DEVELOPMENT OF A DATABASE FOR THE COMPREHENSIVE STUDY OF THE A.C.F./A.C.T. RIVER BASINS

James Talian[1] and Tom Littlepage[2]

AUTHORS: [1]Water Resources Engineer, Sverdrup Civil Inc., Suite 310, 3354 Perimeter Hill Drive, Nashville, Tennessee 37211; [2]Environmental Engineer, Alabama Dept. of Economic and Community Affairs, Office of Water Resources, 401 Adams Avenue, Montgomery, Alabama 36103.
REFERENCE: Proceedings of the 1995 Georgia Water Resources Conference, held April 11 and 12, 1995, at The University of Georgia, Kathryn J. Hatcher, Editor, Carl Vinson Institute of Government, The University of Georgia, Athens, Georgia.

Abstract. The development of the comprehensive study database has addressed many different topics since the effort began. These have included software selection, software development, format standards, database structures, dissemination media and general coordination. This paper addresses a sampling of the types of issues confronted and discusses the relationship between evolving technology and a multi-faceted effort such as the comprehensive study.

INTRODUCTION

The Comprehensive Study of the Apalachicola-Chattahoochee-Flint and Alabama-Coosa-Tallapoosa river systems will produce many volumes of reports and a myriad of tabular data bases. In order to simplify distribution and increase the usability of this large amount of information, CD-ROM technology will be used to supplement traditional report distribution techniques.

One or several compact disks will be produced incorporating data compiled by the Comprehensive Study. The disks will include text, spreadsheets, databases and maps and related geographic information system data in a variety of formats. A common indexing and file access system will be developed to direct users to the documents which meet their individual needs.

Background

As we are aware, the methodologies for disseminating information in our society are drastically changing, and the magnitude of information is dramatically increasing. Our ability to keep up with these changes is so taxed that we inevitably find ourselves "playing catch up" with the technology. Since water resources in general is a data intensive field, practitioners are caught up in the information technologies explosion.

Related Research

Much of the technical work on database development is being done outside of the water resources arena. In many cases, we are merely using information technologies developed in the computer software business and applying them to our discipline. The computer technology developers know more about their field than we do, but they do not know how to use that technology in water resources applications without input from water resources experts.

EXPERIMENTAL DESIGN

Software alternatives addressed in the development of this database effort include: time series software, GIS software, database software and multimedia software. The application of each of these to the field of water resources could be the topic of a separate paper.

Selecting standards for these various types of data has had an interesting range of results. In one case there was almost universal agreement; in other cases there was, and remains, a wide variety of conflicting preferences; in still other cases there was almost nothing to chose from.

Time Series

The time series data has been of considerable focus. This type of data, while by no means unique to water resources, is a major consideration of water resources planning and engineering. The daily records of water movement; flows, withdrawals, rainfall or quality parameters; have been and remain a standard data set used by water resource managers.

A second reason for a focus here was the lack of a recognized standard product in the marketplace to deal with such data. By the lack of a standard product, I do not mean a lack of a general product. There are many products to choose from, but there is no single system which has been unofficially accepted by a broad base of water managers as the means for storing such data. Much of the first year of the database effort focussed on time series data. Windows based software was developed to provide easy access to time series data for the study partners. It is interesting to note that while there was a lack of adequate software at that time, today comparable products can be bought virtually off the shelf.

Geographic Information Systems (GIS)

The GIS data disseminated within the comprehensive study will all be in ERSI's Arc/Info format and in the UTM Zone 16 coordinate and projection system. This was probably the easiest format decision because there were standard packages and usage compatibility. The advantage of having a recognized leader in a field was manifest by the ease of this decision. For others in this situation, this decision may not be so straightforward because there are competing systems which may have a major impact on how individuals decide to handle geographic data.

Fortunately for the comprehensive study, this was one area where consensus agreement was a fairly easy process.

