98-DI
PP. 630-638

A NEW VISION

ASPRS-RTI

1998 ANNUAL CONFERENCE

MARCH 30 -
APRIL 3, 1998
Tampa Convention Center
Tampa, Florida

PROCEEDINGS

ASPRS – RTI
1998 Annual Conference
Technical Papers
March 30–April 3, 1998

About this CD-ROM

In order to enhance your ability to use **Technical Papers – ASPRS – RTI 1998 Annual Conference** as a research tool, the book has been provided to you in electronic format. The book has been rendered in a "portable document format" (PDF) using Acrobat® software, a product of Adobe Corporation.

The Adobe Acrobat format will allow you to search the entire contents of the document for key words and will then display the pages containing those key words on your computer screen. Free versions of the Acrobat Reader® software have been included on this CD-ROM for Windows 3.1, Windows 95, Windows NT, UNIX, and Macintosh platforms. Versions of Acrobat® for other operating systems can be downloaded from the Adobe World Wide Web site. This software will allow you to perform simple searches based on key words, display the results of the search, or browse the document interactively.

If you desire a more sophisticated search tool, this CD-ROM is compatible with Adobe Acrobat Exchange®. Acrobat Exchange features a more sophisticated search tool that allows you to make use of word stemming, "sounds like," and complex search patterns when using this product as a research tool. Adobe Acrobat Exchange is available from your local computer software retailer or directly from Adobe Corporation. The Adobe Acrobat Exchange program is included in the **Adobe Acrobat® version 3.0** software package.

Adobe Acrobat Reader® Installation Instructions
Please read the terms of the Adobe® License Agreement before you proceed.

Windows 95 / Windows NT

Installation Instructions
Locate directory **ACROREAD\WIN\RDR_SRCH\32BIT**.
Execute program SETUP.EXE to install software. Follow instructions on your screen. View the file README.TXT for additional information.

Windows 3.1

Installation Instructions
Locate directory **ACROREAD\WIN\RDR_SRCH\16BIT**.
Execute program SETUP.EXE to install software. Follow instructions on your screen. View the file README.TXT for additional information.

Macintosh

Installation Instructions
Locate the folder **Reader : Reader**.
Double Click on the program **Install Acrobat Reader 3.0** to install the software. View the file **ReadMe-Reader** for additional information.

UNIX

Installation Instructions
Locate the directory /READER/UNIX/READER.
Locate the .TAR file that is appropriate for your version of UNIX. Install the software using the commands that are appropriate for your version of UNIX. View the file INSTGUID.TXT for additional information. Further information is available at the Adobe® WWW site. URL=http://www.adobe.com

NOTICE: Other Operating Systems including DOS and OS/2 Warp

Adobe Acrobat Reader® is available for other operating systems as well as those included on this disk. To obtain other versions of Adobe Acrobat Reader® you may set your Web browser to URL: **http://www.adobe.com** and download the version that you need. The complete list of supported operating systems can be found at URL: **http://www.adobe.com/prodindex/acrobat/alternate.html**

IMPORTANT INFORMATION

Adobe® Acrobat® Reader 3.0 Electronic End-User License Agreement

ADOBE SYSTEMS INCORPORATED ELECTRONIC END-USER LICENSE AGREEMENT FOR ADOBE ACROBAT READER

NOTICE TO USER: THIS IS A CONTRACT. BY INSTALLING THIS SOFTWARE, YOU ACCEPT ALL THE TERMS AND CONDITIONS OF THIS AGREEMENT.

This Adobe Systems Incorporated ("Adobe") End User License Agreement accompanies the Adobe Acrobat Reader product and related explanatory materials ("Software"). The term "Software" also shall include any upgrades, modified versions, or updates of the Software licensed to you by Adobe. Please read this Agreement carefully. At the end, you will be asked to accept this Agreement and continue to install or, if you do not wish to accept this Agreement, to decline this Agreement, in which case you will not be able to use the Software.

Upon your acceptance of this Agreement, Adobe grants to you a nonexclusive license to use the Software, provided that you agree to the following:

1. Use of the Software.
   - You may install the Software on a hard disk or other storage device; install and use the Software on a file server for use on a network for the purposes of (i) permanent installation onto hard disks or other storage devices or (ii) use of the Software over such network; and (iii) make backup copies of the Software.
   - You may make and distribute unlimited copies of the Software, including copies for commercial distribution, as long as each copy that you make and distribute contains this Agreement, the Acrobat Reader installer, and the same copyright and other proprietary notices pertaining to this Software that appear in the Software. If you download the Software from the Internet or similar on-line source, you must include the Adobe copyright notice for the Software with any on-line distribution and on any media you distribute that includes the Software.

