-capable computer hardware is an option to replace the paper based survey recording. Because the data input requirements are fairly low, the benefits of reducing data entry will need to be weighed against the expense and exposure to breakdown that a FDR would introduce.</p> <p>In the absence of a FDC, an office-based Data Entry Application is required. Because of the relatively simple nature of the data set at present, accepting data sets as standardize format CSV or Excel files is sufficient. A more robust solution would be to provide field crews with an MS Access based application for Data Entry.</p> <p>An application is required for the Data Administrator to import either the FDR-generated data set, or the office-based Data Capture Applications data set.</p>	
Data Analysis	<ol style="list-style-type: none"> 1) Species Inventory 2) Population Estimates - Presence/Not Detected, Relative Abundance, and Absolute Abundance population estimates are typically calculation done using specialized applications for that purpose. 3) Spatial Distribution 	
Data Query	Public, species presence/absence, spatial distribution of species.	
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.5.3 Vegetation

4.1.5.3.1 Kanab Amber Snail Vegetation

Data set Name	Kanab Amber Snail Vegetation Layer
Data set Area	Terrestrial Biology
Data set Subject(s)	Invertebrates, Vegetation

GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Jeff Sorensen, Clay Nelson	
Designer ERD / Entity Names(s)		
Description of Project	Population densities of KAS, , along with habitat area and condition measured at Vasey's Paradise since 1994.	
Description of Data sets	Data Format	<ul style="list-style-type: none"> • TerraModel polygons. • Tabular data in Excel spreadsheets.
	Dates of Collection	1995 – 2002
	Quantity of Data	50 – 100 KAS surveys. 100 MB in tabular data, 1994-2001.
	Location of Data	AGFD
	Major Tasks	<ul style="list-style-type: none"> • Convert TerraModel polygons to ARC/INFO polygon layer. • Define table structure for data. • Renumber historical patch_ids. • Convert Excel to csv. • Import csv to Oracle.
	Time required to complete tasks	
	List of Entities & Attributes	Station_id? Date/time Patch_id Plot# Plot area zone Snail count
	Search Keys and tables	
Current Collection Method	<ul style="list-style-type: none"> • Four sampling trips per year from 1994 through 2000. • Polygons (patches) are identified based on Habitat/vegetation types. • Polygons (patches) are mapped using total station survey instruments. • Slope and aspect of polygons is important. • Within each of the 30-40 habitat (patch) polygons that are mapped on each trip, typically 10 plots are sampled for vegetative composition, soil type. • Patch_ids prior to 2000 are inconsistently numbered. 	
Future Collection Method	Two sampling trips per year from 2001 on. Three KAS augmentation sites are sampled only for snail numbers since 1997? Patch_ids from 2000 on will be consistent. Field data capture application. Data entry GUI.	
Data Analysis	KAS population estimates by year (methods being re-evaluated). Habitat change by year. Habitat change as function of yearly discharge patterns. Changes in spatial distribution of population by year.	
Data Query		
Priority		

GCMRC Archives? Yes
 Metadata? No
 GCMRC Pilot? No

4.1.5.3.2 Stage Related Vegetation Transects

Data set Name	Stage Related Vegetation Transect	
Data set Area	Terrestrial Biology	
Data set Subject	Vegetation	
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Michael Kearsley	
Designer ERD / Entity Names(s)	Veg/Transect	
Description of Project	<p>The purpose of these transect studies is to monitor vegetation composition and dynamics related to river stage.</p> <p>A total of 60 stage related vegetation transects are sampled each year. Of these, 20 transects are re-visited from the previous year, and 40 will be newly selected at random.</p>	
Description of Data sets	Data Format	MS Excel
	Dates of Collection	2000 – 2001
	Quantity of Data	
	Location of Data	NAU, GCMRC Archive
	Major Tasks	Convert Excel files to csv. Define Oracle table structure. Import csv files into Oracle tables.
	Time required to complete tasks	

	<p>List of Entities & Attributes</p>	<p>Transect</p> <ul style="list-style-type: none"> • Station_id • Station_type • Trip_id • Date • Stage • Species Code (gcmrc code) • %Cover • Substrate (%sand or finer) <p>Veg Density</p> <ul style="list-style-type: none"> • Site • Transect • Month • Sample# • 1stNative • 2ndNative • 3rdNative • 1stTami • 2ndTami • 3rdTami <p>Plant Species List (Natural Resource Conservation Service (NRCS) Plants database)</p> <ul style="list-style-type: none"> • Species Code (gcmrc code) • Common Name • Kingdom • Phylum • Class • Order • Family • Genus • Species • Sub Species <p>Transect Stations</p> <ul style="list-style-type: none"> • Station_id • Station_type • Start_X • Start_Y • End_X • End_Y • Photo • Description
	<p>Search Keys and tables</p>	
<p>Current Collection Method</p>	<p>Transects run perpendicular to river, crossing 5 stage zones, from waters edge to the 60,000 cfs stage. River mile and side reference Station_ids (transect ids). Transect end points are monumented (marked), and locatable by coordinates and air photos. That is, they are repeatable sites. Stage = 15,000; 25,000; 35,000; 45,000; 60,000 cfs</p>	

	Where a transect intersects each stage, 4 plots are laid out perpendicular to the transect, using 1m square sighting frames. Within each plot, the percent cover of each plant species present is recorded, along with the soil texture within the plot.
Future Collection Method	Same as current methods.
Data Analysis	<ol style="list-style-type: none"> 1) Total cover by stage 2) Richness by stage 3) Biodiversity (number of species) stage 4) Wetland indicator score by stage 5) Compositional status by stage 6) Year to year variation by site and zone 7) Power analysis?
Data Query	Must be able to link to integrated (bird, bug) sites.
Priority	

GCMRC Archives? Yes
 Metadata? Little
 GCMRC Pilot? No

4.1.5.3.3 Integrated Terrestrial Vegetation Patch

Data set Name	Integrated Veg Patch	
Data set Area	Terrestrial Biology	
Data set Subject	Vegetation, Avifauna, Mammal, Invertebrate, Reptile & Amphibian	
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Michael Kearsley, Helen Yard, Neil Cobb	
Designer ERD / Entity Names(s)		
Description of Project	<p>These are integrated vegetation, bird, invertebrate and vertebrate sample transects done at bird patch sites. Bird patch boundaries are based on vegetation type.</p> <p>Transects are run parallel with water's edge, in each of three zones, OHWZ (> 90k cfs), NHWZ (30-40k cfs), and Water's edge (WAT). Transects are laid out in order to characterize each zone within the patch. Minimum 100m habitat length is required for a given transect. Transect locations are uniquely identified, with recorded coordinates. Some transects are re-visited.</p> <p>Samples (plots) are taken at 10 m intervals along each transect. At each plot vegetation structure is measured by setting a structure pole perpendicular to the transect line at each sample point. Traps are set for bugs, mammals and herps. Bird point counts are performed at plots.</p> <p>A total of 14 bird patches have such transects.</p> <p>All Bird Patches (64/yr) have associated vegetative structure measurements.</p>	
Description of Data sets	Data Format	ARC/INFO polygon layer of patch boundaries. Excel files of tabular data
	Dates of Collection	1998 - 1999
	Quantity of Data	
	Location of Data	NAU

	Major Tasks	<ul style="list-style-type: none">• Convert Excel to csv files.• Create ARC/INFO coverage of patch polygons.• Program database tables in Oracle.• Import csv files into Oracle tables.
	Time required to complete tasks	

	<p>List of Entities & Attributes</p>	<p>Vegetation structure (summarized for each transect)</p> <ul style="list-style-type: none"> • Trip id • Patch id (rm, side, zone) • Date • Transect id • Zone • Species • Total # hits/20 pole plants • Total Vegetative Volume (TVV) <p>Summary</p> <ul style="list-style-type: none"> • Sample(MMMYPPP) • Species • Cover • Trip • Site • Zone • TVV <p>Plant Species List</p> <ul style="list-style-type: none"> • ID • FieldAbbr • NtlAbbr • Division • Family • Genus/Species • CommonName • Exotic <p>Bird point counts</p> <p>Walking bird survey</p> <p>Invertebrate trap points</p> <p>Mammal trap points</p> <p>Herp trap points</p> <p>Walking herp survey</p> <p>Bird patch</p> <ul style="list-style-type: none"> • Patch id • Polygon area • Vegetation type <p>Summary counts</p> <ul style="list-style-type: none"> • Patch id • Date • Zone • Species • Number of individuals • Sample method <p>Bird Species List</p> <p>Mammal Species List</p> <p>Invertebrate Species List</p> <p>Reptile & Amphibian Species List</p>
	<p>Search Keys and tables</p>	

Current Collection Method	
Future Collection Method	
Data Analysis	Total vegetative volume by site, zone and date.
Data Query	
Priority	

GCMRC Archives? Yes
 Metadata? Little
 GCMRC Pilot? No

4.1.5.4 Mammals

Data set Name	Mammals	
Data set Area	Terrestrial Biology	
Data set Subject	Mammal	
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Mike Kearsley, Jennifer Frye	
Designer ERD / Entity Names(s)		
Description of Project		
Description of Data sets	Data Format	
	Dates of Collection	
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	Same as in Integrated Bird Patch
Search Keys and tables		
Current Collection Method	Sherman Live Traps	
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.5.5 Invertebrates

Data set Name	Invertebrates	
Data set Area	Terrestrial Biology	
Data set Subject		
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	David Lightfoot, Neil Cobb, Sandra Brandtley	
Designer ERD / Entity Names(s)		

Description of Project		
Description of Data sets	Data Format	MS Excel
	Dates of Collection	Arthropods: 8/29/2001 – 9/12/2002
	Quantity of Data	
	Location of Data	NAU, UNM. Some in GCMRC archive
	Major Tasks	Sorting and ID of invertebrates
	Time required to complete tasks	
	List of Entities & Attributes	<ul style="list-style-type: none"> • Period • Date • RM • Zone • Counts (by Taxon code).
	Search Keys and tables	
Current Collection Method	Sweep Nets, Pitfall and Light Traps set at Integrated Bird Patch sites.	
Future Collection Method		
Data Analysis		
Data Query		
Priority		

GCMRC Archives? Some
 Metadata? No
 GCMRC Pilot? No

4.1.5.6 Reptiles and Amphibians

Data set Name	Reptiles and Amphibians	
Data set Area	Terrestrial Biology	
Data set Subject		
GCMRC Contact	Barbara Ralston	
Principal Investigator(s)	Geoff Carpenter, Neil Cobb	
Designer ERD/Entity Names(s)		
Description of Project		
Description of Data sets	Data Format	
	Dates of Collection	
	Quantity of Data	
	Location of Data	UNM
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	Same as Integrated Bird Patch sites
	Search Keys and tables	
Current Collection Method	Walking surveys	
Future Collection Method	same	
Data Analysis		
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No

GCMRC Pilot? No
 4.1.5.7 Taxonomy

Data set Name	Taxonomy	
Data set Area	Biology, Hydrology	
Data set Subject	Aquatic Species, Terrestrial Species	
GCMRC Contact	Lew Coggins, Barbara Ralston, Susan Hueftle, Mike Yard	
Principal Investigator(s)	GCMRC	
Designer ERD/Entity Names(s)	BIO GENERAL / TAXONOMY	
Description of Project	<p>All biological data collected, whether aquatic or terrestrial, is species specific, and therefore requires a lookup table of species codes and taxonomic information, so that consistency can be maintained in data collection efforts.</p> <p>It has been suggested that a single table of species codes be maintained, containing those terrestrial and aquatic species encountered in the Grand Canyon. From this master table, sub-tables can be generated that only contain species specific to a certain study, such as fishes. A subject area code would allow one to extract these specific species lists.</p> <p>This table will be based on existing resources provided by such organizations as ITIS - Integrated Taxonomic Information System, and the American Ornithologists Union, among others.</p> <p>Care must be given to organize this table in such a way to eliminate duplicate species codes between disciplines.</p>	
Description of Data sets	Data Format	
	Quantity of Data	
	Location of Data	
	Major Tasks	1) Obtain taxonomic information from governing organizations. 2) Define unique species codes for each discipline. 3) Create table in Oracle 4) Populate table with GCMRC specific species.
	Time required to complete tasks	
	List of Entities & Attributes	Taxonomy <ul style="list-style-type: none"> • Species_ID • Species_Code • Subject Area • Common Name • Kingdom • Phylum • Class • Order • Family • Genus • Species • Sub-species • Photograph • Native (Y/N) • Date Identified in GC • Invasive (Y/N) • Endangered (Y/N)
	Search Keys and tables	

Current Collection Method	
Future Collection Method	
Data Analysis	
Data Query	List of species found in GC, species native to GC, non-native fishes, etc.
Priority	

GCMRC Archives? Some species

Metadata? No

GCMRC Pilot? No

4.1.6 4.1.6 Aquatic Biology

4.1.6.1 Native and Exotic Fish

Data set Name	Fishery								
Data set Area	Aquatic Biology								
Data set Subject	Fish (Native, Non-native)								
GCMRC Contact	Lew Coggins								
Principal Investigator(s) (Current)	Bill Persons (AGFD, Trout) Scott Rogers (AGFD, FMS) Dave Speas (AGFD, LF Trout) Dennis Stone (USFWS) Randy VanHaverbeek (USFWS) Melissa Trammel (SWCA) Randy VanHaverbeke (USFWS) Various Historical PIs								
Designer ERD / Entity Names(s)	BIO FISH / FISH SAMPLE, FISH SPECIMEN, FISH PIT TAG								
Description of Project	<p>Native and non-native fish studies for the assessment of stock, size and habitat have been ongoing in the Colorado River and tributaries since 1978. Legacy fishery data have been consolidated into a MS Access database.</p> <p>A review of sampling methods took place during the fall of 2000 in order to define the minimum effort required to monitor absolute and relative fish population trends.</p> <p>Fish Studies include sampling for all species, including the following critical ones:</p> <table border="1"> <thead> <tr> <th>Category</th> <th>Species</th> </tr> </thead> <tbody> <tr> <td>Endangered Native Fish</td> <td>Humpbacked chub Razorback sucker</td> </tr> <tr> <td>Non-endangered Native Fish</td> <td>Flannel-mouth sucker Bluehead sucker Speckled Dace</td> </tr> <tr> <td>Non-native fish</td> <td>Rainbow trout Brown trout Brook trout</td> </tr> </tbody> </table> <p>Approximately 30 different species are represented in the roughly 400,000 specimens collected to date. Sampling techniques include Electrofishing, angling, netting, seining, and Transect sampling. Sampling techniques used vary depending on target species and habitat type.</p> <p>A standard component of fish sampling is tagging. Three types of fish tagging are commonly used – Floy, PIT (Passive Integrated Transponder), and Visual Implant Tag (Color). Support for 2 types of tags per fish is required. Floy tags</p>	Category	Species	Endangered Native Fish	Humpbacked chub Razorback sucker	Non-endangered Native Fish	Flannel-mouth sucker Bluehead sucker Speckled Dace	Non-native fish	Rainbow trout Brown trout Brook trout
Category	Species								
Endangered Native Fish	Humpbacked chub Razorback sucker								
Non-endangered Native Fish	Flannel-mouth sucker Bluehead sucker Speckled Dace								
Non-native fish	Rainbow trout Brown trout Brook trout								

	<p>were used historically, but most (if not all) floy-tagged fish are now expired. Floy tagging may be revived for future studies, as it presents a more economical alternative to PIT tags.</p> <p>Historically, fisheries survey data was delivered to GCMRC in a number of electronic forms, including DBase and spreadsheet formats. Within the last year the GCMRC has developed a MS Access application that mirrors the paper survey forms that have typically been used in the field. A template version of this MS Access application for Data Capture (DC) is provided to contractors, with the intent that they will populate this template with catch data, and submit the database to GCMRC as a deliverable. Data in the DC are then integrated with a master version of the MS Access database, which contains all available historical fish records.</p> <p>A considerable number of errors and exceptions exist in the MS Access database of historical data. Exceptions arise from different investigators using a variety of codes to characterize different sample parameters. Errors also arise from data recording and entry techniques that do not enforce integrity relationships.</p> <p>One example of errors introduced in historical catch rate data arises from the fact that certain investigators recorded multiple fish specimens in a single record, instead of a single record for each specimen. This deviation in the database model of one record per fish specimen results in an inaccurate calculation of Catch per Unit Effort, which in turn produces inaccurate population estimates for certain species.</p> <p>Considerable effort should be given to re-working historical data to remove these errors and exceptions.</p>	
Description of Data sets	Data Format	MS Access
	Dates of Collection	1977 - present
	Quantity of Data	410,000 specimens caught in 92,000 sampling events.
	Location of Data	GCMRC
	Major Tasks	<ol style="list-style-type: none"> 1) Design table structure in Oracle. 2) Export Access tables to CSV. 3) Import CSV files into Oracle structure. 4) Revise data input application to work with Oracle via Access ODBC. 5) Write batch data import application from CSV files to Oracle. 6) Write field data capture application. 7) Link samples to GIS centerline layer for display in A/I
	Time required to complete tasks	
	List of Entities & Attributes	See ER diagrams
	Search Keys and tables	
Current Collection Method	Paper data collection sheets, unique for Netting, Electrofishing, Seining, and Snorkeling.	
Future Collection Method	<p>Field Data Recorder</p> <p>A Field Data Recorder (FDR), comprised of a specialized Data Capture application running on field-capable computer hardware is a viable option to replace the paper based survey recording in the field. The Fisheries data set has a fairly complex set of input parameters depending on Sampling method, and represents a good argument for a FDR application, with the following advantages:</p> <ul style="list-style-type: none"> • Reduce effort and errors due to converting field records and notes to electronic form back in the office. • Allow for field based, real-time validation of input parameters, based on such things as known species ranges for weight, length, location, etc. • Allow for validation of input parameters based on tag number and last recorded values. Make sure the user doesn't change a fish's sex, or drastically change weight or length from previous capture. 	