Database

The database formats have provided and will continue to provide the greatest array of difficulty. About 14 different contractors are working on various tasks within the comprehensive study and will be contributing to this collection of data. Most of the contracts were written before the database standards were set up, so various groups have been well underway using formats which may be difficult to truly integrate into a smooth flowing product. This diversity of data will be incorporated into the final product, however the linkages may not be as smooth as desirable.

Conflicts in database deliveries will fall into two different categories: diversity of formats and diversity of data structures. By diversity of formats I mean, some data will be in dBase formats, some in FoxPro formats and some in a variety of spreadsheet formats. While conversion between formats is possible, it does not always work as well as vendors would like you to believe. In addition, including spreadsheet formats adds a major degree of complication. Spreadsheets can be converted to database formats, but only if the spreadsheet has been rigorously constructed to the constraints that a database will require.

A major advantage of spreadsheets is the enormous amount of flexibility they provide the user, flexibility which is not present in database formats. However this flexibility can make conversion to database formats exceedingly difficult, and completion within time and budget estimates virtually impossible. Much of the data developed within spreadsheets will have to be presented as spreadsheets. This precludes some data linking that would have been possible had standards been established and followed earlier in the process.

Text

To date no deliveries of text in electronic format have occurred. The most significant problem with text will probably result from the fact that it was not originally intended to be a part of the database effort. It is only within the past year that the concept of including the text of the reports on the CD's has arisen. Its late entry into the arena may however cause a variety of formats to be delivered, and many contracts probably do not specify that the text must be delivered in electronic format at all. Since all text is probably being produced in electronic format, most contractors will probably be willing to deliver that electronic format even if it was not in their contract. The problem is that deliveries will most likely be in a variety of formats.

At the minimum, all text delivered will be included on the CD in what ever format it was delivered. This will of course be better than not including it at all, but not as good as having it all in the same format and having it linked together.

Multimedia

Multimedia software is the newest and, from some perspectives, the least understood of the software packages to be delivered. Literacy with these packages is still in development, but such packages might provide some relief from the problems cited above. Our vision at this time is to present an outline of the entire comprehensive study developed within the multimedia software which will then electronically point users to text, databases, spreadsheets, GIS and time series data of interest. The minimum result should be pointers to the data of interest, with the user then having the responsibility to provide their own software to access the data. While not perfect, of course, this will be better than providing no electronic access at all.

METHODS

The database will contain data developed by the database contractor and by contractors for the other elements of the Comprehensive Study. The data developed by the database contractor consists of four major portions. These are A) time series data, B) time series related mapping, C) base map data and D) a GIS coverage of information about dams and reservoirs.

Time Series Data

The time series data will be delivered on a separate CD which will be available primarily to the study partners. The reduced availability is a function of the proprietary nature of the software which accesses this data. The data itself consists of the time series histories of all stream and rainfall gauges in the ACT and ACF basins.

Time Series Related Mapping

The mapping associated with the time series data will also be on the Time Series CD and it will consist of a GIS coverage with a database of information regarding all of the gauges featured in the time series histories. The database will consist of data such as name, ID number, parameter and statistic collected, state and county location, latitude, longitude, UTM coordinate, hydrologic unit code, elevation, stream gauge drainage area and the period of record. In addition, statistical summaries of the historic data collection are available in this database.

Base Map Data

The base map data will consist of five map coverages which were derived from the USGS DLG's. These will be delivered on a second CD which will have a much wider distribution. The five base map coverages are: RIVER, LAKE, HUC, COUNTY and INTERSTT.

The RIVER coverage consists of a subset of the USGS hydrographic DLG which includes all features which are named as a 'River' or a 'Fork' of a river. This reduces the mass of data to a volume which will make it manageable for more casual GIS users. The LAKE coverage duplicates some of these rivers, but is a more reduced subset of the USGS DLG which incorporates only the lakes featured in the comprehensive study. This is a polygonal coverage meaning data is associated with the area of the lake as opposed to the RIVER coverage where the data is associated with the lines which make up the lake, river or stream shorelines.