2. Copyright and Trademark Rights. The Software is owned by Adobe and its suppliers, and its structure, organization, and code are the valuable trade secrets of Adobe and its suppliers. The Software also is protected by United States Copyright Law and International Treaty provisions. You may use trademarks only insofar as required to comply with Section 1 of this Agreement and to identify printed output produced by the Software, in accordance with accepted trademark practice, including identification of trademark owner's name. Such use of any trademark does not give you any rights of ownership in that trademark. Except as stated above, this Agreement does not grant you any intellectual property rights in the Software.

3. Restrictions. You agree not to modify, adapt, translate, reverse engineer, decompile, disassemble, or otherwise attempt to discover the source code of the Software. Although you may customize the installer for the Software to install additional files (such as plug-in and help files), you may not alter or modify in any way the installer program or create a new installer for the Software. The Software is licensed and distributed by Adobe for viewing, distributing, and sharing PDF files. Any plug-in or enhancement that permits you to save modifications to a PDF file is authorized only for use with Acrobat Exchange® program and may not be used with the Software.

4. No Warranty. The Software is being delivered to you AS IS and Adobe makes no warranty as to its use or performance. ADOBE AND ITS SUPPLIERS DO NOT AND CANNOT WARRANT THE PERFORMANCE OR RESULTS YOU MAY OBTAIN BY USING THE SOFTWARE OR DOCUMENTATION. ADOBE AND ITS SUPPLIERS MAKE NO WARRANTIES, EXPRESS OR IMPLIED, AS TO NONINFRINGEMENT OF THIRD-PARTY RIGHTS, MERCHANTABILITY, OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL ADOBE OR ITS SUPPLIERS BE LIABLE TO YOU FOR ANY CONSEQUENTIAL, INCIDENTAL, OR SPECIAL DAMAGES, INCLUDING ANY LOST PROFITS OR LOST SAVINGS, EVEN IF AN ADOBE REPRESENTATIVE HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES, OR FOR ANY CLAIM BY ANY THIRD PARTY. Some states or jurisdictions do not allow the exclusion or limitation of incidental, consequential, or special damages, or the exclusion of implied warranties or limitations on how long an implied warranty may last, so the above limitations may not apply to you.

5. Governing Law and General Provisions. This Agreement will be governed by the laws of the State of California, USA, excluding the application of its conflicts of law rules. This Agreement will not be governed by the United Nations Convention on Contracts for the International Sale of Goods, the application of which is expressly excluded. If any part of this Agreement is found void and unenforceable, it will not affect the validity of the balance of the Agreement, which shall remain valid and enforceable according to its terms. You agree that the Software will not be shipped, transferred, or exported into any country or used in any manner prohibited by the United States Export Administration Act or any other export laws, restrictions, or regulations. This Agreement shall automatically terminate upon failure by you to comply with its terms. This Agreement may only be modified in writing signed by an authorized officer of Adobe.

6. Notice to Government End Users. If this product is acquired under the terms of the following: GSA contract - Use, reproduction or disclosure is subject to the restrictions set forth in the applicable ADP Schedule contract; U.S. DoD contract - Use, duplication, or disclosure by the Government is subject to restrictions as set forth in subparagraph (c) (1) (ii) of 252.227-7013; Civilian agency contract - Use, reproduction, or disclosure is subject to 52.227-19 (a) through (d) and restrictions set forth in the accompanying end-user agreement. Unpublished rights reserved under the copyright laws of the United States.

Adobe Systems Incorporated, 345 Park Avenue, San Jose, CA 95110-2704 USA.

YOUR ACCEPTANCE OF THE FOREGOING AGREEMENT WAS INDICATED DURING INSTALLATION.

ISBN-1-57083-054-1

Published by
American Society for Photogrammetry and Remote Sensing
5410 Grosvenor Lane, Suite 210
Bethesda, Maryland 20814

Printed in the United States of America

pp 630 - 638

on CD-ROM

4747

98-DI

DESIGN AND IMPLEMENTATION OF THE ENP ENVIRONMENTAL DATABASE USING ORACLE*

N. Rishe, D. Barton, M. Chekmasov, M. Chekmasova, K. Madhyanapu, R. Rodriguez

High Performance Database Research Center, School of Computer Science, Florida International University, U.S.A.