	<ul style="list-style-type: none"> Standardize Survey methodology by forced collection of consistent data among all surveyors and all agencies – a paper form allows all parameters to be optional, whereas a FDC application can insist on certain parameters always being collected. <p>The FDR must also interface with a PIT Tag scanner, to fully automate the recording of Tag numbers. Validations based on last capture weight, length, and sex shall be built into the FDR, to eliminate common data entry errors.</p> <p>The benefits of reducing or eliminating in-office data entry will need to be weighed against the expense and exposure to breakdown that a FDR would introduce. An important aspect of the FDC decision is the choice of appropriate hardware that will withstand the rigors of the river environment – water, sand, heat, and physical abuse. One device that is currently under evaluation by GCMRC is the Juniper Systems Allegro (http://www.junipersys.com/allegro_pc.htm).</p> <p>Data Capture Application</p> <p>In the absence of a FDC, an office-based Data Capture Application is required.</p> <p>Data Transfer Application</p> <p>Finally, an application will be required to transfer data into the appropriate Oracle tables. The Data Administrator will perform these imports, either of the FDR-generated data sets, or the office-based Data Capture Applications data sets.</p> <p>Quality Assurance (QA)</p> <p>In order to support a quality control / quality assurance process throughout the collection and management of Fish Survey data, the following mechanisms will be required:</p> <ul style="list-style-type: none"> Mechanism for editing old data, in conjunction with some form of Audit Trail – Who/What/When edit performed. Mechanism for record based Quality Indicator or Confidence Level, and Quality comment. Standard set of reports summarizing ranges of data contained within each field, to be used to proof data before inclusion in database. These can be incorporated into a data entry application and used to enforce integrity of records when they are entered.
<p>Data Analysis</p>	<p>The fisheries data set represents a long (since 1977) and broad (30 species) data set. The consolidation of historical data sets allows GCMRC staff to determine how much baseline information is actually available, and to aid in the design of future sampling programs that can be compared to historical index information.</p> <p>Analysis requirements include:</p> <ul style="list-style-type: none"> Stock Assessment for native and non-native species Calibration of gear types Catch Rate / Unit Effort – by species, by river mile Species Population Estimates- Spatially and temporally, Relative Abundance, Absolute Abundance Fish Habitat characteristics- spatially and temporally Fish recapture history. Growth patterns. <p>Offline Analysis Queries</p> <p>A number of analysis queries were identified as a result of discussions with Lew Coggins, and are detailed somewhat in his Example_queries.xls. The implementation of these queries could be done either in the Web application or in a MS Access application specific to the Fisheries data set. Because of the relatively small number of users requiring use of these queries, and because of the need to support evolving and adhoc queries, the MS Access application path</p>

will probably be the favored solution.

Supertag

This query is meant to generate an input file for the program Supertag, an open population estimator. Supertag is a Visual Basic application authored by Carl Walters, and Lew Coggins, and has a handful of users in GCMRC and Arizona Game and Fish. The input to Supertag is currently generated manually in Excel, using a combination of macros and user operations, and is an error-prone and laborious process. The input file type must be CSV.

Details:

Query Parameter	Specified as	Detail
Species	Species Code	
Tag location	River and River Mile	Can specify single or multiple rivers, or Single or multiple reaches within a specific river
Recapture location	River and River Mile	Can specify single or multiple rivers, or Single or multiple reaches within a specific river
Capture Gear	Specify single, multiple, or all gears used to catch fish	
Recapture Gear	Specify single or multiple gears	
Starting Month and year/		
Ending Mo/yr		
Count recaps within tagging month	Y or N	N is default

Growth and Movement

Growth and Movement conveys the capture history of fish, and is a more serial format than Supertag. It is used as an input to Excel or SPSS for further analysis. A couple dozen users would potentially use this export file.

Query Details:

Query Parameter	Specified as	Detail
Species	Species Code	
Capture location	River, RM	Specific river or All
Recapture location	River, RM	Specific river or All

Mark Recapture #1

This query conveys capture/recapture over a short period of time, and is performed on smaller sets of data – typically 2 trips worth. The query result is typically analyzed further in Excel. The following conditions also apply to the query:

- For a particular trip or file
- Retain only unique fish by species
- Detect and remove duplicate pit tags within a data set. Report only unique and first appearance in data set
- Find any Pit tags that do not have 10 digits
- Combination of netting and specimen tables
- Query is specific to Species and Trip ID

Mark Recapture #2

This query is used to calculate Instantaneous Mark/Recapture Population Estimate, and assumes a short hiatus between trips. The query compares two trip data sets and finds recaptures between events.

	<p>Query Details:</p> <table border="1" data-bbox="558 226 1422 363"> <tr> <th>Query Parameter</th> <th>Details</th> </tr> <tr> <td>Species</td> <td></td> </tr> <tr> <td>Trip ID (Mark)</td> <td>Trip ID for Mark</td> </tr> <tr> <td>Trip ID (Recap)</td> <td>Trip ID for Recapture</td> </tr> </table>	Query Parameter	Details	Species		Trip ID (Mark)	Trip ID for Mark	Trip ID (Recap)	Trip ID for Recapture
Query Parameter	Details								
Species									
Trip ID (Mark)	Trip ID for Mark								
Trip ID (Recap)	Trip ID for Recapture								
<p>Data Query</p>	<p>PIT Tag</p> <p>Because the PIT (Passive Integrated Transponder) Tag ID represents a business key that transcends trips, studies, and contracts, making it widely and easily available to fisheries surveyors would allow faster and more accurate validation of individual fish captures. The PIT Tag ID can be used to look up the capture history of an individual fish, providing the user with not only the data, time, and location of capture, but also previous fish metrics such as weight, length, sex, etc.</p> <p>The consolidated fisheries data set includes approximately 6500 PIT tags, with some fish being recaptured up to 21 times. Some glaring errors exist in these data, such as drastic changes in weight, length, or sex of a particular fish through time. These no doubt indicate data recording or entry errors, which could be minimized by ready access to existing PIT Tag information in a field data recorder application.</p> <p>A web query by PIT tag ID is required. Query output would display Capture history of the tagged fish, including when, where, length, weight, and sex for each Mark and Recapture.</p> <p>Export full data set</p> <p>A number of users would like access to the complete fisheries data set. Because this data set is relatively complex, with a large number of tables and associated relations, it may be necessary to provide access to the complete fisheries data set. A number of options exist to perform this task, including:</p> <p>MS Access Export</p> <p>One method to export the complete Fisheries data set would be to provide access to a MS Access version of data set. This would require taking a snapshot of the data set as maintained in Oracle, and then making it available for download via a web interface. This may be difficult to do in a totally automated fashion, thus presenting a maintenance issue for the data administrator.</p> <p>CSV File Export</p> <p>Another method to export the complete Fisheries data set would be to provide easy access to the Oracle export CSV files. The proposed generic Oracle table interface allows the user to export one table at a time. Because of the large number of tables in the fisheries data sets, this becomes somewhat laborious for the user. A shortcut would be provided that would export all the pertinent tables in one shot, and package them as a single unit, possibly as a zip file. The next step would possibly entail importing them into a MS Access template file. This import could be automated to a degree – what files to look for and where to put them. The advantage of this process is that table relationships and business logic would be encapsulated in the MS Access file, and wouldn't have to be recreated every time a refreshed data set was downloaded. This would allow the user to have custom queries or business logic in their own personal version of the MS Access file.</p>								
<p>Priority</p>	<p>High</p> <p>Because of the advanced stage of modeling efforts for the Fish Data set, the consolidation of most legacy data sets that has already taken place, and the relative importance of the Fisheries data set in the GCMRC business, it is a likely candidate for inclusion in the Phase I Pilot project.</p>								

	Discussions with GCMRC IT personnel have also placed the development of a Field Data Capture application as a high priority. A Fisheries trip in March 2002 presents a good opportunity to Beta test a version of this application under real-world conditions.
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

GCMRC Archives? Yes
 Metadata? Some (Report by BioWest)
 GCMRC Pilot? Yes

4.1.6.2 Fish Meristics

Data set Name	Fish Meristics	
Data set Area	Aquatic Biology	
Data set Subject	Fish	
GCMRC Contact	Lew Coggins	
Principal Investigator(s)	Richard Valdez, BioWest	
Designer ERD / Entity Names(s)		
Description of Project	BioWest measured physical parameters of Humpback Chub from 1990 - 1994.	
Description of Data sets	Data Format	DBase III
	Dates of Collection	1991 - 1995
	Quantity of Data	6,594 records in two files.
	Location of Data	GCMRC Archive. P:\biological\legacy\FishStudies\BIOWEST\Delivery19951211\chub9393 and chub94.
	Major Tasks	<ul style="list-style-type: none"> • Define Oracle tables • Import dBase files
	Time required to complete tasks	2 days
	List of Entities & Attributes	Chub Meristics <ul style="list-style-type: none"> • Fish ID • Tagnum • TL • FL • SL • WT • P1_P2 • ND • CPL • CPMAXXD • CPMIND • HEAD_LN • SNOUT_LN • DORSAL_FB • ANAL_FB • BODY_DEPTH • DORSAL_RAY • ANAL_RAY

	Search Keys and tables	
Current Collection Method	HBC caught in netting efforts were measured for a variety of physical parameters in addition to length and weight.	
Future Collection Method	Not currently collected, but may again be collected in the future.	
Data Analysis		
Data Query		
Priority		

GCMRC Archives? Yes
 Metadata? Yes
 GCMRC Pilot? No

4.1.6.3 Fish Genetics

Data set Name	Fish Genetics	
Data set Area	Aquatic Biology	
Data set Subject	Fish Genetics	
GCMRC Contact	Lew Coggins	
Principal Investigator(s)	Michael Douglas (ASU)	
Designer ERD / Entity Names(s)		
Description of Project	DNA sequencing for HBC using fin clip samples.	
Description of Data sets	Data Format	
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Fish are caught in traditional netting samples within the LCR and 30 mile populations of HBC. Fin clips of HBC are taken and sent to lab for genetic analysis. Genetic makeup of individuals is compared to those of other geographically distributed populations.	
Future Collection Method	Same	
Data Analysis	Can the 30 mile population of Chubs be characterized as independent of the LCR population.	
Data Query		
Priority		

GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

4.1.6.4 Fish Stomach Samples

Data set Name	Fish Stomach Samples	
Data set Area	Aquatic Biology	
Data set Subject	Fish	

GCMRC Contact	Lew Coggins	
Principal Investigator(s)	BioWest, AGFD, USFWS	
Designer ERD/Entity Names(s)		
Description of Project	Analysis of stomach contents to determine eating habits of native and non-native fishes.	
Description of Data sets	Data Format	DBaseIII, CSV ASCII text
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Stomachs of certain non-native fishes are preserved for later analysis. Stomachs of HBC are pumped and preserved for later analysis (legacy only).	
Future Collection Method		
Data Analysis		
Data Query		
Priority		

4.1.6.5 Aquatic Food Base (Drift & Benthic)

Data set Name	Benthic	
Data set Area	Aquatic Biology	
Data set Subject	Aquatic Food Base (drift & benthic)	
GCMRC Contact	Mike Yard	
Principal Investigator(s)	Joseph Shannon, Dean Blinn (NAU)	
Designer ERD/Entity Names(s)		
Description of Project	<p>Aquatic microorganisms are sampled at nine different locations in the Colorado River, typically twice per year. Historically, these were sampled with greater frequency.</p> <p>Data collected by NAU from 1997 to present is due to be delivered to GCMRC by mid summer 2002. Data collected prior to this may be unavailable to GCMRC.</p>	
Description of Data sets	Data Format	Systat
	Quantity of Data	Unkown
	Location of Data	NAU
	Major Tasks	1) Receive data, 1997 to present, from Joe Shannon, in CSV format. 2) Design Oracle table structure. 3) Import CSV files into Oracle stucture.
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Four different gear types are used to sample both drift and benthic microorganisms. Identification is limited to Family.	
Future Collection Method		
Data Analysis		
Data Query		
Priority		

4.1.6.6 Fish Food Base

Data set Name	Fish Food Base	
Data set Area	Aquatic Biology	
Data set Subject	Aquatic Food Base	
GCMRC Contact	Mike Yard	
Principal Investigator(s)	William Persons, David Speas (AGFD), Dennis Stone (USFWS)	
Designer ERD/Entity Names(s)		
Description of Project	Aquatic food base has been sampled by the AGFD at different sites in the Lees Ferry reach of the Colorado River, as well as the Little Colorado River. The USFWS also collected food base data from the LCR.	
Description of Data sets	Data Format	
	Quantity of Data	Unknown
	Location of Data	AGFD, USFWS
	Major Tasks	1) Receive data 2) Design Oracle table structure. 3) Import CSV files into Oracle structure.
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

4.1.6.7 Fish Elemental Analysis

Data set Name	Fish Elemental Analysis	
Data set Area	Aquatic Biology	
Data set Subject	Fish and Water	
GCMRC Contact	Lew Coggins	
Principal Investigator(s)	Brian Kennedy (U of Michigan)	
Designer ERD/Entity Names(s)		
Description of Project	This is a new project that attempts to utilize differences in strontium isotope ratios present in tributaries of the Colorado river to determine natal stream of fishes	
Description of Data sets	Data Format	CSV
	Quantity of Data	Unknown
	Location of Data	U of Michigan
	Major Tasks	1) Collect data 2) Design Oracle table structure. 3) Import CSV files into Oracle structure.
	Time required to complete tasks	Unknown

	List of Entities & Attributes	Unknown
	Search Keys and tables	Unknown
Current Collection Method	Water proof notebooks	
Future Collection Method	Fin clips or scales collected to determine Sr ratios in fish specimens. Sub-samples identified by bottle_number, similar to genetic analyses. Standard data sheets would most likely be employed.	
Data Analysis	Can natal stream be determined by Sr ratios observed in fish specimens	
Data Query		
Priority		

4.1.7 4.1.7 Cultural Resources

4.1.7.1 Archaeology

Data set Name	Archaeological Sites	
Data set Area	Cultural Resources	
Data set Subject	Archaeology	
GCMRC Contact	Ruth Lambert	
Principal Investigator(s)	Ruth Lambert, Nancy Coulam (BOR), Jan Balsom (GCNP), GCRA	
Designer ERD/Entity Names(s)		
Description of Project	<p>Grand Canyon NP has mapped approximately 100 archaeological sites in ARC/INFO, and monitors site condition on a bi-annual basis. Attributes such as inventory, site condition, use, and erosion are recorded in a MS Access database that is maintained by GCNP personnel. Mitigation activities are also logged by the NPS. These site maps could be appended into a single polygon layer of archaeological sites for the entire Grand Canyon and Glen Canyon Recreation Area.</p> <p>GCMRC has mapped some of these sites using conventional total station survey techniques.</p> <p>Other potential providers of cultural data are the Hopi, Havasupai, Hualapai, Navajo, and Paiute tribes, as well as the Glen Canyon Recreation Area.</p> <p>Access to this information is considered highly confidential, and must be restricted to the data providers themselves, and to select others with permission of the providers.</p> <p>Final reports of investigation are archived at the GCMRC library. Any information about site location is censored from these reports.</p>	
Description of Data sets	Data Format	ARC/INFO, Terra-Model?
	Quantity of Data	
	Location of Data	GCNP (NAU), Tribes
	Major Tasks	<ol style="list-style-type: none"> 1) Distribute data questionnaire to cultural data providers and users. 2) Derive a security model for providers and users of cultural data. 3) Demonstrate security model and access methods to cultural data. 4) Arrive at consensus on what, if any data are to be stored in a consolidated database at GCMRC. 5) Program spatial data tables. 6) Import data from NPS. 7) Import data from tribal providers.
	Time required to complete tasks	

	List of Entities & Attributes	Arch Sites <ul style="list-style-type: none"> • Site_id • Station_id • X_COORD • Y_COORD • Elevation • Stage • Comments Arch Site Condition <ul style="list-style-type: none"> • Site_id • Date • Condition • Comments Arch Site Details <ul style="list-style-type: none"> • Site_id • Date • Artefact • Description Arch Site Actions <ul style="list-style-type: none"> • Site_id • Date • Action_taken • Comments
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis	GCMRC: Select all sites inundated by a given stage.	
Data Query	Data provider: Full administrative access. PI and/or committee participants: Access by permission of the provider. Public: Access is restricted to final reports.	
Priority		

4.1.7.2 Tribal Traditional Resources

Tribal representatives must be consulted in order to develop the format and parameters that will be used in developing these data.

