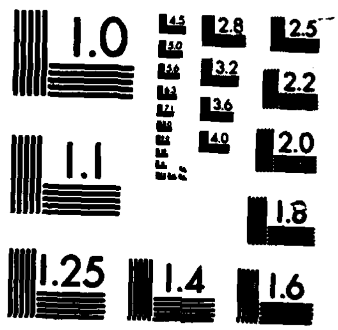
AD-A150 835 COMMERCIAL DATA BASE MANAGEMENT SYSTEM FOR GEONAMES DATA BASE(U) PLANNING SYSTEMS INC SLIDELL LA G LANGRAN ET AL. DEC 84 NORDA-73 N00014-83-C-0657 1/1

UNCLASSIFIED F/G 9/2 NL

END

FILMED

DTIC




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

**Naval Ocean Research and Development Activity**

NSTL, Mississippi 39529 NORDA Report 73 December 1984

# Commercial Data Base Management System for Geonames Data Base

Final Report

nsored by Defense Mapping Agency HQ/STT

AD-A150 835

DTIC ELECTE MAR 4 1985 S D B


John Campbell
Edward Gough

Planning Systems, Inc.
Slidell, Louisiana

Gail Langran

Mapping, Charting, and Geodesy Division
Ocean Science Directorate


DTIC FILE COPY



85 02 19 103

# Foreword

Currently available commercial data bases are screened according to their ability to meet DMA's future geographic names processing requirements. Vendor comments, current applications, and level of user satisfaction are reported. These criteria will be applied when selecting a data base management system to support DMA's geographic names data base.

R. P. Onorati, Captain, USN
Commanding Officer, NORDA

# Executive summary


The Defense Mapping Agency ~~(DMA)~~, with the help of the Naval Ocean Research and Development Activity ~~(NORDA), Code 351~~, is conducting a large effort to automate production of maps, gazetteers, charts, and similar products. One element of this effort is the creation of a single, controlled source for the placenames used in these products: the Geonames Data Base. Because of its size (up to $10^{10}$ bytes) and user requirements (up to 25 simultaneous users), the Geonames Data Base belongs to a special subset of current data base management practice, the Very Large Data Base.

From requirements generated earlier by DMA, NORDA, and others, this task surveys commercial data base management systems (DBMS) to determine candidates for the Geonames DBMS. These systems are then ordered by objective and subjective measures of performance to help in selecting the Geonames DBMS. Among the systems which may satisfy the requirements of the Geonames Data Base, two are used widely today and are exceptionally well regarded by current users. These systems are

- IDMS/R — Cullinet
- ADABAS — Software AG of North America.

# Acknowledgments

This work was sponsored by DMA under Program Element 64701B, with the subtask title, "Geonames Processing System." The DMA Program Managers were Lt. Col. Tom Baybrook and Dennis Franklin. Gail Langran of the Pattern Analysis Branch, Mapping, Charting, and Geodesy Division (Code 350) of NORDA was the principal investigator. John Campbell and Edward Gough, both of Planning Systems, Inc., were under contract #N00014-83-C-0657 to NORDA Code 350 for a significant portion of the work presented in this report.

# Contents

# Tables

# Commercial data base management systems

## 1. Introduction

### Background

The Defense Mapping Agency (DMA) is developing automated tools to aid production of maps, charts, gazetteers, and similar products. One key element to successfully developing these tools is having a single, controlled source of digital placenames: the Geonames Data Base.

Currently there is no analog to the Geonames Data Base. Gazetteers are primarily collected from the Foreign Place Names File (FPNF) containing about 4.5 million names and associated data stored on index cards. These names represent features appearing on 1:250,000 scale maps. The names for new maps are taken primarily from old maps, at scales as large as 1:100,000 or 1:50,000, accounting for features at far greater detail and, thus, many more names. The total number of names from these and other sources is expected to be about 60 million.

As envisioned, the Geonames Data Base will collect these names into a single digital resource of up to 153 data bases, depending on the implementation scheme adopted. All cartographic and toponymic applications will use this common base, thus providing DMA the opportunity to control accuracy, consistency, and agreement among these data products. The job of controlling the base, maintaining data integrity, and other functions to be discussed belongs to the data base management system (DBMS), the software that manipulates and provides access to the data base.