ABSTRACT

The joint project of the Everglades National Park and Florida International University is aimed at constructing a set of databases which contains the environmental data that has been collected in the Park from 1911 to date. These data-sets describe trends in vegetation, land and marine fauna as well as the physical parameters of the area. The design stage of the project is based on the semantic binary model database approach which is being developed at the High Performance Database Research Center at FIU. The implementation stage consists of table creation, data loading, forms, user views and reports, as well as Internet access support using ORACLE RDBMS.

The completion of the project will make valuable environmental data accessible to the public through the World Wide Web and encourage new environmental investigation projects both within the Park and in the area encompassing South Florida. These investigations can include areas such as water management, fisheries and air quality control. Furthermore, this project would serve as an ideal ground for incorporating satellite and aircraft data with existing textual data to give a more comprehensive view of the development of South Florida.

---

* This research was supported in part by DoI (CA-5280-4-9044 of the National Park Service), NASA (under grants NAGW-4080, NAG5-5095, and NRA-97-MTPE-05), NSF (CDA-9711582, IRI-9409661, and HRD-9707076), ARO (DAAH04-96-1-0049 and DAAH04-96-1-0278), NATO (HTECH.LG 931449), and State of Florida.

INTRODUCTION

The Everglades National Park is currently considered one of the most environmentally sensitive regions in the United States. The problem of monitoring this vast territory is increasingly important. Environmental surveys have been performed in the Park for more than half a century, and the results of these surveys need to be stored and managed in an efficient manner that is also easy to use. To achieve this goal a joint project between Everglades National Park (ENP) and the High Performance Database Research Center (HPDRC) at Florida International University has been established.

The joint project is aimed at constructing a general environmental database for the Park. This general database is being implemented as set of 22 databases containing about 300 tables with approximately 2000 fields that will store and manage at least 1GB of plain text data. Each of the 22 databases stores information on a different type of survey that has been done in the Park. Further, since the surveys are continually under way in the Park, new data will be also placed into these databases.

Along with the improved use and management of this environmental data, the ongoing research contributes towards better understanding and documentation of historical data. It also will aid in future survey designs by allowing the opportunity to check the data sets and correct the errors in them. Finally, through the use of modern database technology, survey results can be made available to public through Internet. Thus, new environmental surveys will have a stronger research background and the data will be widely available for urban planning and educational purposes.

The future continuation of the present project is even more exciting. Plans are currently underway for the construction of a semantic database which will integrate the present textual data with satellite images of the territory. This will give a broader view of the development of South Florida.

The rest of the paper is organized as follows. Section 2 describes the main data sets to be populated in the databases. Section 3 corresponds to the semantic approach for the database design. Section 4 describes the implementation part of the project. Section 5 discusses the applications of the environmental database and considers further directions of project development.

THE DATA SETS AND DATABASES

In the scope of one paper it is impossible to describe all the data sets that will be implemented in the project. Therefore, in this section we will concentrate on the discussion of the largest data sets and the corresponding databases.

Environmental surveys typically contain some general information such as a description of the survey project, the names of investigators and technicians performing the work, and the instruments that are used. For storing this type of data a special Common database has been designed and implemented. There are several tables in this database worth mentioning.

The Project table assigns a code to the survey and describes the project. The Personnel table contains information about the people involved in the research. This includes the ENP employees (i.e., biologists, laboratory assistants and technicians) as well as personal information on contractors such as computer specialists or airplane pilots. Although each person can be uniquely identified by Social Security Number, the first, middle and last names are stored as well. The Organization table contains information on organizations of interest to ENP. It includes the names of organizations that order and perform the surveys as well as Contact-information.

The special Equipment table describes instruments and tools used within the survey. It includes serial and model numbers, accuracy, precision, resolution, and the manufacturer's name. Location is among the most important general tables. The results of the survey are highly dependent upon the position of the measuring instruments or the place for collecting data. We have agreed to use the Universal Transverse Mercator (UTM) system to store the location information. Nevertheless, the scientists sometimes use other coordinate systems for recording data. To deal with this, special triggers are used to implement algorithms that transform information from one coordinate system to the other. They are designed in such a way that the user need not be concerned with how the coordinates are stored in the database.