4.1.7.3 Tribal Sacred Resources

Tribal representatives must be consulted in order to develop the format and parameters that will be used in developing these data.

4.1.7.4 Mitigation and Monitoring

Nancy Coulam of the U.S. Bureau of Reclamation is contracting for a future monitoring and mitigation plan.

4.1.8 4.1.8 Recreation

4.1.8.1 Campsites

Data set Name	Campsite Condition	
Data set Area	Recreation	
Data set Subject	Campsite Condition	
GCMRC Contact	Ruth Lambert	
Principal Investigator(s)	Matt Kaplinski	
Designer ERD/Entity Names(s)		
Description of Project	<p>Campsite sandbar conditions have been monitored for the past 10 years, in conjunction with NAU sandbar studies.</p> <p>Camp-able surface areas are measured annually, as well as such subjective attributes as condition, use and suitability.</p>	
Description of Data sets	Data Format	
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Total station survey techniques, resulting in topographic maps of campsites.	
Future Collection Method		
Data Analysis		
Data Query	<p>Camp-able area as a function of river stage.</p> <p>Camp-able area as a function of year.</p>	
Priority		

4.1.8.2 User Attitudes

Data set Name	User Attitudes	
Data set Area	Recreation	
Data set Subject	User Attitudes	
GCMRC Contact	Ruth Lambert	
Principal Investigator(s)	William Stuart (U of Illinois)	
Designer ERD/Entity Names(s)		
Description of Project	<p>Questionnaires are formulated to determine the attitudes and opinions of various recreational users of both the Glen Canyon Recreation Area and the Grand Canyon river system. These users have included fishermen, and both recreational and commercial river runners. Studies of this type have been done since the mid-1980s.</p> <p>Typically, questionnaires are constructed with specific unique study in mind, and thus it might be difficult to combine raw data from disparate studies into a single database table structure.</p> <p>Final Reports of these studies will be included in the data management framework, however.</p>	
Description of Data sets	Data Format	CSV ASCII text, MS Word
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	

	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query	Public: Web access to final reports, searchable on study name, contract number, and subject. Pis: May want to download the CSV file for a specific study. No reason to consolidate raw data from multiple studies into a single database.	
Priority	Low priority prior to year 2000, when reports were delivered in hard copy form.	

4.1.8.3 Safety

Data set Name	Safety	
Data set Area	Recreation	
Data set Subject	Safety	
GCMRC Contact	Ruth Lambert	
Principal Investigator(s)	Linda Talbert (NPS)	
Designer ERD/Entity Names(s)		
Description of Project	River safety has been monitored periodically, in order to evaluate the effects of river dam releases on the frequency of accidents at certain rapids on the Colorado River. Final Reports of these studies will be included in the data management framework.	
Description of Data sets	Data Format	CSV ASCII text, MS Word
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query	Public: Web access to final reports, searchable on study name, contract number and subject. The public may want to download the CSV file for a specific study. No reason to consolidate raw data from multiple studies into a single database.	
Priority	Low priority prior to year 2000, when reports were delivered in hard copy form.	

4.1.9 Base Data

GCMRC collects and manages many base data sets that act as spatial reference frameworks for other data collection efforts.

Base Data sets require a variety of access methods.

- Base data sets should always be available to the public for download as individual files.
- Suitable data sets should also be available via direct connection from PC GIS applications (such as ArcGIS). This capability will be restricted to Intranet and Extranet users.
- Some data sets should be visible as base layers through the Web based spatial query application.
- Some data sets should be available to Intranet, Extranet and Internet users via an ArcIMS Feature Server

Base Data sets are described in the following tables:

4.1.9.1 Survey Control

Data set Name	Survey Control Points	
Data set Area	Base Data	
Data set Subject	Survey Control	
GCMRC Contact	Mark Gonzales, Keith Kohl	
Principal Investigator(s)	Keith Kohl	
Designer ERD/Entity Names(s)	SURVEY / CONTROL	
Object Class	Control	
Description of Project	A network of control points is maintained throughout the canyon in the Arizona State Plane coordinate system. This network of points will be imported into an ArcSDE feature class that can be queried both spatially and textually.	
Description of Data sets	Data Format	MS Excel, ARC/INFO
	Quantity of Data	
	Location of Data	GCMRC
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	Control Points <ul style="list-style-type: none"> • PID# • X_COORD • Y_COOR • ELEVATION • Datum • Comments • Site Photograph (TIFF image)
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis	Control point locations are recomputed in a variety of coordinate systems, for convenience of use. Some data sets are located using these control points, such as NAU sandbar topographic surveys, or Kanab Amber Snail habitat surveys. Any refinement in a control point location would therefore affect spatial data collected from that point, and provision must be made to later alter that data as well.	

Data Query	Subsets of control points are used to load GPS units before the start of a field trip. There should be the ability to query and extract these subsets in the appropriate format for field use.
Priority	

4.1.9.2 Integrated Long Term Monitoring Reaches

Feature Class	Integrated Long Term Monitoring Reaches
Feature Data set	Base Data
Feature Class Subtypes	
Description	Eleven Reaches are defined, by polygonal boundaries, in which intensive historical data collection has been focused.
Spatial Data Type	Polygon
Other Feature Attributes	REACH_ID Name UP_RM DOWN_RM Length (m)
IMS Base Layer	Yes

4.1.9.3 USGS Quarter-Quad Boundaries

Feature Class	USGS Quarter-Quad Boundaries
Feature Data set	Index
Feature Class Subtypes	
Description	USGS Quarter-Quad boundaries are used as a spatial reference system for other data sets (such as orthophotos).
Spatial Data Type	Polygon
IMS Base Layer	Yes

4.1.9.4 River Mile Index

Feature Class	Rivers
Feature Data set	Base Data
Feature Class Subtypes	<ul style="list-style-type: none"> • Main Stem Colorado • Lake Powell • Perennial Tributary • Intermittent Tributary
Description	<p>Several different River Mile systems have been used in the Grand Canyon:</p> <ul style="list-style-type: none"> • GCMRC Centerline (in development) • Stevens River Guide • Bellknap River Guide • USGS • Bureau of Reclamation • Lake Powell centerline <p>Most subject area data sets use a river mile index as their primary spatial reference. That is, the river mile constitutes a repeatable sampling "station". However, each river mile mapping system in use is slightly different.</p> <p>The GCMRC standard centerline is currently being developed using modern orthophotography obtained during the year 2000 over flights. When complete, this coverage will consist of an ARC/INFO Network defining the river centerline according to GCMRC, with nodes positioned in 1/10 mile increments, and a Node Attribute Table with</p>

	<p>columns for each river guide mileage. Each river mile "station" will be assigned State Plane coordinates based on this coverage.</p> <p>In the Grand Canyon, river mileages are measured downstream from Lees Ferry, with zero located at the origin defined by the 1923 USGS topographic mapping expedition.</p> <p>In the Colorado River between Glen Canyon Dam and Lees Ferry, river mileages are measured as negative numbers, increasing upstream from Lees Ferry.</p> <p>Lake Powell centerlines are mapped from the historical Colorado River and tributaries, which are now inundated. These are measured in river kilometers upstream from Glen Canyon dam.</p> <p>All tributary centerlines are mapped with river kilometers measured from their intersection with the Colorado River.</p>
Spatial Data Type	Network/Line
IMS Base Layer	Yes

4.1.9.5 LIDAR

Name	LIDAR
Description	Data acquire using LIDAR produces georeferenced elevations for ground surface features.
Spatial Data Type	Points, Lines
Format	Base file: DXF
Derived Products	TIN, DEM
IMS Base Layer	No

4.1.9.6 Orthophotos

Name	Orthophotos
Description	A georeferenced, orthorectified aerial photo coverage of the entire river corridor.
Spatial Data Type	Raster
Format	TIFF, MrSID
IMS Base Layer	No

4.1.9.7 Aerial Photographs

Name	Aerial Photographs
Description	The base data from which orthophotos are derived.
Spatial Data Type	Raster
Format	TIFF
Derived Products	Orthophotos
IMS Base Layer	No

4.1.9.8 Aerial Flight Line Index

Name	Aerial Flight Line Index
Description	Each Aerial photo data set has an associated coverage of flight line and exposures showing the footprints (extents) of each photograph.
Spatial Data Type	Polygon
Format	Feature Class Index, Subtype Flight Lines
IMS Base Layer	Yes

4.1.9.9 High-Gain Aerial Photography

Feature Class	High-Gain Aerial Photography
Feature Data set	Base Data
Feature Class Subtypes	
Description	High gain photography penetrates the water surface to reveal details of the riverbed.
Spatial Data Type	Image
Format	TIFF
IMS Base Layer	No

4.1.9.10 Cross Section Locations

Name	Wilson-Stars Cross Section Locations
Description	A line coverage of the cross-section locations of the Colorado River channel, including the Wilson, Stars Interpolated, and NAU cross-sections.
Spatial Data Type	Line
Format	Feature Class Channel Geometry, Subtype Cross Sections
IMS Base Layer	Yes

4.1.9.11 USGS Permanent Gage Locations

Name	USGS Permanent Gage Locations
Description	A line coverage of permanent USGS gage locations on the Colorado River and tributaries.
Spatial Data Type	Line
Format	Feature Class Gages, Subtype USGS Permanent Gages
IMS Base Layer	Yes

4.1.9.12 GCMRC Stream and Weather Gage Locations

Name	GCMRC Stream/Weather Gages
Description	Locations of GCMRC maintained stream and weather gages on the "un-gaged" tributaries of the Colorado River
Spatial Data Type	Point
Format	Feature Class Gages, Subtype GCMRC Stream/Weather Gages
IMS Base Layer	Yes

4.1.9.13 GCMRC Temperature Gage Locations

Name	GCMRC Temperature Gages
Description	Locations of the continuous water temperature monitors placed by GCMRC
Spatial Data Type	Point
Format	Feature Class Gages, Subtype GCMRC Temperature Gages
IMS Base Layer	Yes

4.1.9.14 Geomorphic Reaches (Graff and Schmidt)

Name	Geomorphic Reaches
Description	Graff and Schmidt geomorphic reaches
Spatial Data Type	Polygon
Format	Feature Class Geomorphology, Subtype G&S Geomorphic Reaches
IMS Base Layer	Yes

4.1.9.15 Geomorphic Reaches (Melis)

Name	Melis Geomorphic Reaches
Description	Geomorphic reaches refined by Melis
Spatial Data Type	Polygon
Format	Feature Class Geomorphology, Subtype Melis Geomorphic Reaches
IMS Base Layer	Yes

4.1.9.16 Sediment Yield Reaches

Name	Sediment Yield Reaches
Description	Reaches defining sediment yield characteristics (Webb, Griffiths, Melis)
Spatial Data Type	Polygon
Format	Feature Class Geomorphology, Subtype Sediment Yield Reaches
IMS Base Layer	Yes

4.1.9.17 Sediment Yield Basins

Data set Name	Sediment Yield Basins	
Data set Area	Sediment Transport, Hydrology, Climate	
Data set Subject	Fine Grained, Coarse Grained, Stream Flow, Modeling, Rainfall, Weather Service	
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Robert Webb, Peter Griffiths, David Topping	
Designer ERD / Entity Names(s)		
Description of Project		
Description of Data sets	Data Format	ARC/INFO polygon layer
	Quantity of Data	None
	Location of Data	Non-existent
	Major Tasks	1) Create polygon coverage of drainage basins.
	Time required to complete tasks	
	List of Entities & Attributes	Basins <ul style="list-style-type: none"> • Basin_id • Tributary_name • Sediment_yield_reach • Drainage_area (sq km) • Sediment_yield (mg/yr)
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

4.1.9.18 HUC Boundaries

Name	HUC Boundaries
Description	USGS Hydrologic Unit Boundaries
Spatial Data Type	Polygon
Format	Feature Class Geomorphology, Subtype HUC Boundaries
IMS Base Layer	Yes

4.1.9.19 Shoreline Habitat (Korman)

Name	Korman Shoreline Habitat
Description	Fish habitat as defined by Korman (after Stevens).
Spatial Data Type	Line
Format	Feature Class Korman Shoreline Habitat
IMS Base Layer	No

4.1.9.20 Shoreline Habitat (Yard, Wyse and Mietz)

Name	Shoreline Habitat
Description	Fish habitat as defined by Yard, Wyse, and Mietz
Spatial Data Type	Line
Format	Feature Class GCMRC Shoreline Habitat
IMS Base Layer	Yes

4.1.9.21 Fish Habitat Transect Locations (USFWS)

Name	USFWS Fish Habitat Transects
Description	Fish habitat transects surveyed by the USFWS on the Little Colorado River
Spatial Data Type	Line
Format	Feature Class USFWS Fish Habitat
IMS Base Layer	No

4.1.9.22 Stage Related Vegetation Transect Locations

Name	Stage Related Vegetation Transects
Description	A line coverage of transects along which stage related vegetation inventories are made.
Spatial Data Type	Line
Format	Feature Class Vegetation, subtype Stage Related Transects
IMS Base Layer	No

4.1.9.23 Integrated Structure and Composition Transect Locations

Name	Integrated Vegetation Structure and Composition Transects
Description	A line coverage of the integrated bird/bug/veg sampling transects done in bird patches
Spatial Data Type	Line
Format	Feature Class Bird Patch, Subtype Integrated Transects
IMS Base Layer	No

4.1.9.24 Bird Patch Boundaries

Name	Bird Patches
Description	A polygon coverage of the bird sampling patches along the Colorado River
Spatial Data Type	Polygon
Format	Feature Class Bird Patch, Subtype Integrated Patch
IMS Base Layer	Yes

4.1.9.25 Vegetation Boundaries

Data set Name	Vegetation Polygon Layer	
Data set Area	Terrestrial Biology, Base Data?	
Data set Subject	Vegetation	
GCMRC Contact	Barbara Ralston, Steve Mietz, Phil Davis	
Principal Investigator(s)		
Designer ERD/Entity Names(s)		
Description of Project	Vegetative units along the river corridor are mapped from aerial photography. Changes in these vegetative polygons will be monitored over 5 year time scale.	
Description of Data sets	Data Format	ARC/INFO polygon layer
	Quantity of Data	
	Location of Data	GCMRC?
	Major Tasks	
	Time required to complete tasks	To be conducted over the next 2 years.
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

Name	Vegetation Types
Description	A large scale coverage dominant vegetation types in the Colorado River ecosystem.
Spatial Data Type	Polygon
Format	Feature Class Vegetation, Subtype Vegetation Types
IMS Base Layer	Yes

4.1.9.26 Kanab Amber Snail Habitat Boundaries

Name	Kanab Amber Snail Habitat
Description	A small scale coverage of the vegetation patches defining Kanab Amber Snail habitat
Spatial Data Type	Polygon
Format	Feature Class Vegetation, Subtype KAS Habitat
IMS Base Layer	No

4.1.9.27 Campsites

Name	Campsites
Description	Locations of monitored campsite beaches
Spatial Data Type	Polygon
Format	Feature Class Sandbar Sites, Subtype Campsites
IMS Base Layer	Yes

4.1.9.28 Sand Bar Sites

Name	Sand Bar Sites
Description	Locations of NAU long term sand bar monitoring sites
Spatial Data Type	Polygon
Format	Feature Class Sandbars, Subtype Long Term Monitoring Sites
IMS Base Layer	Yes

4.1.9.29 Surficial Geology

Name	Surficial Geology
Description	Holocene geologic deposits
Spatial Data Type	Polygon
Format	Feature Class Geology, Subtype Surficial
IMS Base Layer	Yes

4.1.9.30 Archaeological Site Boundaries

Name	Archaeological Sites
Description	A layer mapping known archaeological sites in the Grand Canyon
Spatial Data Type	Polygon
Format	Feature Class Archaeology, Subtype Archaeology Sites
IMS Base Layer	No (restricted access)

4.1.10 Modeling

4.1.10.1 Hydrologic Modeling - CRFSSGUI

Data set Name	CRFSSGUI or FlowGUI
Data set Area	Modeling
Data set Subject	
GCMRC Contact	Ted Melis
Principal Investigator(s)	Josh Korman (Ecometric), S.M. Wiele (USGS), T. Randle and E. Pemberton(BOR)
Designer ERD/Entity Names(s)	
Description of Project	<p>The intent of the Colorado River Flow, Stage, and Sediment Input Graphic User Interface (CRFSSGUI) is to combine three predictive hydrologic models into a single, easy-to-use application that will model river discharge, stage, and sediment input in the Colorado River system.</p> <p>The three models combined in CRFSSGUI are:</p> <p>UNSTEADY, a reach-averaged one-dimensional model of diurnal discharge wave propagation. The model was developed by S.M. Wiele at the US Geological Survey (USGS) to predict the progression and evolution of discharge waves released from Glen Canyon Dam (Wiele and Smith, 1996). A primary purpose of the model is the calculation of hydrographs at points along the river in support of field studies. The model also predicts reach-averaged water velocity that can be used to estimate travel time and distance.</p>