Data base management systems may be either designed and built from scratch or acquired and adapted from commercial sources or agencies that have developed similar systems. Developing a DBMS from scratch for Geonames Data Base is a high-risk task, particularly when the designer must deliver high-speed transactions and a high level of data integrity with a very large data base. Well-established DBMS products have long since completed this phase of development and associated risk.

For most applications, and certainly for the Geonames Data Base, adapting an existing system is more cost efficient. The marketplace for DBMS has grown at a rate of 50% per year over the past seven years (McClellan, 1984), supplying a large number of products that have been developed, tested, and refined on other users' applications. This market has been relatively efficient in turning academic ideas into specific realizations when doing so has meant satisfying customer demand and improving performance. Many products (as discussed later) are able to satisfy most of the demands of the Geonames Data Base, and a few handle all the requirements articulated so far.

The characteristics of the Geonames Data Base have been described by Brown et al. (1983), Langran (1984) and Langran et al. (1984). From these descriptions we know the base to be

- large,
- multiuser,
- dominated by production functions.

We also know that building the data base is the largest technical and practical problem facing the Geonames Data Base. Very few of the projected 60 million placenames are now in digital form. Data capture techniques are being developed in parallel with this project, which will increase the rate of data entry—currently to be accomplished by hand. Still, Langran (1984) estimates 10 years to bring the base to a full 60 million records. This slow rate of entry of new data implies

- *data independence is critical*: Everything—hardware, software, etc.—may change over the next 10 years, but the data base must not be allowed to become obsolete.
- *data integrity is crucial*: Users must always trust the base.
- *backup is imperative*: The base must never be lost.
- *the INSERT function of the DBMS is initially important*: A consideration for benchmark.

The high value of the data base is a common element to almost all DB applications and large commercial systems address these issues.

### Purpose

The purpose of this task is to determine which, if any, commercial DBMS products satisfy the stated requirements for the Geonames Data Base.

### Approach

The task can be considered to be composed of five subtasks:

- Determine if commercial DBMS practice will support Geonames Data Base applications.
- Establish a list of candidate DBMS for the Geonames Data Base from the first subtask.
- Rank candidate systems accordingly if objective performance measures are available.
- Incorporate subjective performance measures where available.
- Recommend a course of action.

The methods employed in accomplishing the subtasks are described in Section 3 of the report. Briefly, these methods included literature search, vendor interviews, vendor promotional literature, published user surveys, and Planning Systems, Inc. (PSI) user surveys.

## Organization of this report

This report is organized to follow the subtasks described in the preceding paragraph. First, we establish that the Geonames DBMS as presently described is within current commercial DBMS practice, although because of size and access requirements belongs to a special subset of the market. To do this we describe how the views of users, administrators, and designers impose constraints on any DBMS that is used with very large data bases, not just the Geonames application. The commercial data base market is then briefly introduced and shown to meet similar concerns.

In Section 3 the list of all DBMS available on the market is trimmed by two levels of discrimination, leaving a list of systems that appears to support the application. In Section 4 we attempt to order the remaining systems by objective and quantitative measures of performance, but find that no such measures exist independently of a specific data base design and implementation. We are thus limited in our attempt to make a quantitative evaluation. We then rely on subjective assessments, such as published user surveys, and our subjective reading of the literature to rank candidate systems and make our recommendation.

# 2. Commercial DBMS support for the Geonames Data Base application

## The Geonames Data Base

The Geonames Data Base belongs to a special subset of data base practice because of its projected very large size and multiuser environment. These two factors impose most of the meaningful constraints that must be used when selecting a system. The eventual size of the data base has been projected to be

- $10^{10}$ bytes,
- 60 x $10^6$ records,
- 153 different bases corresponding to different geographic areas or gazetteers.

The size of this data base suggests many potential problems as discussed by Brown et al. (1983). Among these problems are

- difficulties in building a large base,
- difficulties in maintaining a large base,
- controlling potentially long search times,
- access and maintenance problems caused by scattering the base over many disk storage devices.