The HUC, or hydrologic unit code, coverage is a polygonal coverage which includes all the drainage basins as delineated by the

USGS. The COUNTY coverage is a series of polygons representing all the counties in the comprehensive study. The INTERSTT coverage is a linear coverage of the Interstate highway in or near the comprehensive study basins. It is provided primarily as background reference for maps.

Database of Dams

The GIS and database of Dams includes 65 fields of information about each major dam featured in the comprehensive study. The data in this database includes information such as owner, location, purposes, locks, power generation, capacities and structural information.

Other Contractors Products

The data developed by other contractors is not under the control of the database contractor. Each contractor has their own specific mission. When completed, they are to deliver their data to the database contractor whose role relative to these efforts is twofold. The first role is to establish standards as described earlier in this paper. The second role is to collect and distribute the databases. The other scopes of work which will be contributing data to the master database are:

Population & Employment
Groundwater
Surface Water Modeling
Navigation
Power
Recreation
Agricultural
Municipal & Industrial
Environmental
Apalachicola Bay

The data delivered will be in the form of databases and/or GIS coverages. Many of the databases will be attachable to the GIS coverages developed in the database effort. For example, several tasks are planning to deliver data on a county by county basis. These tables can be attached to the COUNTY coverage and and information contained in those databases can be displayed in the GIS without the necessity of developing a GIS within the individual tasks. Other tasks are delivering data groups in amalgamations of counties. This too can use the COUNTY coverage as a basis for data display.

Many of the tasks will be developing GIS coverages within their individual efforts. By adhering to a uniform coordinate system, all coverages will overlay each other and spatial data developed in one scope can readily be compared with spatial data developed by others. The use of the GIS capabilities in this way will allow individual contractors' data to be visually compared and analysized, without the need for a rigidly controlled database structure.

CONCLUSIONS

The overriding conclusion is that for effective linkages of products, any database effort must have more emphasis at the beginning of a study of this magnitude. At this point we are attempting to develop linkages in the middle of the effort, and we still see a strong potential for success. However, were the process to be initiated again, the establishment and agreement on electronic standards would best be done prior to the issuance of the first contract.

DISCUSSION

In the current era of exploding technology, it is difficult to produce a result from a five year study which meets the technological capabilities of the time of completion. To do this one would have to have known at the planning stages what the technologies would be at completion. Let me remind you that in 1990, computers were not on virtually everyone's desk, few people had CD-ROM drives; few were on the Internet; nobody had 486's; not many people were using Microsoft Windows; and GIS was a world exclusively for GIS guru's. All of these changes have occurred since this study began.

The question arises whether, in the process of a five year study, managers should decide up front what is going to be done and deliver a product based on that technology, or whether managers should react along the way and try to incorporate new capabilities as they present themselves. While arguments can be made for both sides, managers of this study have tended to incorporate new capabilities as they presented themselves. The evolution of technology has made a successful integration of the data produced by the study more possible than before. Almost every advance of technology that came along was greatly needed, and we are fortunate to have been able to incorporate so many of them.

SUMMARY AND RECOMMENDATIONS

An obvious recommendation is to do as much of the database design as soon in the process as is feasible. Even with the continually changing technology that we are faced with, it is rare to have done too much database planning too soon.

Another key recommendation is to seek and employ those tools that improve our ability to integrate data. Our goal is to provide users of the database with the tools and ease of use that will allow them to find what they need as quickly as possible, regardless of whether the data is in text, spreadsheet or database format.

ACKNOWLEDGMENTS

Acknowledgement is given to the four study partners of the comprehensive study for their foresight to include a database effort within the study. Acknowledgement is also given to every contractor of the comprehensive study, all of whom have been extremely cooperative in meeting the needs of the database effort, even when their contracts did not specifically require such efforts.