	<p>STARS, is a one-dimensional, steady state model designed to calculate water surface profiles (river stage) and simulate streambed response to varying water and sediment discharges (Randle and Pemberton 1987). The model was developed by the Bureau of Reclamation and applied to the Colorado River in Grand Canyon by T. Randle and E. Pemberton. While STARS simulates both stage and sediment movement, only the stage component of the model has been integrated within this application.</p> <p>Paria and LCR Sediment Input Models</p> <p>To be integrated in FY2001. Do details on this model are known at this time.</p> <p>The above model requires the following inputs:</p> <ul style="list-style-type: none"> • USGS Gage Data - Historical unverified unit value data (15 min. interval) for Glen Canyon, Lees Ferry, Grand Canyon, National Canyon, Diamond Creek on the Colorado River mainstem, and the Paria and Little Colorado River tributaries. These data will be available from the USGS Arizona District WWW server. • Glen Canyon Dam Historical Hourly - Historical hourly discharge data from Glen Canyon Dam (GCD), available from the GCMRC FTP and WWW servers • Glen Canyon Dam 6 month Hourly Forecast from Hydro LP Model - Projected Glen Canyon Dam hourly discharges based from Hydro LP model output. This model is used by the Western Area Power Administration (WAPA) to project power production from GCD. • Cross Sections: USGS-Wilson, STARS Interpolated, NAU <p>Some of the above inputs are candidates for DMF-provided Programmatic Access, detailed later in this document.</p> <p>This model is detailed in the FlowGUI UsersGuide.doc.</p>	
Description of Data sets	Data Format	MS Access database, VB application
	Quantity of Data	35 MB
	Location of Data	GCMRC archive
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

Flow GUI Model: GCMRC Archives? Yes
 Metadata? Yes
 GCMRC Pilot? No

Stars Model: GCMRC Archives? No
 Metadata? No
 GCMRC Pilot? No

Wilson Cross Sections: GCMRC Archives? No

Metadata? No
GCMRC Pilot? No

4.1.10.2 Grand Canyon Ecosystem Model

Data set Name	GCM	
Data set Area	Modeling	
Data set Subject		
GCMRC Contact	Ted Melis	
Principal Investigator(s)	Josh Korman (Ecometric)	
Designer ERD/Entity Names(s)		
Description of Project	The primary objective of developing the Grand Canyon Ecosystem Model (GCM) is to help define and prioritize research programs and experimental monitoring design.	
Description of Data sets	Data Format	
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method		
Data Analysis		
Data Query		
Priority		

4.1.11 Imagery

4.1.11.1 Imagery - Misc

Data set Name	Imagery
Data set Area	Library
Data set Subject	
GCMRC Contact	Stephanie Wyse
Principal Investigator(s)	
Designer ERD/Entity Names(s)	LIBRARY/LIB IMAGERY
Description of Project	<p>The Library has both hardcopy and digital collections of imagery that are uncataloged. These data include photos on CD, Videos, Prints to scan, maps to scan, and historic aerial photography on paper prints. An effort is being made to catalog and convert these to electronic form where possible. Depending on the image type, the resultant data files will either be stored offline on CD or DVD, or online using hard drive file storage.</p> <p>A catalog of these imagery data sets is being constructed in an MS Access database by the librarian.</p>

	The Lib Imagery entity has some of the same attributes as the Information Source entity. It contains some information, such as Scale or Resolution, or Dam Release CFS that won't fit into the Information Source entity. A Design decision will need to be made as to how best to deal with this image metadata.	
Description of Data sets	Data Format	Paper images, hardcopy slides, video footage
	Quantity of Data	
	Location of Data	GCMRC Library
	Major Tasks	Scan, catalog, convert images and catalog to Oracle
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	An Imagery data catalog is being constructed in MS Access, which points to TIFF file locations on disk storage.	
Future Collection Method	Conversion to Oracle.	
Data Analysis		
Data Query		
Priority		

4.1.11.2 Photography - Campsite

Name	Campsite Repeat Photography
Description	Northern Arizona University (NAU) has a collection of repeat photography for 34 sites, taken on a daily basis. These records are categorized as campsite monitoring, but may also be categorized as Fine-grain Sediment, as they were taken in conjunction with NAU sandbar monitoring. Some are stereoscopic.
Spatial Data Type	Feature Class
Format	TIFF
IMS Base Layer	No

4.1.11.3 Photography - Stanton Repeat

Name	Stanton Repeat Photography
Description	The Stanton Repeat photographic set includes 100's of photographs taken in the 1890's during two Stanton trips down the Colorado River. Starting in the 1990's, photos have been taken at the sites of the original Stanton photos, allowing for comparison of vegetation and topographic features. This may be an ongoing effort, with repeats performed on a regular basis.
Spatial Data Type	Feature Class
Format	TIFF
IMS Base Layer	No

4.1.11.4 Photography - Other Repeat

Name	Repeat Photography
Description	Repeat photograph is used to monitor changes in archaeological site condition, geomorphology, and terrestrial vegetation.
Spatial Data Type	Feature Class
Format	TIFF
IMS Base Layer	No

4.1.12 Library

4.1.12.1 Follet Card Catalog

Data set Name	Follet Card Catalog	
Data set Area	Library	
Data set Subject		
GCMRC Contact	Stephanie Wyse	
Principal Investigator(s)		
Designer ERD/Entity Names(s)	LIBRARY/LIB FOLLET CATALOG	
Description of Project	<p>GCMRC uses a Windows application Catalog Plus from Follet Software Co. to catalog their library resources. They currently have 2000+ resources in this catalog. These resources include journal article abstracts, technical reports, book and book chapter pertaining to areas of research in the Grand Canyon, electronic journal subscriptions, Study Final Reports, among others</p> <p>The Follet application is comprised of a suite of Windows applications, a Faircom c-tree database server, and a web application available at http://130.118.164.151/wx/s.exe.</p> <p>An ODBC Driver is currently available for both the Faircom Server as well as for Follet, to facilitate accessing and exporting the Follet data set. GCMRC does not seem to currently own the Follet ODBC driver.</p> <p>The Follet Catalog has an almost identical structure to that of the Information Source entity, which is part of the DMF Data Registry. This will allow for a straightforward export of the Follet Catalog records to the DMF, and allow us to leverage the query capabilities implemented for Information Sources.</p> <p>At the simplest level, having the Follet Catalog records imported to Oracle would allow the user to have a simple search interface to the Follet Card catalog. However, this is redundant, as query functionality is already provided by the existing Follet web interface as it already provided by the Web Collection Plus web application http://130.118.164.151/wx/s.exe.</p> <p>To leverage the catalog further would require relating the Follet records to Studies, using either Study ID, or some other business key.</p> <p>In the future, the GCMRC library may utilize the Library of Congress subject categorization, which may have application to the Information Source entity within the DMF.</p>	
Description of Data sets	Data Format	
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Follet Windows application Catalog Plus.	
Future Collection Method	<p>The Follet Catalog Plus application will continue to be used for the near future. A MS Access application would be required to export catalog records from Follet into Oracle.</p> <p>Follet may not be a long-term solution for GCMRC. This should probably be ascertained before a lot of effort is put into supporting it within the DMF.</p>	
Data Analysis		
Data Query	<p>A web interface is required that will search on all Information Source attributes. This query interface is a standard function proposed for all Information Sources in the DMF, and would not need to be specific to Follet Catalog derived records.</p> <p>The use of the Follet Web application would not likely be supplanted, as it provides a more sophisticated query interface than likely to be incorporated into</p>	

	<p>the DMF. The disadvantage of the Follet Web application is that it does not know anything about the GCMRC data model, and can't query for Catalog resource appropriately.</p> <p>Another option to tie the two systems together would be to query the Follet web application from the DMF web application, based on such known Study attributes such as Species, or PI name. This can be done with the appropriate URL with parameters. This assumes that the parameter interface is stable and consistent. This will be verified during the design phase.</p>
Priority	

4.1.12.2 Contacts

Data set Name	Contacts	
Data set Area	Library/Administration	
Data set Subject		
GCMRC Contact	Stephanie Wyse/Serena Mankiller	
Principal Investigator(s)		
Designer ERD/Entity Names(s)	See Administration Contacts data set	
Description of Project	Library staff needs a contact mailing list. This list is used when publications are sent out for peer review. This contacts list should draw from the Administration Contacts data set. Refer to this section for details.	
Description of Data sets	Data Format	MS Word
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method		
Future Collection Method	See Administration Contacts data set	
Data Analysis		
Data Query	<p>Web access to the Admin Contacts table.</p> <p>Address labels with bar codes, as per Administration user requirements, by a contact's area of expertise.</p>	
Priority		

4.1.13 Administration

GCMRC Administrative personnel track agreements and contracts with organizations that are responsible for conducting research and monitoring programs. Each element of information within the database must be traceable to the Principal Investigator and organization that collected it, and the intent of the study it was collected under. The administrative Agreement Number forms the basic attribute that will track this information within a set of administrative tables in the database.

4.1.13.1 Administrative Agreement

Data set Name	Agreements	
Data set Area	Administration	
Data set Subject		
GCMRC Contact	Serena Mankillier, Vickie Kieffer	
Principal Investigator(s)		
Designer ERD/Entity Names(s)	ADMIN/ADMIN AGREEMENT	
Description of Project	<p>A Microsoft Access file contains a list of all current GCMRC contacts by agreement number, and includes Agency, Description, Contract Monitor, PI, Expiration Date, Final Cost, Deliverables, etc.</p> <p>This data set is important in that the Agreement Number is a top-level business key to which most Studies are related.</p> <p>Current business processes only require one Agency, Contract Monitor, and Costing per Agreement. It is un-necessary to further break down costs associated with each study.</p> <p>More detailed cost tracking is performed in a separate USGS Accounting application, and at this time it is not necessary to interface with this application.</p>	
Description of Data sets	Data Format	MS Access/MS Word
	Quantity of Data	2 files < 250 records
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	<p>Agreements</p> <ul style="list-style-type: none"> • Agreement number • Description/name • Contract value • Start Date • Close Date • Principal Investigator • Organization • GCMRC COTR • Deliverables
	Search Keys and tables	
Current Collection Method	Vickie Kieffer maintains this data set in a MS Access database called Contracts.mdb. Serena Mankillier maintains another version of this information in MS Word, along with information specific to the records filing requirements of DOI.	
Future Collection Method	GCMRC Admin personnel will interface an Oracle table using a simple MS Access application and ODBC, and maintain a combined summary data set.	
Data Analysis	A simple project cost analysis could be performed using the value information contained in Contracts. At this time, it has not been determined whether the general public should have the ability to do this through the Web application.	
Data Query	Web access to the Admin Agreement Oracle table is required.	
Priority	High	

4.1.13.2 Contacts

Data set Name	Contacts
Data set Area	Library/Administration

Data set Subject		
GCMRC Contact	Stephanie Wyse/Serena Mankiller	
Principal Investigator(s)		
Designer ERD/Entity Names(s)	ADMIN/ADMIN CONTACT	
Description of Project	<p>This data sets contains a list of GCMRC contacts. Historically this data set has only contained current PI name and address information. This data set will be expanded to include all GCMRC research and monitoring contacts, and will serve both mailing and PI tracking needs at GCMRC.</p> <p>This data set is important as it provides a common Contact lookup source for a number of other entities within the GCMRC data model.</p>	
Description of Data sets	Data Format	MS Word/MS Access
	Quantity of Data	
	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Serena Mankiller currently maintains this data set in a MS Word document for personal or department use.	
Future Collection Method	GCMRC Admin personnel will interface an Oracle table using a simple MS Access application and ODBC, and maintain a combined summary data set. The Data Administrator, as well as the GCMRC Librarian could also maintain this list.	
Data Analysis		
Data Query	<p>Web access to the Admin Contacts Oracle table is needed.</p> <p>Serena Mankiller would like the capability to generate mailing labels, complete with barcodes, based on the contents on this data set.</p> <p>Stephanie Wyse would like the ability to generated mailing lists when papers are published for peer-review.</p>	
Priority	High	

4.1.13.3 Reports

Data set Name	Reports	
Data set Area	Library/Administration	
Data set Subject		
GCMRC Contact	Stephanie Wyse/Serena Mankiller	
Principal Investigator(s)		
Designer ERD/Entity Names(s)	ADMIN/ADMIN REPORT	
Description of Project	<p>Under its previous relation with the Bureau of Reclamation, GCMRC used their ARMS (Automated Records Management System) system to keep track of deliverables associated with Contracts. Under USGS, the ARMS moniker has been informally retained. Admin personnel keep track of all hardcopy permanent deliverables and file folder locations.</p> <p>Users of this information are probably limited to GCMRC Admin personnel, as access to the physical documents is restricted to Admin personnel.</p> <p>A subset of this data set is that detailing the Final Reports associated with a Contract. Whether the final reports need to be detailed in a separate entity, or whether they can be sufficiently detailed and identified in the Information Source entity will be determined during the design process.</p>	
Description of Data sets	Data Format	MS Word
	Quantity of Data	

	Location of Data	
	Major Tasks	
	Time required to complete tasks	
	List of Entities & Attributes	
	Search Keys and tables	
Current Collection Method	Serena Mankiller maintains this data set in a MS Word document for personal or department use.	
Future Collection Method	GCMRC Admin personnel will interface an Oracle table using a simple MS Access application and ODBC, and maintain a combined summary data set. The data administrator, as well as the GCMRC Librarian may also maintain this list.	
Data Analysis		
Data Query	List of Final Reports by <ul style="list-style-type: none"> • PI • Cost • Year • Subject Web access to the Oracle table is required.	
Priority	High	

4.2 GIS DATA STANDARDS

GCMRC has well defined standards for GIS data, detailed in the document [Data Standards and Delivery Requirements](#). Aspects of these standards that have an impact on the requirements for the DMF are elaborated here.

4.2.1 Coordinate system

GCMRC standards specify that spatial data sets be in the following map coordinate system:

Projection:	Arizona State Plane Central, FIPS 0202
Datum:	NAD83
Units:	Meters
Spheroid:	GRS1980
Vertical datum:	NGVD 29

Some of the data that GCMRC manages falls outside the limits defined by the Central zone of the Arizona State Plane projection. However, the accuracy implications of this are less important than the consistency that is gained by combining all spatial data sets within the same projection.

4.2.2 Data Formats

Currently spatial data are stored at GCMRC either in ARC/INFO, or in image formats on local file systems. Supported formats have been standardized for the sake of compatibility and manageability. At a minimum, any spatial technology used in the DMF must be compatible with these formats:

Data type	File Format
Vector data	ARC/INFO coverages and shape files, ARC/INFO export format files (point, line, polygon and annotation), ArcSDE export format.
Imagery	TIFF, MrSID
Tables	CSV
Reports	MS Word 7.0; PDF
Metadata	XML

Certain large-scale geographic coverages will be moved into ArcSDE format, with Oracle as the relational database engine. The use of SDE with Oracle removes the limitations of INFO, allows for a multiple-user spatial database of very large capacity, and provides for a tighter integration of tabular and spatial data. However, the introduction of SDE does change the way in which GIS data are stored, retrieved, and viewed, and its use cannot be allowed to limit potential analyses or displays of spatial information. Many of the broad implications of using ArcSDE are unknown at this time; therefore GCMRC staff will carefully evaluate its use during the Pilot phase of development.

Initial tests have shown that it is feasible to store Canyon wide image mosaics within Oracle using ArcSDE technology. Certain efficiencies are gained by doing so; centralized data management and backup and faster retrieval and display times are two. It is anticipated that some legacy imagery may continue to reside in file system based storage, particularly photography that is accessed infrequently. All data sets, spatial, non-spatial, ArcSDE, and file system based, will be registered through ArcCatalog, thus making them accessible to Intranet users of ARC/INFO software tools. In addition, tabular data sets so registered will

be accessible to the variety of database connectivity tools described in the Functional Requirements section.

4.2.3 Software Standards

GCMRC uses the ESRI suite of GIS software included in ArcGIS version 8.2. Most external users of GCMRC GIS data also use ESRI software. The DMF must be compatible with and support the use of this software for viewing and spatial analysis.

One of the goals of the DMF is to take advantage of current technology in order to improve the management of the spatial data sets maintained by GCMRC. The Oracle Spatial Option is software technology that can be used extend the functionality of ArcSDE, though it is not required to build a Geodatabase with ArcSDE. Specifically, Oracle Spatial allows GIS data to be stored in a way that permits direct query of spatial attributes using standard (non-ESRI) SQL clients. This could be of benefit to a custom Web application, as a mechanism to perform spatial queries without using an ESRI API to the database. It may also provide more flexibility in managing spatial data stored in the Oracle system, by not requiring ESRI tools for routine administrative tasks. There is a tradeoff, however; the Spatial Option requires an Enterprise Edition of the Oracle license, which is significantly more expensive than a Standard Edition in a Web enabled publicly accessible version of the software.

The pros and cons of using Oracle Spatial will be carefully investigated during the Pilot phase of this project.

4.3 DATA FILE ARCHIVE

GCMRC manages a repository containing a large number of GIS data sets, collected by many different researchers and projects over a number of years. Just as with non-spatial legacy data, these have varying levels of accuracy and standardization, and also form a valuable archive of legacy data. Importing many of these individual data sets into a uniform GIS is of questionable value, and it is likely that they will continue to be stored in file-based formats.