These and many similar unidentified issues must be raised and specifically addressed in the data base design process before a practical solution can be implemented in a DBMS. The DBMS is not the solution to these problems; rather, it is the tool through which the solution obtained in the design process is implemented. Without that solution the tools can only be judged either abstractly or with respect to their performance in other similar applications. The DBMS can be separated into two sets: those able to efficiently operate on data volume as large as the Geonames Data Base in a multiuser, multiple storage unit environment, and those unable to do so. Answers to more specific questions must await design and benchmark.

In addition to the size of the base, up to 25 users must be able to simultaneously access the system without undue delay. This suggests many more problems (Brown et al., 1983) including

- data base integrity (concurrency controls),
- security (in the weak sense of individual user files and system function),
- conflicts in the design process between ease of use, flexibility, and speed.

The system must also support such I/O devices as printers, plotters, work stations, and displays. The level of support provided by the target operating system (in terms of user scheduling, LOGON/LOGOFF, and access to the CPU and storage devices) will affect the degree of administrative processing overhead the DBMS will absorb. In addition, the system must provide utilities for DB maintenance, security, access, reporting, etc., for all users. Again, without a specific design it is difficult to quantitatively assess the efficiency of commercial DBMS products. Instead, we classify systems as either capable of such tasks at such a scale or otherwise not qualified.

Other constraints on the Geonames DBMS derive from the *collective expectations of the people involved with the Geonames* Data Base rather than from the necessary form of the base. These people include

- users—toponomysts, cartographers;
- designers;
- data base administrator.

Their views have a practical impact on DBMS selection that cannot be ignored and must be accounted for in the DB design process. Some of these concerns have been suggested by Brown et al. (1983) and by Langran et al. (1984).

## The user's view of the Geonames Data Base

The hierarchy of users of the Geonames Data Base is described by Langran et al. (1984). We are particularly interested in system performance as viewed by the Applications Analysts, i.e., the cartographers and toponomysts. Brown et al. (1983) described five classes of users and 15 example queries. These users include

- toponymic queries—concerned with lexical attributes of a geoname or sets of geonames.
- gazetteer production—concerned with all geonames in a country whose feature attributes meet certain well-defined criteria (e.g., population size).
- map production—concerned with name, location, and ancillary information needed to determine type size, format, etc., of geonames.

- outside queries—sundry requests by non-DMA personnel. Queries may be any sort regarding names.
- data base builders—massive updates to the data base corresponding to a newly digitized map, etc.

The 15 queries collectively exhibit a broad range of expectations from the data base. Large extractions supporting production of maps or gazetteers suggest standard applications packages written to navigate the base quickly over predetermined paths. Other queries such as "Extract all items in a 2° x 2° box, except for waterways" or "Extract all geonames of a certain form which came from a given reference source" seem to demand a DBMS sufficiently flexible as to allow the user to interact with the Geonames Data Base directly without formal help from an applications programmer.

Whether these classes of users and sample queries accurately address the real or perceived needs of the applications analysts should be determined in detail as an early part of the Geonames Data Base design process. The Comprehensive Coordination Plan (Brown et al., 1983) emphasizes the many assumptions made to achieve estimates of performance for planning. While these assumptions were appropriate for preliminary analysis, they may or may not accurately reflect the expectations of the Applications Analysts. Closing this loop between ultimate user and DB designer has proven to be absolutely vital to implementing and operating a successful DBMS. Although we do not expect our conclusions to change as a result of this process, the DBMS should be reviewed with respect to the new list of transactions to be performed.

## Constraints imposed by the user

Collectively, the user base seems to demand
- ease of use,
- fast response.

Unfortunately these attributes are normally mutually exclusive and in a large, complex environment such as Geonames Data Base, tradeoffs between them are hard to establish. Instead, strategies must be developed that tend to support both objectives. Relational data structures and powerful query languages are generally conceded to make user interaction easier, though slower. Navigation aids such as inverted pointer tables, hashing, or binary trees can be developed to enhance speed for predetermined, production-oriented queries regardless of the data model chosen.