There are several look-up tables which contain helpful support information. For example, the Measurement-unit table contains the description of all possible units in which the data can be measured during investigations. These tables are not related to the particular data sets.

The Common database plays an important role in synchronizing the other databases. To illustrate this point we turn to the tables of part of the Common database called the Taxonomy sub-schema. The Taxonomy table contains a set of classification codes for animal and vegetation species. A composite taxonomy code includes information on Category, Family, Genus, Species and Subspecies for each particular species. This is accompanied by the common and scientific, mainly Latin, name of the species. Usually each survey has its own coding system for the species under investigation so one species may be represented by several codes, each belonging to different systems. Since the species code and the project code uniquely identify the particular species, we are able to turn to the Taxonomy tables to find out the classical, worldwide recognized name of the animal or plant. Thus, information about the object of interest can be received from different databases using the Common one.

Now we turn to the discussion of the Physical data sets. At present approximately 200 monitoring stations are found throughout the territory of the Park. These stations automatically record the values of the physical parameters at their locations on a daily, hourly and even 10-minute basis. Depending on the equipment installed, the following physical parameters are measured: water temperature, salinity, pH, depth, conductivity, air temperature, relative humidity, rainfall, evaporation, radiation, and wind speed and direction. As a result, a set of valuable and precise data is growing every day. This continuous investigation requires well-organized equipment maintenance and a good system of data storage.

The process of collecting data and maintaining the instruments was implemented in the Physical database. Here the concepts of Physical-observation-data-set and Physical-observation play a key role in the design. This implementation has resulted in the construction of corresponding tables which are, in turn, related to other tables of both the Physical and Common databases. In particular, Physical-observation-data-set is collected by implementing some analytical and sampling methods by using sensor equipment and intermediate storage devices on a particular location. Each has a person responsible for the data set. Physical-observation is related to the corresponding data set and is determined by the date-time and the recorded value. Depending on the physical parameter measured, some specific characteristics are also recorded.

The next data set to be discussed here is the Creel Census. This data set describes the fishing activities within the boundaries of the Park. Florida Bay is an essential part of the Everglades National Park and until 1989 commercial, recreational and guide fishing was allowed. Commercial fishing, however, was subsequently forbidden. The data is usually collected by interviewing fishermen and by counting, identifying and measuring the various fish species. The boating activities in this area are also estimated. The compilation of this information led to the construction of the Creel Census database. To formalize the data collection process we used Interview-report as a key concept. Each report is uniquely identified by a number and date. This makes it possible to tie the fish measurements with information on the location, personnel and equipment used. Moreover, the Park is issuing guide permits which are identified by unique numbers related to the date of issue. This provides the means to trace information about the activities of fishermen who are guiding others during boat trips.

Shrimp is another database that is related to the marine activities in the Park. The corresponding data set is collected at the marine observation stations. The recording of the physical parameters plays an important role in this part of the investigation. These parameters include surface water temperature, turbidity, conductivity, salinity, pH, a measure of the dissolved oxygen, wind speed and direction, cloud covering percentage and moon phase. Marine field observations are performed with the special methods and devices which are described in the database. Collection of a sample at a particular station and on a specific date is a major source of information for the recording. It includes invertebrate, fish and vegetation species. The quantity, weight, length, and gender of the species caught are all stored in the database.

In addition, marine vegetation observations are performed with special sampling devices and methods. The logic of the data collecting process was implemented in the database design. The logical schema of the Shrimp database consists of four sub-schemas that correspond to Marine Field Observations and Observation Methods as well as Invertebrate, Fish and Vegetation observations.

Data is populated in several other databases as well. The Panthers database stores the data received during the investigations of Florida panther behavior. The Manatee database contains information on manatee aerial observations. The SRF (Systematic Reconnaissance Flight) database stores data on white tail deer, wading birds and alligator nests observations. Separate databases have been created for Water and Soil Quality, Air Quality, Pine Demographic Study, Water Quality Effects on Vegetation, East Everglades Exotic Plants, Wildlife, Fire Ecology and other environmental surveys.

SEMANTIC APPROACH TO THE DATABASE DESIGN

Semantic design of a database is a construction of its conceptual schema. One of the key advantages of the semantic approach is that it highlights the functional logic of the constructed database. Another strength of the semantic database design is in its clear and straightforward appearance. It proves to be understandable by a person who is not familiar with the theory of databases and storing data in computers. In the present project it serves an intermediate language between the database designers and the biologists. The biologists, being specialists in the environmental surveys and the main users of the database, take an active part in the database design.