The current contents of this Spatial Data File Archive are detailed in the document entitled "Legacy Data Inventory".

A proposed directory structure for archiving Electronic data received by the GCMRC is detailed in Appendix II. Legacy spatial and non-spatial data sets will eventually be moved into the Archive, indexed, and made available to query by the Data Registry application. The DMF will provide this mechanism for populating the Archive with both legacy and newly collected data, and creating a searchable index of data sets stored within it. Once this mechanism is in place, the tasks can be assigned to temporary or student employees for completion.

5. SECURITY REQUIREMENTS

5.1 GCMRC GUIDELINES FOR RELEASE OF DATA SETS

A data set consists of GIS, tabular and meta-data that are collected during, or result from a single research or monitoring project at the GCMRC (Grand Canyon Monitoring and Research Center). Data sets managed by the GCMRC are classed and released according to the following guidelines.

5.1.1 Provisional Data

All data sets delivered to the GCMRC are considered provisional until they have been checked for errors and incorporated into the GCMRC database. Certain data associated with core research projects will be considered provisional until the project has been finalized. These time frames are generally outlined by the individual project contract. The purpose of this policy is to allow researchers the opportunity to adequately analyze their data, and to allow time for peer review and publication of their results in professional journals.

Provisional data will only be available to the Public by specific written request to GCMRC, and with permission of the Principal Investigator who provided it.

5.1.2 Publicly Available

Most data sets generated by GCMRC-sponsored projects ultimately become Publicly Available, after a period of time with Provisional status.

The time frame for Provisional Monitoring data to become Publicly Available is generally two months from collection, however some data sets require more or less effort than others to be prepared. The time frame for data associated with special research projects to become Publicly Available is one year from date of delivery to the GCMRC, unless a longer time-period is specified by the individual contract.

Those data sets considered Publicly Available would be accessible through the GCMRC Web site using the DMF with standard browser technology. These include long-term monitoring data associated with core research areas as well as meta-data documentation for all GCMRC data sets.

5.1.3 Restricted

Some data are restricted from public access. The lead PI or data contributor will document these restrictions, which will be available through the GCMRC Web site in the meta-data for each project. Restricted data generally include, but are not limited to the following types of information.

- Data related to Cultural Resources
- Location information about endangered species
- Data generated by the research of others, who are not funded by the GCMRC, and who have contributed information conditionally.
- Data that are subject to copyright restrictions

Restricted data sets are available only to the Principal Investigator or Provider of the information. Metadata descriptions of these data are considered Publicly Available.

5.2 MEDIA AND FORMAT FOR RELEASE OF GCMRC DATA

5.2.1 Internet

All Publicly Available data sets and accompanying metadata will be made available for query and download via the DMF application running on the GCMRC web site. Archived data sets will be indexed in the GCMRC data catalog. This Internet data presentation tool will constitute satisfaction of any publication requirement for data sets and metadata.

5.2.2 Removable Media

Certain data sets may be compiled and released on CD-ROM, or similar removable media. This will be done on a case-by-case basis, depending on funding, need, and applicability of the medium to dissemination of data.

5.3 USER REGISTRATION

Public users of GCMRC data will not be required to complete a registration form or to "sign in" each time they access publicly available data from the GCMRC website. When the user enters the data view/download application, the GCMRC Guidelines for acceptable use will be displayed, and his/her acceptance of those guidelines will be required in order to move forward. This acceptance will be implemented merely to remind the user that the intellectual rights of those who generated the data are respected. A free user account will be created to avoid having to fill out any registration information for the general public.

Specific user groups, defined in Chapter 2.1, who need access to Provisional or Restricted data sets will be required to log in to the DMF system. Oracle Security will control access to data stored within the RDBMS, and NT Security will control access to data sets stored within the file based Archive. This security model is detailed in Chapters 5.4 and 7.4.

5.4 DATA ACCESS CONTROL

The following table outlines access restrictions that will apply to identifiable groups of DMF users. The system must also enforce appropriate security for the process of adding data to the Archive and RDBMS, and for editing data once it has been imported into the RDBMS.

It is envisioned that GCMRC IT staff will be responsible for importing new data sets to the system, and routine editing tasks required for maintenance of the system. However, it is expected that some Principal Investigators will retain the rights to edit and manage data that they provide to the system.

User Group	Privileges
General public	Read access to all data that is of Publicly Available status. No access to specific location information about Endangered Species. Read access only to the metadata related to archaeological studies. There will be no access to any real data, results, or location information of archaeological studies. The General Public will have full read access to any recreational studies.
AMWG/TWG	Same as general public.
GCMRC staff	Full read access to all data that is of Provisional, as well as Publicly Available status.

	Full read access to all Restricted information regarding Endangered Species. Read access only to the metadata related to Restricted archaeological studies. There will be no access granted to any real data, results, or location information of archaeological studies.
GCMRC Cultural staff	Same rights as GCMRC staff. Full read access to all Restricted culturally sensitive data.
GC Natl. Park, GC and LM Recr. Areas	Full read/write access to all, including Restricted culturally sensitive studies performed within their respective boundaries.
Individual Native American Tribes	Full read/write access to their respective data, including Restricted culturally sensitive information. No access to Restricted information belonging to other tribe. There may be groups of users within each tribe, such as administrators and general users, who have differing access rights. Specific user Roles will be determined after a general security model can be demonstrated to the group of tribal participants.
Principal Investigators	Full read/write access to studies that they have conducted, including Restricted Culturally Sensitive information. Read access to Provisional data of others.
GCMRC Database Administrator	Full read/write access to all classes of data stored in the database.

6. HARDWARE AND SOFTWARE REQUIREMENTS

GCMRC has an existing system infrastructure that supports file, database and Internet serving. The DMF will integrate with the existing infrastructure wherever possible. However, GCMRC does not intend that the design of the DMF be limited by existing system hardware and software.

6.1 EXISTING INFRASTRUCTURE

6.1.1 Server Hardware

GCMRC has a large-capacity system that is used for file and database serving, as well as the Arc License Manager process (system name: UNKAR). UNKAR is a Compaq rack-mount system, outfitted with dual-processor CPU and a disk farm with 2 terabytes of capacity. This system has excess CPU capacity for its current function, and will be used to initially host the DMF with no changes to its configuration. As time progresses, and data sets are added to the DMF, disk capacity of this machine will be added as required. This computer will house the Production Instance of the ArcSDE/Oracle database. In addition, a Development Instance of this database will be housed on a smaller workstation class computer (GCMRC-DBS). This development server allows experimentation with database configurations, ArcSDE and Oracle installation options without affecting a production version of the database.

GCMRC also has a computer used for serving static Web pages and hosting an ftp site.

6.1.2 Server Software

Server O/S	Windows 2000 Server
Web / FTP Server	Microsoft IIS 5.0
Database	Oracle 9.0.1.3 Enterprise Edition
Internet Map Server	ESRI ArcIMS 4.0

	Tomcat 3.2.3 Servlet Container
GIS	ESRI ArcSDE 8.2

6.1.3 Network Configuration

The GCMRC is running a Microsoft Windows 2000 Server, TCP/IP network.

Windows 2000 Security Groups provide Network Authentication, and access to network file and print resources is controlled by ACL (Access Control Lists) security.

6.1.4 Firewall

The USGS runs a Cisco router that provides extensive and flexible firewall capability to the GCMRC subnet. The architecture of the DMF must provide as many resources as possible inside the firewall, in order to prevent unauthorized or malicious access by external users.

6.1.4.1 VPN (Virtual Private Network)

USGS/GCMRC intends to provide VPN access to its Intranet. VPN access will be used during the development phase to provide access to GCMRC resource behind the firewall. As well, VPN access can be used as a mechanism for providing secure extranet access to the DMF for user groups who are physically located outside of the GCMRC Intranet.

6.2 CLIENT SOFTWARE REQUIREMENTS

Client O/S	Windows 2000 Professional, Windows XP Professional
Web Browser	Microsoft Internet Explorer 5.0 (IE5). There may be a requirement to support Macintosh users, via the support of IE5 (Internet Explorer) for Macintosh. Netscape support is not required.
Office Suite	Microsoft Office 2000 Professional
GIS	ArcInfo 8.2 with 10 floating licenses. Standard GCMRC machine will have ArcInfo installed, with floating licenses shared among users.
DBMS Client	SQLNet 9.0.1.1.0, SQLPlus 9.0.1.1.1, Oracle Discoverer 3.1.36
Statistical Software	SAS 6.12
DBMS Management	Embarcadero DBArtisan 5.4.1, RapidSLQ 5.7.1 Oracle Enterprise Manager 9.0.2.0.0
DBMS Design	Embarcadero ERStudio 5.0, Oracle Designer/Developer 6.5.52.1

7. HIGH LEVEL DESIGN

General architectural elements proposed for the DMF system are defined in this section. A detailed and refined design will result from work completed during the Pilot phase of this project. See *Design Priorities*.

7.1 SYSTEM ARCHITECTURE

The following diagram illustrates the system architecture anticipated for the DMF. Deployment of servers, major software components of the system, and varieties of client access pathways to the DMF are diagrammed. Each component is specified in greater detail in the following sections.

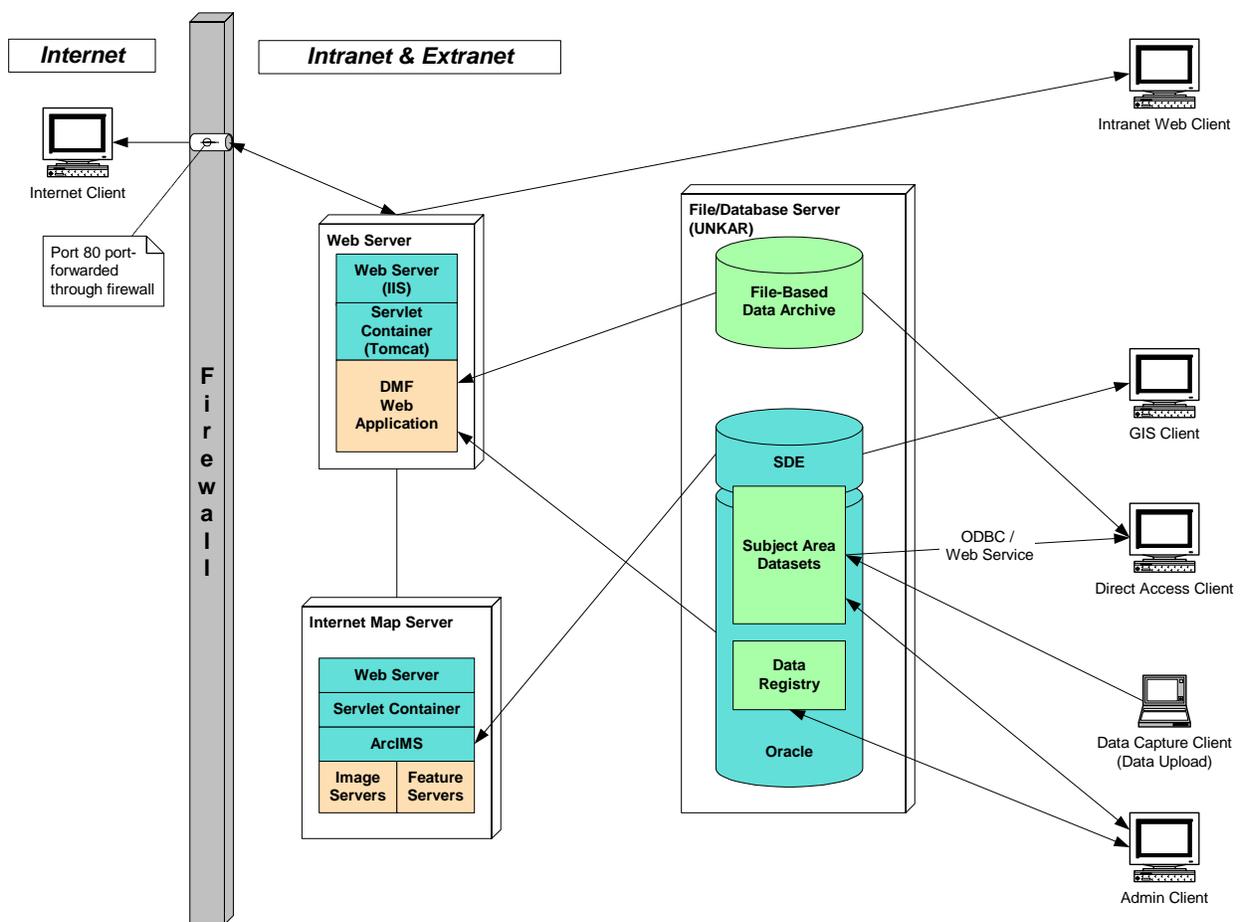


Figure 7-1 – System Architecture Diagram

7.2 SERVER DEPLOYMENT

The follow sections detail the specifications for servers necessary to support the DMF.

7.2.1 File And Database Server

7.2.1.1 Software Specifications

File sharing: Windows 2000

Attribute Database Oracle 9.0.1.1.1 Enterprise Edition with Spatial Data Option.
Future use of Enterprise Edition will be based on the evaluation of Oracle Spatial Option, and other technical issues investigated during the pilot phase.

Spatial Database ESRI ArcSDE 8.2

7.2.1.2 Hardware Specifications

The existing UNKAR server, currently hosting file system and database services has sufficient hardware capacity to continue in this role. Following completion of the Pilot phase of database development, the disk capacity requirements of UNKAR will be re-evaluated.

7.2.2 Web Application/Internet Map Server

A Web Application/IMS Server delivers static HTML pages, as well as dynamically generated HTML to both Intranet and Internet users. It will also run middle tier code for the DMF Web applications, and an Internet Map Server process that will render maps to be presented to users through the DMF Web application. These are moderately CPU intensive tasks, but require little disk capacity or I/O resources. These server processes may be separated on independent systems if testing shows that performance is an issue.

The Web Application and IMS servers are depicted as two independent systems in Figure 7-1, and are located inside the Firewall for security purposes. Because the DMF Web application must retrieve information from both Windows 2000 file system and Oracle database located on UNKAR, this configuration limits security issues that would arise if these tasks were consolidated on the GCMRC Web server. The GCMRC Web server will remain outside of the GCMRC firewall, and will exchange HTML packets with the Web Application/IMS Server through the firewall.

7.2.2.1 Software Specifications

Web Server IIS 5.0 or Apache
Servlet Container Tomcat 4.0
Internet Map Server ESRI ArcIMS 3.1

7.2.2.2 Hardware Specifications

CPU 1.5 GHz Pentium III, single processor
Disk 3X72 GB RAID 5

7.3 FUNCTIONAL ARCHITECTURE

The following diagram illustrates the relationship of the major functional components of the DMF. The following sections detail the design of each of these components.

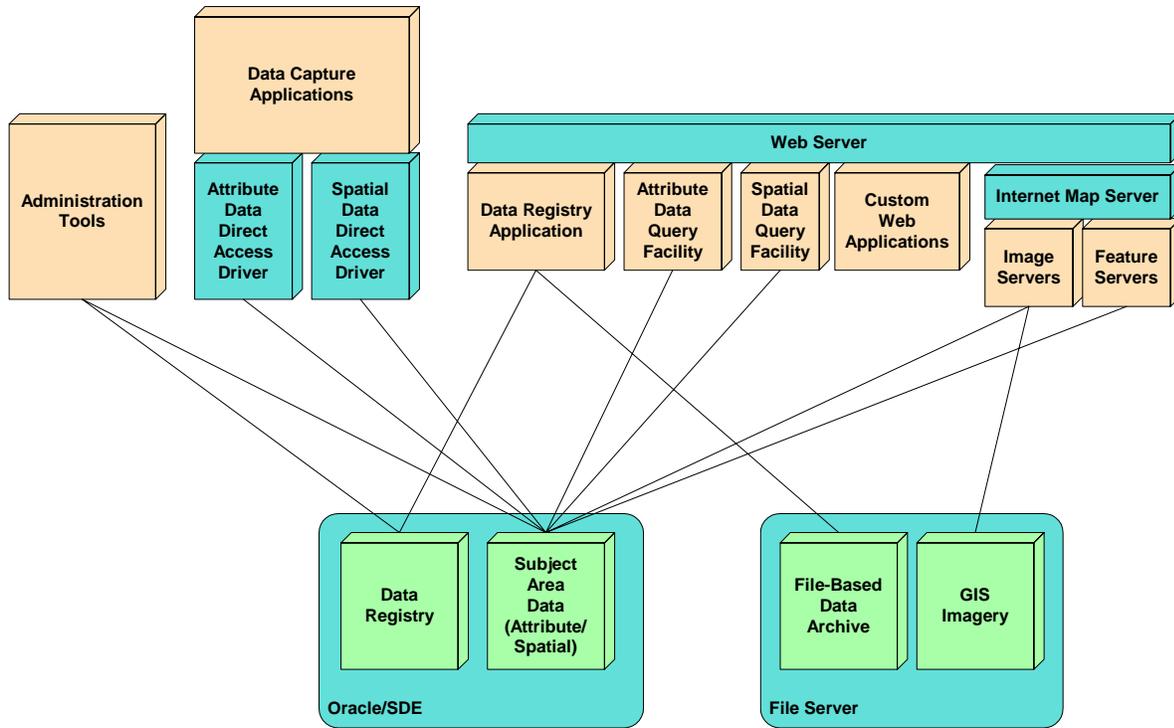


Figure 7-2 - Functional Diagram of the Data Management Framework

7.3.1 Data Registry

The Data Registry is a database and administration application that contains information about all of the data available to the DMF. This is the mechanism that DMF will use to provide extensibility to the system once the project is complete. For example, Web applications will query the data registry to find out what legacy data are available for download, and then present the user with the option to download or view them. As new data sets are created they will be registered in this application, and then be available for discovery by various Web applications and publishing components.