The central notion of semantic database design is the concept of an object. An object is any real world entity that we wish to store information about in the database. Objects are categorized into classes according to their common properties. These classes, called categories, need not be disjoint - that is, one object may belong to several of categories. Further, an arbitrary structure of subcategories and supercategories can be defined.

We assume that the representation of the objects in the computer is invisible to the user who perceives the objects as real-world entities. These objects can be tangible, such as persons or cars, or intangible, such as observations, meetings, or desires. At the design stage the database is perceived by its user as a set of facts about objects. These facts are of three types: facts stating that an object belongs to a category; facts stating that there is a relationship between objects; and facts relating objects to data, such as numbers, texts, dates, images, tabulated or analytical functions. The relationships between objects can be of arbitrary kinds. For example, stating that there is a many-to-many relation 'address' between the category of persons and texts means that one person may have an address, several addresses, or no address at all.

The semantic design can be implemented in either a semantic or relational database. In the present project we are constructing the relational database since the initial data to be populated is represented in plain text format with a simple structure. However, in situations where the structure of information is complex, greater flexibility is required (objects with unknown identifiers, or objects moving from one category to another, etc.), or non-conventional data is involved (long texts, images, etc.) implementation in the semantic database is preferable. We refer to [1] for a thorough discussion of the semantic modeling approach.

Let's turn to the example of semantic database design. Figure 1 represents a sub-schema of the semantic design for the SRF survey database. The categories

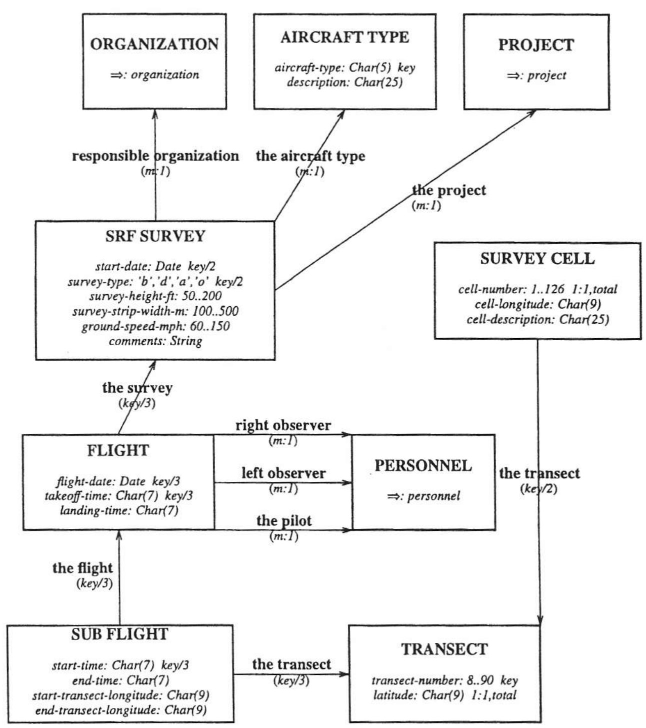

Figure 1: Semantic sub-schema for Systematic Reconnaissance Flight survey.

are drawn in the boxes. The name of a category is written in capital letters and the name of its attributes and ranges are written in lowercase italic. The relations between the categories are represented by the arrows. Each relation has its own name.

Briefly this schema can be explained as follows. The SRF-SURVEY is identified by its start-date and survey-type. The survey is performed within the PROJECT by the responsible ORGANIZATION. The survey consists of a set of FLIGHTs, performed on the aircraft of some particular AIRCRAFT-TYPE. The flight crew is represented by the pilot and two observers, who are described within the category PERSONNEL. The flight, in turn, is a set of SUBFLIGHTs. Each subflight is performed along some particular TRANSECT, which has a unique number and is determined by latitude. The transect also determines a set of SURVEY-CELLs. The location of each particular cell is uniquely identified by the cell-longitude and the latitude of the corresponding transect. Thus, the cells determine a grid which covers the territory of the Park.

IMPLEMENTATION OF THE DESIGN USING ORACLE

We want to emphasize that a construction of a correct semantic design of the database is a key task in the whole development process. The design should completely represent the ongoing survey and should be flexible enough for adopting any changes which might affect the performance of future surveys of this type.