The Data Registry allows for the communication of the availability, delivery, and use of data within the GCMRC. Data collections (sets) are referred to in this Registry as Information Sources. These sources will include consolidated Oracle databases, legacy data files, hardcopy documents, Mylar maps, physical collection of items, etc.

Information Sources exist in the following formats:

Category	Existing Formats
Document	Html, WordPerfect, RTF, PDF, MS Word, ASCII, text, Other
Data	MS Access, XML, Oracle Table or View, CSV, Dbase, Excel, Lotus 123, Systat, Other

GIS	DEM, TIN, MrSID, PCARC/INFO, ESRI SDE Geodatabase, ARC Export, ARC Shapefile, ARC/INFO, ArcStorm, Other
Image	TIFF, Jpeg, gif, bitmap, png, Other
Hardcopy	Periodical, Library item, Journal, Photo, Video, Slide, Other
Other	http, ftp, MS PowerPoint, Other

For consistency, when legacy data sets are added to the Data Registry (i.e. indexed) they will be converted to the supported formats that are required of newly collected data, as described in the GCMRC Data Standards.

The Data Registry can be compared to a Library Card Catalog. The catalog provides a list of all Information Sources that are available, and Meta data describing each. Query screen(s) within the Data Registry web application will provide users a list Information Sources of interest, and allow for view, download, or query as desired.

The preliminary data model for the Data Registry can be found in Appendix I.

7.3.1.1 Data Categorization

Data will be categorized according to Subjects and Areas as they are defined in Chapter 4.1. These categories will provide attributes upon which data sets can be retrieved by searching the Data Registry.

7.3.1.2 Index Map

The DMF will provide an index map for the Data Registry, to in order to allow the user to spatially locate data sets. This will be accomplished by using ArcIMS "Services" to deliver both reference and scenery layers to the user of a Web browser, along with the capacity to turn each of them on or off. The scenery layers will not be query-able. Reference layers, however, representing ways that data sets have been organized or indexed, could be used to query the Data Registry for information. For example, a user might wish to display all of the studies that have been conducted within a certain range of river miles, or studies that target a specific species within a Monitoring Reach.

Spatial indexes will be displayed in the standard reference systems used by GCMRC, with an extent including Lake Powell, Upper Lake Meade, and tributaries of the Colorado River. These will be a sub set of the layers stored within the ArcSDE Geodatabase. The capability to add new layers from the ArcSDE Geodatabase to the ArcIMS "Service" is required functionality.

Reference Layers

- USGS Quarter-quadrangle boundaries
- Colorado River and tributary centerline network, including river mile index
- GIS Study Site boundaries
- New Long Term Monitoring Sites polygons
- Study specific sample Stations, e.g. Fish Polygons, Kearsley Transects, NAU Beaches
- GCMRC and NGS survey control points
- Index layers for base data sets that are not organized by Quarter quad, e.g. LIDAR flightlines, raw imagery flightlines, etc.

Scenery Layers

- Elevation relief map
- Primary, secondary and primitive roads
- Annotation

Cities

7.3.1.3 Data Registry Web Application

A Data Registry Web Application will provide widespread and detailed Public access to GCMRC data sets. This application will be written to compliment the ArcIMS map services listed in the previous section, by allowing users to drill down into attribute and tabular data selected by spatially querying the Index maps. Most GCMRC data are somehow spatially referenced, though much exists in non-spatially enabled tables. The Web Application interface must support both interactive map viewing and query of ArcSDE layers, and attribute query of non-spatially enabled tables within the database.

An attribute query interface will be integrated with the ArcIMS interface, to allow the user to select records from various data sets by their non-spatial attributes. From this interface, users will be able to perform ad-hoc queries of individual data sets, or combinations of data sets based on related attribute fields.

Attribute Query

The user should be able to query the Data Registry for Information Sources based on a number of predetermined attributes. These attributes include the following:

- Area /Subject
- Agreement Number
- Study Title
- Principal Investigator
- Trip ID
- Date Range
- River Mile (Station_ID and Station_type)
- Physical or Biological Parameter
- Keywords
- Species
- Ad-hoc Spatial Query
- Other Study Specific Attributes

The results of this type of query will be a list of Information Sources that match the query attributes.

Information Source List

The Information Source List will display a list of the query hits found in the Attribute Query form. Each Information Source will display the following information:

- Name
- Description
- Agreement Number
- Information Source Type
- Note
- Posted Date
- Contact
- Location by River Mile (Station_ID and Station_type)
- Metadata

A hyperlink to the Information Source itself may also be retrieved, if appropriate. This hyperlink would be specific to the Information Source Type; that is, it will provide a mechanism to query and display non-structured information such as subjective passages, photographs, etc, or to enter a set of ad hoc query screens that allow retrieval of information from the Oracle database. Some examples of hyperlinks include:

- A link to a document on the ftp site
- A link to a MS Word file stored in the Archive on UNKAR, such as a Final Report for a specific Contract or Agreement. Clicking this link would cause the Word file to download and display, if supported by the user's browser.
- A link to another part of the web application supporting generic display and query of Oracle database tables. If the Information Source is an Oracle table, this link would allow us to query, display, and export individual records from the Oracle table, based on the completion of an ad-hoc query screen
- A link to a related Website. This would allow the Information Source to point to a peer-reviewed journal containing a paper written based on research done for the GCMRC. This provides a way for GCMRC to store links to pertinent information without violating individual copyrights.
- A link to an image file or other non-structured data stored on the file server, allowing that file to be displayed on the user's browser.

7.3.2 Subject Area Database

The Subject Area database contains all of the data sets that have been identified in the Data Inventory that have been consolidated into a set of related Oracle tables. The Subject Area database will be implemented within the same Oracle database as the Data Registry.

The Subject Area database is modelled in the Entity Relationship Diagrams described in Appendix I. This model consists of a set of sub models, one for each type of data set that has been identified to date. Each sub model contains one or more entities, with the appropriate measured and derived attributes. High-level data models have been completed for most Subject Area data sets, though many are incomplete. The purpose of completing a high level design of all data sets is to look for common attributes and similarities between what initially appear to be disparate data sets. During the Physical Design phase of this development these ER diagrams are refined, Entities are simplified and combined, specific attributes are defined, and relationships between Entities are established.

7.3.2.1 Common Attribute Data

Each sample, measurement or observation collected by GCMRC has certain common attributes types associated with it. These attributes must be captured in order to adequately group and store the information so that it can be retrieved when needed.

Some of these common attributes are:

Agreement

A number used to identify the contracting organization, time frame, and cost accounting of a particular scientific effort at GCMRC. This attribute will be used to group deliverables with the Data Archive. It may also be used to retrieve a list of all deliverables received by GCMRC under a specific contractual agreement. The Agreement is used to identify the Principal Investigator responsible for a given data set.

Study	A GCMRC classification for an on-going data collection effort. GCMRC may enter into many contracts or agreements during the course of a particular study. A study is usually focused on a particular Subject or Area, but some may be multidisciplinary.
Trip	A unique sampling or data collection trip. A trip is usually a collection of sampling events. A single deployment of a continuous sampling monitor also constitutes a trip. An Agreement usually results in one or more sampling trips, perhaps over the course of several years.
Location – Station, Site, River Mile	A unique place at which individual samples are taken or measurements made. Data sets are geo-referenced. Depending on the type of measurement or sample, this reference might be to a repeatable well-known sampling location such as a gage station, or a spatial reference framework such as river mile. Either constitutes a sampling “Station” that can be spatially referenced on a map.
Station Type	A grouping of Stations by the kind of sampling activity performed there. Examples include Gage Stations, Fish Sampling Stations, Vegetation Transects, Survey Control Point, etc.
Sample	A sample is collection of material taken at a station, for analysis either on site, or at a laboratory off site. Grouping observations or measurements into samples is often important to their analysis. For example, the numbers and sizes of fish caught in individual sampling “efforts”.
Sample or Measurement Type	A grouping of samples or measurements collected using a certain methodology or technique. Examples include Fish Netting vs. Electro Fishing Samples. Attributes of the sampling effort vary by the type of sample taken.
Observation	An observation is an activity conducted in the field that yields results without the use of instruments, usually relying on the judgment of the observer.
Measurement	An activity conducted in the field, usually involving the use of instruments of some sort, which produce environmental results without the creation of samples to be evaluated elsewhere.
Date and Time	The date and time at which a Sample, Observation or measurement is taken. Sometimes the duration of the sample is an important parameter that must be recorded.
Parameter	A Parameter is a physical attribute that is measured in the field or laboratory. Examples include Species, Length, Temperature, Ion Concentration, etc.
Status, Sensitivity and Security Levels	The observation, measurement, or sample taken must be assigned both a status (Provisional, Available, Restricted) and a security level. Oracle 9i row level security can be employed to limit access

to sensitive database records, without the use of complex application programming logic.

7.3.2.2 Spatial Data

Spatial database technology allows spatial attributes (X, Y, and Z coordinates) to be stored in the same table as non-spatial attributes. For compatibility with the GCMRC ESRI GIS software, ArcSDE will be used to spatially enable the Oracle database. For entities with spatial attributes, such as "Stations", the physical tables will be SDE enabled using ESRI software. This will allow relationships to be built between these and tables of observations and measurements performed at those locations. In this configuration care must be exercised to preserve the integrity of the ArcSDE spatial schema while using non-ESRI tools, such as SQLPlus, to manage the database.

ArcSDE can be implemented on Oracle using one or more of three different schema options. One of these requires the use of the Oracle Spatial Option; two do not, making the SDE installation portable to RDBMS software products other than Oracle. In the following list of options, the perceived benefits and drawbacks of each are discussed. For this discussion, consider the "Stations" table to be an example of a "Business" table to which spatial attributes are to be added.

1. **ArcSDE compressed binary with a LONG_RAW data type.** The ArcSDE geometry is stored in a column defined as a LONG_RAW data type, in a feature table separate from its corresponding business table. When spatial geometry is added to the business table, a foreign key reference is made to the feature table relating the two. This is similar to how feature geometry was stored in the old INFO database technology. The coordinates of a feature are only visible to ESRI client tools.

Advantages:

- Does not require Oracle Spatial or the Enterprise Edition of Oracle (less expensive software license).
- Storage is efficient. Features are stored in 4 byte binary integer pairs with a "scale" multiplier that determines the level of precision.
- Display and spatial queries using ArcMap and ArcCatalog may be faster because fewer bytes are transferred to the client.

Disadvantages:

- Cannot query the coordinates of a feature using non-ESRI query tools such as SQLPlus.
- Support for the LONG RAW datatype, though it is still the default for SDE, has been dropped by Oracle.

2. **ArcSDE compressed binary with a BLOB data type.** Same as the first, except that the data type used is a BLOB (Binary Large Object).

Advantages:

- Allows Oracle replication of database.
- More flexible data type than LONG RAW.

Disadvantages:

- Unknown.
3. **Oracle Spatial Object Relational Model.** Oracle Spatial allows features to be defined with the SDO_GEOMETRY data type, and spatial columns are stored directly in the business tables, not in separate feature tables (F tables).

Advantages:

- Conforms to the OpenGIS Simple Features Specification for SQL.
- Coordinate values are numeric, and can be queried using SQL statements.
- Oracle provides other spatial display and analysis tools, such as “draw”, “buffer” and “overlay” that replace the simpler functions available in ArcGIS.
- SDO geometry is portable to other GIS software vendors.

Disadvantages:

- Requires the use of Oracle Spatial option to the Enterprise Edition (more expensive software license).
- It is a more complicated SDE schema.
- Feature coordinates require more storage space than SDE Binary type.
- Performance may be slower than SDE Binary for a large database.

The most important limitation of the default ArcSDE schema is that its use is practically limited to ESRI client tools. In particular, ArcSDE does not provide any spatial query functionality that is available from a standard SQL client. Oracle Spatial is a technology that does provide SQL spatial query functions, though it comes at a cost of a larger, more complicated, and more expensive database schema. The application design of the DMF might be simplified by using Oracle Spatial in conjunction with ArcSDE, however, because spatial attributes of features could be queried directly with custom Java tools instead of requiring the use of ArcGIS APIs.

Spatial database development will initially proceed using the ArcSDE Compressed Binary LONG RAW data type, unless specific data sets or application programs require the use of Oracle Spatial. The use of ArcSDE Compressed Binary BLOB data types will be investigated, due to the fact that Oracle support for the LONG RAW data type is questionable. GCMRC staff will make these design decisions based on experimentation during the pilot project.

7.3.3 File-Based Data Archive

The DMF must provide a warehouse for file-based data sets. GCMRC has defined a directory structure for legacy file data that will be tested during the pilot project, to ensure it provides needed functionality and flexibility within the DMF.

7.3.4 Attribute Data Query Facility

The DMF must provide, through the web application, the ability to query and display Oracle tables or views containing attribute data. This capability will allow the user to enter common query or filter conditions, such as Date Range, Species, or Study. Query results will be displayed as column heading and data in a browser window, and allow users to refine or change their query. The user can also export query results, for download and use in offline analyses.

7.3.4.1 Support offline analysis of data sets.

Users of GCMRC data sets must be able to analyze data outside of the system using their own software tools. These tools include spreadsheets, custom modeling applications,

statistical packages, or GIS tools such as ArcView or ArcInfo. For instance, if a user needs to perform population estimates on fish species using a tool such as Excel, he or she must be able to export some subset of data from the Oracle database, to the user's desktop.

Common export file formats, such as CSV (Comma-Separated-Value) allow exporting data in form readily usable by other applications.

Supported export types are

Information Source Type	Export File Format
Oracle Database Table or View	CSV (Comma Separate Value)

Note that this is an export-only mechanism for tabular data. The DMF Web application will not support uploaded changes made to these data sets.

7.3.5 Spatial Data Query Facility

7.3.5.1 Image Servers

The DMF will utilize the capability of ESRI ArcIMS to define Image Server Map Services to render maps of spatial data sets. These maps will be available through a Web-based interface. The interface will allow basic GIS viewer functionality such as

- zooming and panning
- layer display control
- legend generation
- scale-dependent layer display
- feature Identify

The maps will also be available by using other ESRI client tools to directly connecting to the ArcIMS Image Server Map Services.

7.3.5.2 Feature Servers

ArcIMS has the capability of defining Feature Server Map Services. A Feature Server Map Service allows ESRI client software to request extracts of feature-based data sets from spatial data sources. Using Feature Server capability to provide data extracts from DMF data sets will be investigated as part of this project. Security and performance issues will be identified during this investigation.

7.3.5.3 Spatial Query Application

The DMF will provide the ability to view and query spatially referenced subject area data, independently of IMS. It will display base layers for spatial orientation, and allow a user to select additional layers to be displayed. The user will be able to select a feature and display all of its available attributes.

Data sets that should be available (but are not necessarily limited to) include the following:

- Survey
 - Control Points
- Water
 - Riverbank edge at selected discharge levels
 - Stream flow Gage sites
 - Lake Powell Water Quality monitoring sites

- Physical
 - Geomorphology
 - Surveyed Sand bars
 - Channel Geometry Profiles
 - Geomorphologic River Reaches
- Biology
 - Surveyed patches for species counts
 - Composition of plant and animal species
- Biology – Aquatic
 - Fish Sampling “Stations”
 - Shoreline fish habitat types
 - USFWS fish habitat transects (LCR)
- Biology – Vegetation
 - Dominant vegetation types
 - Vegetation sample plots
 - Vegetation sample transects
- Biology – Avian
 - Bird patches
 - Bird point count locations
 - Bird walking transects
- Cultural
 - Campsites

7.3.6 Custom Web-Based Applications

The DMF will be designed to provide maximum flexibility and extensibility; however, the specialized requirement of a given discipline might require a custom web application be developed. For example, a fishery biologist might need to display all recaptures of a given PIT Tag, in order to examine the growth history of a certain fish. While it is possible to provide this functionality through the generic Attribute Query, the general public most likely won't need it. A better solution might be to provide this kind of query through a custom Web application that is only available to internal users.

7.3.7 Direct Query Facility

A single generic Web application cannot hope to satisfy all query requirements of all user groups. For this reason, we need the ability to provide direct access to particular data sets that might be contained in the DMF.

7.3.7.1 Access via ArcGIS tools.

More sophisticated users of the GCMRC data sets will require the enhanced spatial analysis capabilities of other GIS tools. This system will not prevent a privileged Intranet user from analyzing GIS data with ArcGIS tools.

7.3.7.2 Access to Oracle data via SQLNet/ODBC (Read/Write)

Users on the GCMRC Intranet may need to query or maintain Oracle tables using software tools such as MS Access, Visual Basic, or SAS. To provide this kind of access, standard client / server technologies such as SQL-Net and Oracle ODBC connectors must be supported. These technologies will impose a maintenance issue for GCMRC System Administration staff, but the number of users requiring this kind of access should be small.

7.3.7.3 Programmatic Access to Oracle Data via a Web Service

A User of the DMF may require programmatic read-only access to Oracle tables thru web-enabled applications. An example of this is the Flow Modeling Application that requires stage-discharge rates for a particular date/time range at a given location. Programmatic access would allow the modeling software to establish an Internet connection to the Web server, and using an established query mechanism, invoke a function whose XML output would be usable to the modeling program. The term for systems that support requesting information via a Web interface is "Web Service". Typically a web service allows client software to establish an Internet connection to the Web Server and make a request for data. These data are delivered in XML format. The requesting software should be able to understand data in this format.

The DMF should support the ability to act as a web service. The Web Service should support a set of queries similar to the ones supported by the Data Registry web application.

7.3.8 Data Capture Applications

Some data capture processes could be automated using of custom applications running on small form factor computers such as the Compac iPAC or the Toshiba laptop. These data capture applications would be used for record data electronically in the field, and allowing information to be directly transferred to the Oracle database when back in the office. Some initial candidates for field data capture applications include:

- Control point GPS site logs
- Fish samples
- Water quality observations

7.3.9 Administration Application

An Administration application is needed to aid GCMRC Administration personnel in the maintenance of various shared aspects of the DMF. These aspects include:

- Contact and PI information
- Contracts and Agreements
- Trip information

7.3.10 Data and System Administration Tools

Data and System Administration tools will be developed to handle the following tasks:

- Adding Information Sources to the Data registry.
- Administering Data Registry Security for User Groups.
- Verifying links between information sources and linked resources.
- Automating repetitive data management tasks such as backup and recovery.
- Maintaining domain tables in the Oracle database
- Importing data sets into the Oracle database.

7.4 SECURITY ARCHITECTURE

7.4.1 Web Applications

Because the DMF Web Application will be deployed to the Internet, a large component of the technical design of this system must deal with Security. Security entails the following:

- Restricting users with Internet access from accessing areas of the GCMRC network that they have no right to.
- Designing a hardware and software architecture that will work within the confines of the GCMRC/USGS firewall strategy.
- Restricting authenticated users to those areas of the web application that they have rights to.
- Restricting user only to Information Sources they should have access to. If the Information Source is an Oracle table, the user may only have access to specific records within that Oracle table.
- Support for multiple groups of users.

The DMF Web application must distinguish between users of the system, and grant them appropriate access privileges. Restrictions focus on Cultural, Endangered Species, and Provisional Information Sources.

The Security requirements for the web application are as follows:

- When appropriate, the application must have an authentication mechanism to determine the identity of the user. The use of NT Groups could simplify and centralize user management by utilizing the security features of the underlying operating system. Oracle accounts could be set up to utilize OS password authentication, eliminating the need for two sets of user passwords, one to access file based systems and one to access database tables.
- Support for secure access using authentication will be provided to users of Microsoft Internet Explorer 5.0 or newer browsers on the Windows platform.
- Support for secure access using authentication will be provided to users of Microsoft Internet Explorer 5.0 or newer browsers on the Apple Macintosh.
- Support for secure access using authentication will possibly be provided to users of Netscape or other browsers on Windows or other operating systems.

Given the security requirements of the system, one solution is to provide security at the application level, meaning that access to the both Oracle tables and file resources would be controlled by the web application itself, simplifying administration of the file system, and meaning that Oracle UserID's would not be required for all users. The web application would log into Oracle with a single secure Oracle UserID with sufficient privileges to access all necessary resources, and would broker this resource access to individual user depending on their group membership.

In this type of system, the web application must support different access rights to individual Information Sources, based on the assignment of a Source to any number of Security Groups, the name of which will coincide with NT Groups.

7.4.2 Data Administration Tools

A small number of users, possibly 1 or 2, will fall into the Data Administrator User group. This group, comprised of GCMRC staff members, will need the highest level of DMF access, in order to perform administration task on the DMF data sets.

Applications that require or use direct access to Oracle thru SQLNet and ODBC will utilize Oracle UserID/password's and Oracle roles to provide security.

Oracle Roles allow the database designer to grant specific access rights to a role. These access rights allow a user who has been granted that Oracle role to perform operations within those rights, such as Select, Update, Delete, Insert, Execute, etc. If a user hasn't been granted a role, that user will not have privileges with respect to a schema of which he is not the owner.

7.4.3 Data Capture Applications

Security requirements of individual Data Capture applications will not be known until each are analyzed in detail.

7.5 DESIGN PRIORITIES

A three-phase development effort is proposed for this Project. The planned phases are:

Phase I	Pilot
Phase II	Production
Phase III	Ongoing Deployment of Data sets

Phase I - Pilot

The Pilot Phase allows for an architectural design and development using a series of iterative steps. Feedback from GCMRC staff and stakeholders will be integrated before work proceeds on remaining data sets, in order to assure that long term needs are met. For the Pilot Project, three representative data sets have been selected to produce an initial database design, including Fishery, Water Quality and Stream Flow/Sediment areas. Other data sets will be examined, and GCMRC staff interviewed, in order to insure that the initial database design will be extendable to remaining data sets. Specifically, the Pilot data sets include information that is described in the following sections of this document:

- Native and Exotic Fish – 4.1.6.1
- Gage Information – 4.1.1.1
- Suspended Sediment at USGS Gages – 4.1.1.2
- Historical Suspended Sediment at USGS Gages – 4.1.1.3
- Dam Release Data (SCADA) – 4.1.3.2
- Stream Flow Gage Data – 4.1.3.4
- Stage-Discharge Relationships– 4.1.3.11
- Chemical/Physical/Biological Water Quality – 4.1.3.12.2

The goal of the Pilot Project will be to produce the following products and features:

- Index map website and map service
- Sample data areas – Fishery, Water Quality & Stream flow gages
- Data Registry – partial population of DR tables
- File Structure for the file data
- Definition of Oracle tables, views, tablespaces, etc
- Detailed definition of H/W, S/W, and network architecture
- Definition of ArcSDE layers
- Data load/ Data Admin tools / applications
- Orthophoto Index

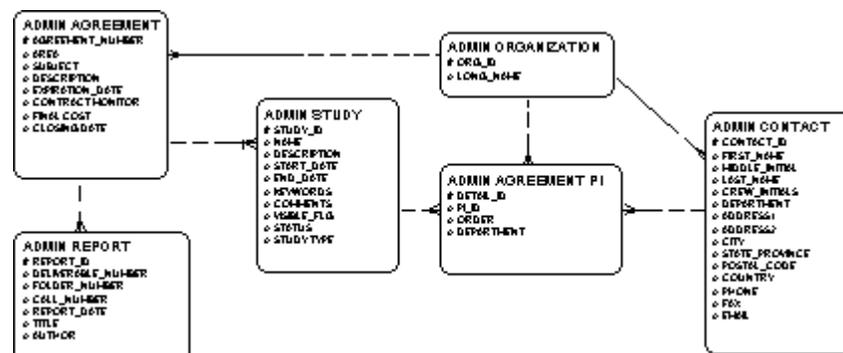
Phase II – Production

On completion of the Pilot Project, GCMRC staff will evaluate results. Phase II of the development effort will seek to put the application into production. This will involve making architectural changes based on user feedback and technical knowledge learned during the Pilot Phase. Phase II will finalize any code developed during the Pilot into a production quality application on the GCMRC intranet. At the end this phase, pilot data sets will be fully accessible using Web and GIS tools, and specific analysis needs of each user group, identified during the Pilot, will be met.

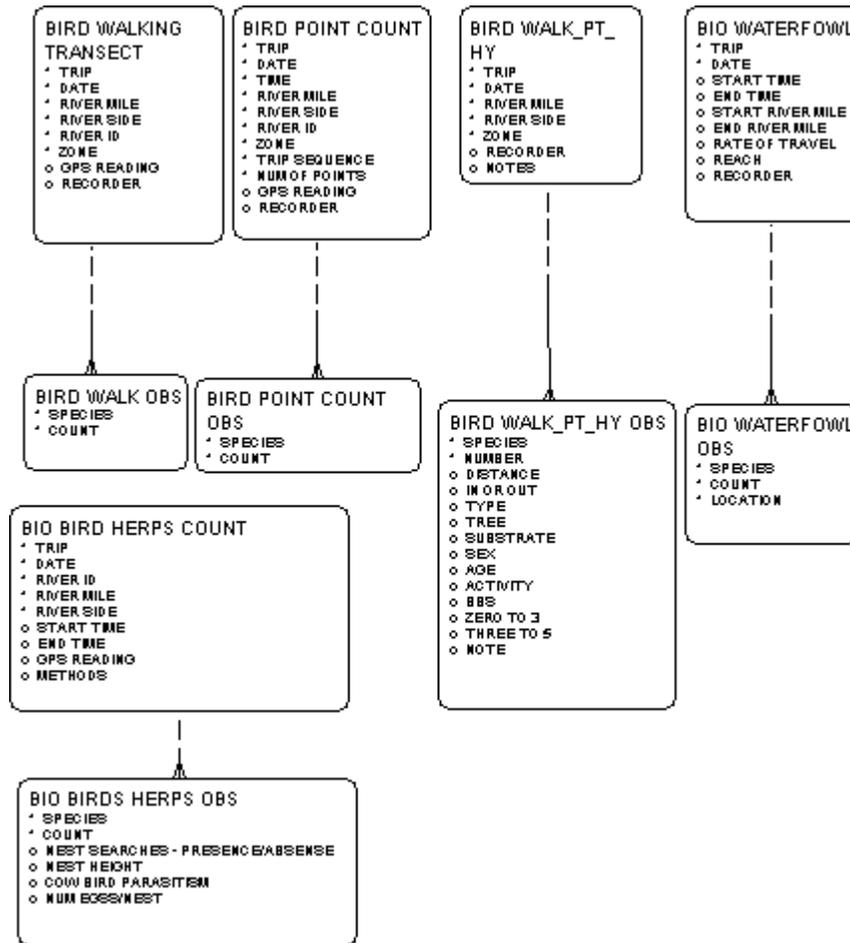
Phase III – Ongoing Deployment of Data sets

Phase III of the development effort will add remaining data sets to the database application, and will build on the product developed during Phase II. Phase III deployment will be driven by priorities defined by GCMRC Program Managers. Once a data set has been analyzed to ensure consistency and value, it will be added to the Data Management Framework.

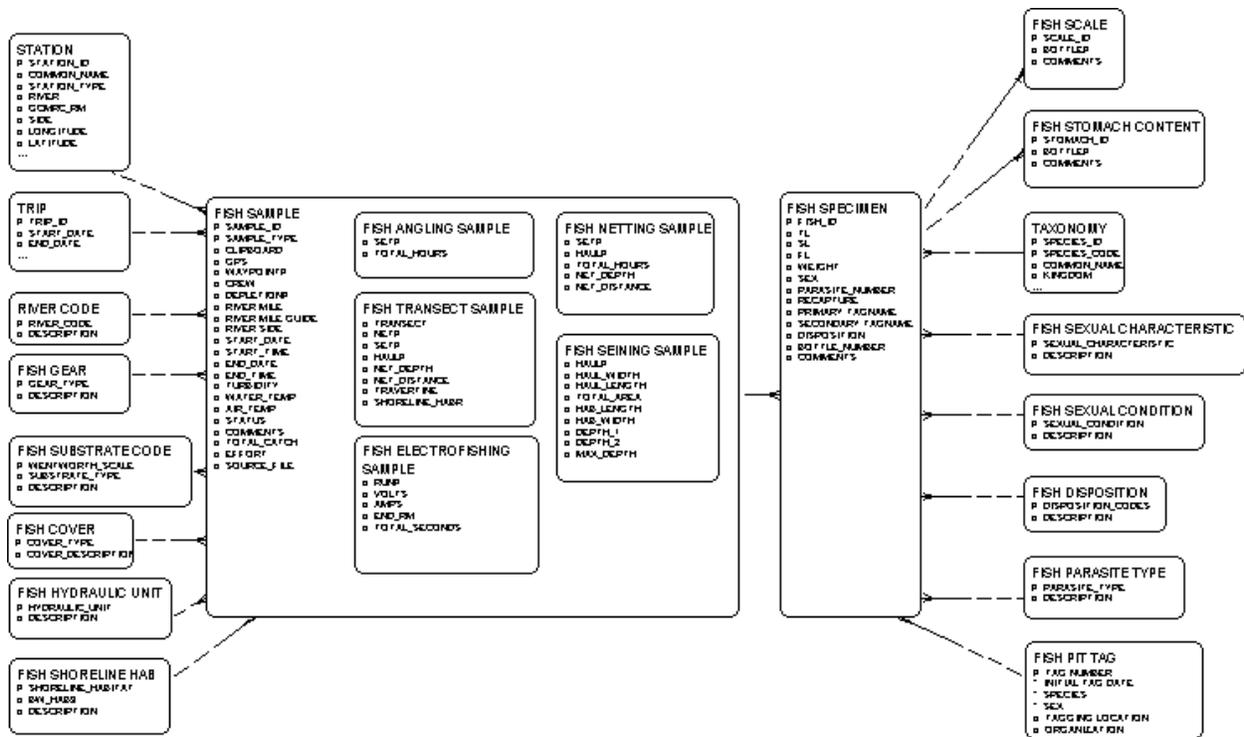
8. APPENDIX I - ENTITY RELATIONSHIP DIAGRAMS (ERD)