The implementation of the semantic design in the tables of the relational database is now an automatic process since the relational database model can be considered a subset of the semantic model. The semantic model is less strict and much more flexible. Available programming tools, created at the High Performance Database Research Center (HPDRC), are used to fully automate the process. These tools take a description of the semantic database schema as input and creates the corresponding relational database including basic forms and reports.

With the semantic design of the databases approved we used the above mentioned conversion tools for creating databases in ORACLE RDBMS. We also used the basic forms and reports for creating custom forms and reports.

Loading data is the next important step in the development process. It is possible to load the data into the ORACLE database in two ways. The first way is manual data loading using the created forms. This approach, however, is almost useless due to the massive amount of data that needs to be populated into the database. Thus, the forms will mainly be used for making corrections in the loaded records and performing queries to the database. A more practical way for our purposes is data loading by means of the ORACLE SQL*Loader tool. It allows an ORACLE user to automate the process by specifying the data file format and the tables to be populated with the data.

As was mentioned previously, we dealt with a lot of historical data, some of which was collected before the computer era and was not properly documented. The SQL*Loader allowed us to find the errors and misprints in this data. While loading data records the loader tool checks all the integrity and checking constraints and rejects the records which do not fit into the database. An essential part of the loading process is a joint analysis of the rejected records. The database designers and biologists need to integrate their skills to find ways to correct and populate these records to the database. This is particularly important since the historical data is very valuable in that there is no feasible means to return back in time and to re-collect this data. Comparison of the historical and modern data gives a broad view of how the rapidly growing human activities in South Florida affect nature.

The forms provide a user-friendly interface for making queries to the database. Together with the forms, the user views are very helpful in retrieving data from the database. Generally, a user view is a virtual table which contains the data from several physical tables and at the same time provides the user with the results of the desired query. We designed and implemented the user views to calculate the elementary statistics on the data retrieved.

In this project there is one more reason for the design of the user views and forms. Historically, different groups of scientists used various units of measurement in the surveys. As a result the data collected includes numbers using different scales. The data should be stored in a unique system of units, however, the user should be able to retrieve data in the convenient form. For example, suppose a person used to work with a latitude/longitude coordinate system, but the data is stored in UTM coordinates in the database. The form retrieves the data from the database and automatically transforms one coordinate system to the other in such as way that the user is not concerned with what format this data is actually stored in the database.

As was mentioned above, the ENP environmental database is physically a set of databases. Nevertheless the databases are closely related to each other. This structure makes the implementation part of the project more complicated in comparison with the case of one database. Nevertheless it has certain advantages from the database administration point of view. First, each particular database is relatively small. Its structure is clear and easily documented. Second, the general performance is increased since the user is mainly interested in one subject at a time. Third, insuring security is a less complicated task since it is evident that only certain users will have rights to change the content of the database.

APPLICATION AND FURTHER DEVELOPMENT OF THE PROJECT

The completion of construction of the ENP environmental database will achieve several goals, and some of the benefits have already been obtained. In particular, the valuable historical data is professionally stored, managed and documented. The biologists are already starting to use the data for their research work and the reports on environmental problems. They are also using the results of the previous surveys to plan further investigations. It is well-known that environmental surveys are expensive and proper planning of the research is essential to reduce cost of the study.

Unfortunately, at present the database is used only by professionals and is not available to public. All the supporting tools are designed to help biologists in their research. At the same time specialists in other fields are interested in this data. The data may help in decision making on water management, urban and road construction, fisheries and air quality control. The data may be used for educational purposes as well.

The most natural way to share the ENP environmental database is to allow Internet access to it. The ORACLE Web Server package could be a supporting tool for implementation. Of course, additional work should be done on the design of a user interface for performing the database queries since, at this level, it would be non-professionals interested in the data.

The present project is also an ideal ground for the future research at the HPDRC. One of the aims at the HPDRC is the construction of a prototype of the semantic environmental database which, in addition to the field observations already obtained, will accumulate the available aerial survey, satellite remote sensing, as well as demographic and cartographic data. The main goal is to integrate this data in such a way that a comprehensive database of all information concerning this important, environmentally critical area is available in a convenient, useful, and timely manner.

AKNOWLEDGEMENTS

We would like to thank Bob Doren, Dave Buker, Darrell Tidwell and other scientists of the Daniel Beard Research Center, Everglades National Park for their cooperation during our joint work on the project.

SELECTED BIBLIOGRAPHY

Rishe, N. (1992). *Database Design: The Semantic Modeling Approach.* McGraw-Hill. 528 p.