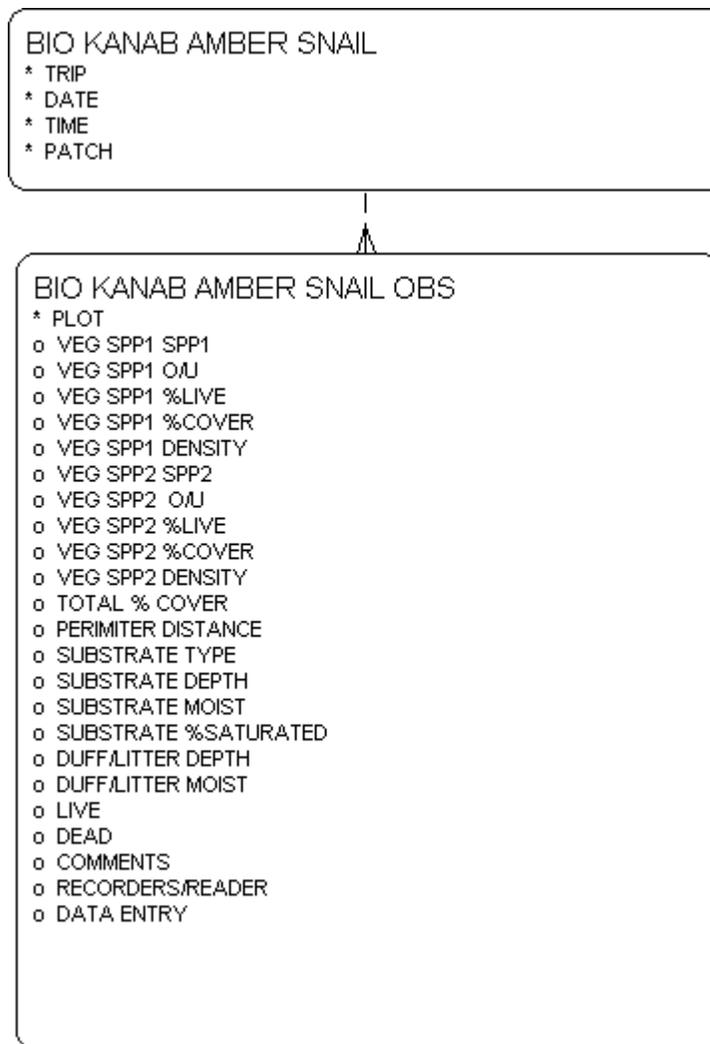
ERD ADMIN

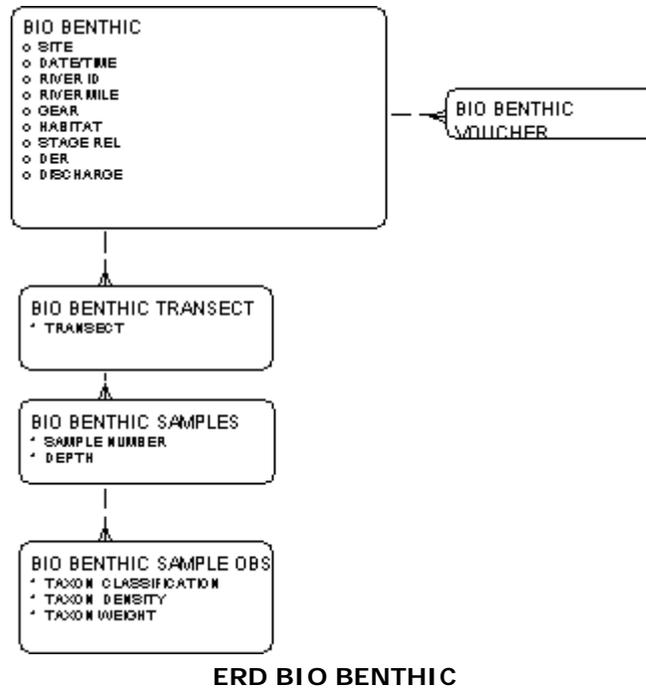


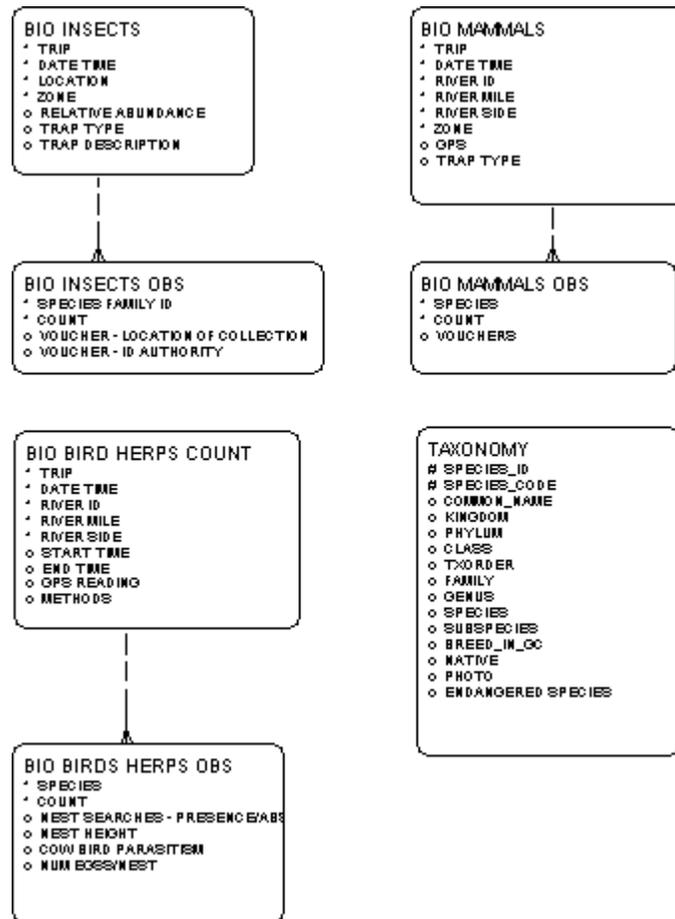
ERD BIO BIRDS



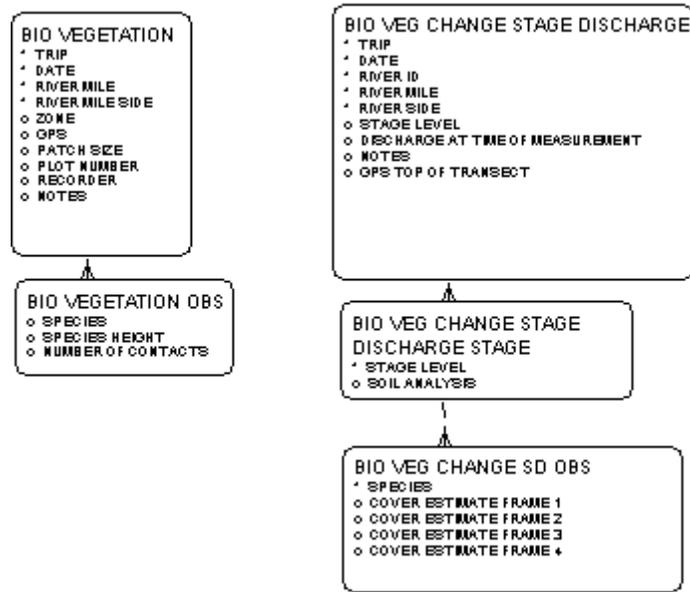
ERD BIO FISH

**ERD BIO KANAB AMBER SNAIL**

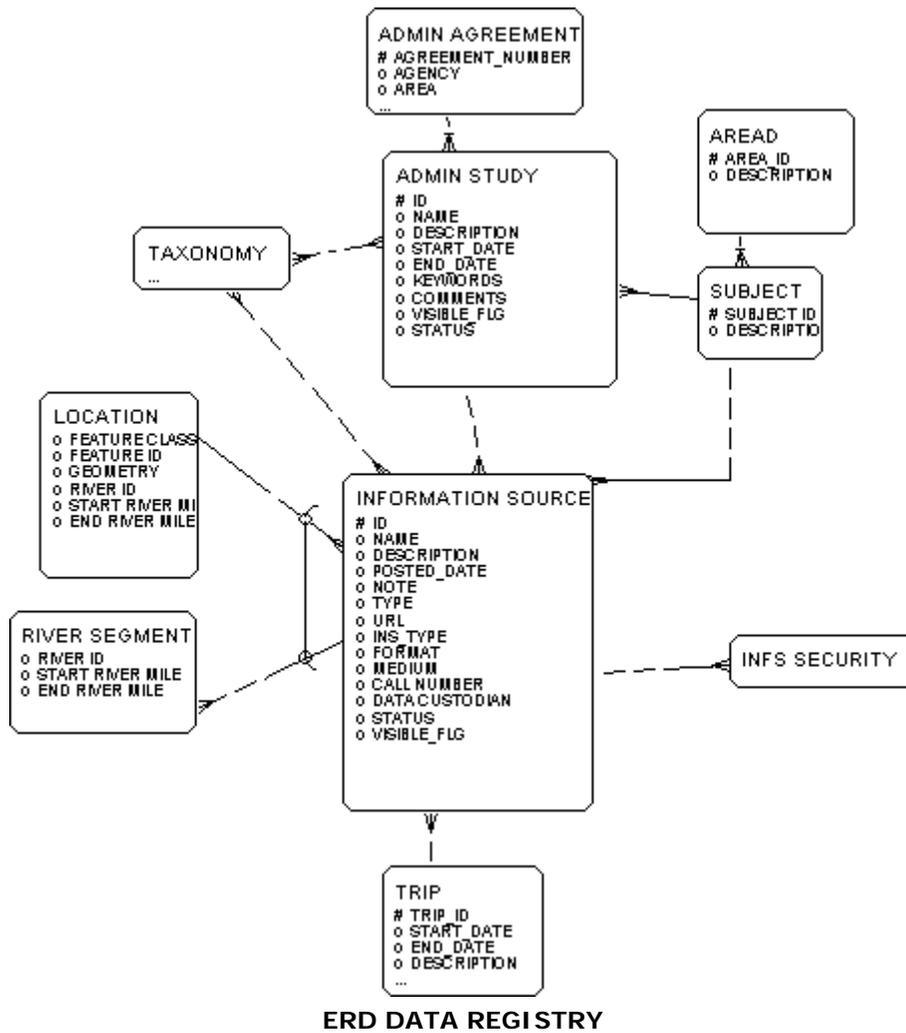




ERD BIO GENERAL



ERD BIO VEG



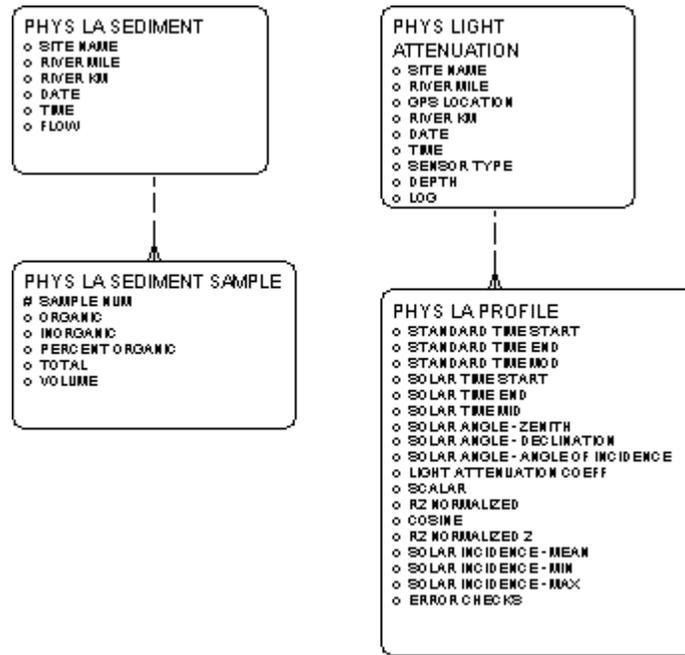
LIB FOLLET CATALOG

- o CALL NUMBER
- o TITLE
- o AUTHOR
- o PUBLISHED
- o DESCRIPTION
- o NOTES
- o ELECTRONIC ACCESS
- o ADDED ENTRY

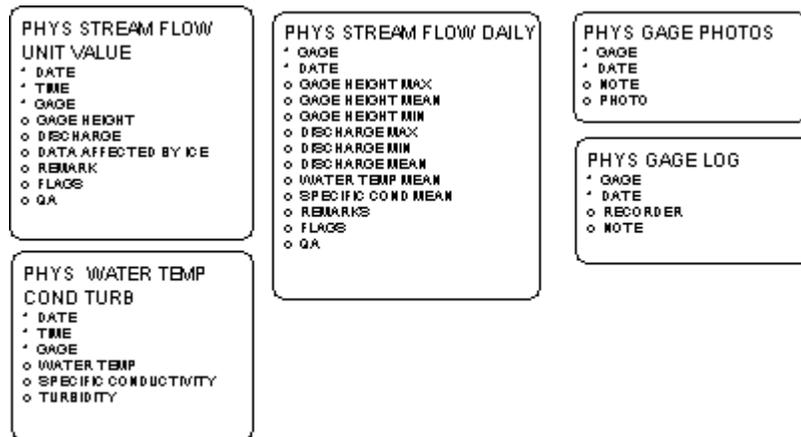
LIB IMAGERY

- o ACQUISITION DATE
- o TYPE
- o PRODUCT
- o START RIVER MILE
- o END RIVER MILE
- o SCALE OR RESOLUTION
- o NUMBER OF IMAGES OR FILES
- o DAM RELEASE CFS
- o COMMENTS

ERD LIBRARY



ERD PHYS LIGHT ATTENUATION



ERD PHYS STREAMFLOW

PHYS RAIN
GAGE
* STATION NAME
* DATE
* TIME
* INCHES OF RAIN

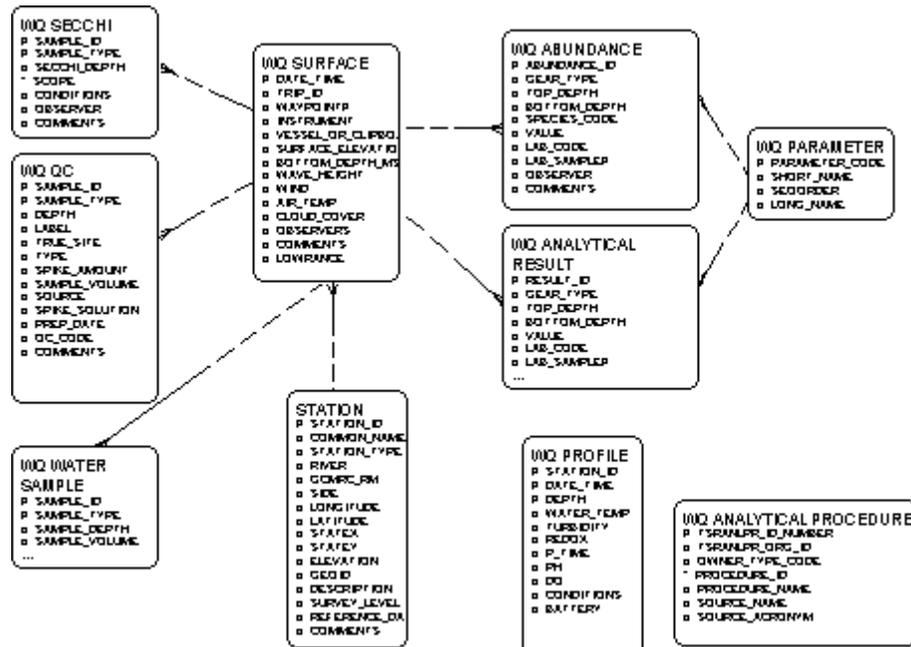
PHYS SANDBAR
* SITE ID
* TRIP ID
* SANDBAR VOLUME
* EDDY VOLUME
* STAGE DISCHARGE

PHYS SEDIMENT
* SAMPLE ID
* STATION NUMBER
* TRIP
* RIVER MILE
* GPS
* NOTES
* SAMPLING METHOD
* DATE
* TIME_MST
* START TIME
* END TIME
* MEAN TIME
* AIR TEMP
* WATER TEMP
* FLOW DEPTH AT SAMPLE LOCATIO
* SAMPLING_TIME
* FLOW RATE
* SILT & CLAY CONCENTRATION
* SAND CONCENTRATION
* SAND D160000
* SAND D500000
* SAND D840000
* SAND % < 0.075000
* SAND % < 0.075000
* SAND % < 0.105000
* SAND % < 0.125000
* SAND % < 0.149000
* SAND % < 0.177000
* SAND % < 0.210000
* SAND % < 0.250000
* SAND % < 0.297000
* SAND % < 0.354000
* SAND % < 0.425000
* SAND % < 0.500000
* SAND % < 0.595000
* SAND % < 0.707000
* SAND % < 0.841000
* SAND % < 1.0000
* SAMPLE WGT
* WGT PASSING THRU 63 MICRON SIEVE
* WGT RETAINED ON 63 MICRON SIEVE
* CONCENTRATION < 63 MICRON SIEVE
* CONCENTRATION > 63 MICRON SIEVE
* CONCENTRATION ANALYSIS LABAT
* GRAIN SIZE LABORATORY

ERD PHYSICAL

CONTROL
* POINT NUMBER
* VERSION
o END DATE
o NORTHING
o EASTING
o WGS-84 ELLIPSOID HT
o NGVD29 GEOID90 ORTHOMETRIC HT
o NGVD29 GEOID99 ORTHOMETRIC HT
o NAVD88 GEOID90 ORTHOMETRIC HT
o NAVD88 GEOID99 ORTHOMETRIC HT
o NVGD29 - NAVD88
o GEOID90
o GEOID99
o N LAT DEC
o N LAT DMS
o E LONG DEC
o E LONG DMS
o W LONG DMS
o DESCRIPTION
o IMAGE

ERD SURVEY



ERD WATER QUALITY

9. APPENDIX II - ARCHIVE DIRECTORY STRUCTURE

Archive Structure for Electronic data received by GCMRC (GCMRC Internal DRAFT)

This structure is internal and will be viewed by Information Technology staff only. In the future other staff, as well as the public, will be able to query and access the archive using a web interface product.

P: This is where the archive is located on Unkar

Data for contracts will be stored by agreement number

/01WRAG0034

Each of these folders will include all files sent by the contractor: e.g. draft reports, final reports, and data.

/draft_20010513

Once a final has been received, peer reviewed, and approved by the program manger, the draft can be deleted from the archive

/final_20010817

/GC20000129

Grand Canyon Trip – departed January 29, 2000

/LP20010506

Lake Powell Trip – departed May 6, 2001

/LF20010310

Lee's Ferry Trip – departed March 10, 2001

/00_FC_40_3400

If a contract has several components or cooperators working on a single contract, the sub-project will be further denoted by the P.I.'s last name

/Topping

/raingagedata

/streamflow_unitvalues

/Rubin

/Schmidt

/98_AA_40_22620

/01_NA_WRD_001

When GCMRC receives reports from outside agencies that it hasn't funded, a contract number can be fabricated to manage the data. Outside users can access this report by author, agency, and date.

/basedata

The base data folder will include vector data and cartographic products produced in house such as the Center Line map, river mile coverage, etc. Imagery will be stored on the Q: drive

/Mar2000_lidar

/NE36111B7

/SW36112C4

These folders would contain xml files and compressed files of processed lidar points, deleted points, a digital elevation model (dem) and tin for the quarterquad

/legacy

This folder will include electronic data for which the contract number is not known. It will be organized by contractor. Once an agreement number is found, the data will be moved into the folder labeled "Agreement #"

/Biowest

/deliveryA19950423

This directory includes the first delivery that was submitted April 23, 1995 from Biowest

/sensitive

This folder will contain sensitive data including cultural and endangered species information; access is limited and protected.

Q: This area includes all of the imagery stored on Unkar

/Mar2000_bw_ortho

Black and white orthos for the March 2000 flight

/SID

This folder will contain sided images for each of the ortho'd images

/TIFF

This folder will contain tiff images for each of the orthophotos

/Sept2000_bw_raw

These are the black and white images for the September flights

/Preflood

All data from September preflood flight including tiff images, eo files, reports, and the footprint

/Postflood

All data from September postflood flight including tiff images, eo files, reports, and the footprint

/July2000_cir_raw**/footprint**

These images are too numerous and large to store on Unkar, therefore a footprint will be available from which individuals can determine which images they need and obtain them from the library.