The DBMS should, therefore, be flexible enough to reconcile such hybrid approaches to improved performance—and be tolerant of different views of the DB among the users.

## The DB Administrator's view of the Geonames Data Base

The Data Base Administrator (DBA) is at the top of the Geonames Data Base personnel hierarchy, and his concerns necessarily include those of all the functions under his authority. He is, however, exclusively concerned with the administration of the DB, including
- building and maintaining the Geonames Data Base,
- controlling user interaction with the Geonames Data Base.

These responsibilities tend to give the Administrator a different view of the data than that taken by the users of the system. The administrator is more likely to view the data base as a capital investment (which it is) rather than as just another mapmaking tool for the cartographer (which it also is). The base's value depends on its accuracy and completeness as a whole, and a substantial effort to assure those qualities is justified.

## Constraints imposed by the DB Administrator

The DBA expects the DBMS to provide the tools which assure the integrity, security, audit trials, back-up utilities, etc., that provide him with the means to measure system performance, trace transaction histories, and recover from minor disasters. Major disasters must never happen.

Langran et al. (1984) discuss the constraints imposed on the DBMS by the DBA:

*Data Integrity*—A data base is useless if the accuracy or validity of the stored data is questionable. The DBMS should allow the definition of data elements both by type (e.g., integer, character string) and by range limits of acceptable values, and have facilities to allow applications programmers to impose additional validation checks as required. Hardware or data transmission errors should be detected and flagged, and failures during a processing sequence should be rolled back to the last correctly processed record with appropriate error message diagnostics. The data relationships (e.g., parent/child one-to-many) established by the DBA should be prevented from unauthorized modifications, and mandatory relationships should be enforced.

*Physical Data Protection*—The volume of data in the Geonames Data Base will make frequent back-up copying infeasible. The DBMS should provide update facilities that do not require manual re-entry of changes for both "father" and "grandfather" back-up copies in the event of media or storage device catastrophe. Accurate records of physical media storage contents should be maintained.

*Data Security*—Facilities to prevent unauthorized users from modifying data or reading sensitive data should be supported. Access restrictions to the level of data item (not just to files or records) is required.

*Data Independence*—To preserve flexibility for future enhancements, the DBMS should provide data independence between applications software and the physical structure of the data base. It should be possible to modify the physical structure without affecting either the logical structure (the user's view of data base) or previously written data access programs.

*Management Information Statistics*—Information on current system functions, changes in data volume, and measures of DBMS efficiency over time is required by the DBA to monitor system performance and to aid in planning system resource allocations. Identification of on-line users, current processing functions, and system hardware allocations should be accessible interactively. Information on volumes (number of elements), access frequencies, and measures of update volatility of various data base components would be made available to system operators.

Tracking trends in data base growth and retrieval efficiency will allow the DBA to trade off storage allocations, access algorithms, and user demands in tuning the DBMS for optimal performance. Advance warning of processing bottlenecks will facilitate system adaptability to change.

Audit trail functions, including records to identify the user who inserted or modified a data element, when this occurred, and references to special circumstances of the event are required to support some types of toponymic inquiries. This information is not required to be stored on-line or available for general access, and occasional requests of this type are expected to be processed in batch mode. System software to automatically capture most of this audit information during data base loading is desirable.

## The designer's view of the Geonames Data Base

The architecture of the Geonames Data Base is the key and missing element that connects the various requirements to the data and involves the practical problems posed by the need to devise logical data structures and translate them to physical structures, to devise access strategies, to assure data independence, etc. How well the design is accomplished probably has more impact on the ultimate system performance than the machine, the operating system, or the DBMS.

Methodologies for data base design abound, and a thorough review of that technology and its application to design of the Geonames Data Base is beyond the scope of this effort. Whether the design data model should be network or relational is a matter for the designer to determine after more detailed specifications of the requirements are developed and the computational environment decided. Only then can intelligent design decisions be made.

## Design constraints

In the absence of specific design requirements, a designer will choose the DBMS that has the most tools and places the fewest restrictions on the DB implementation. The designer will be concerned about

- performance,
- storage efficiency,
- retrieval efficiency,
- data independence, etc.

The DBMS should provide both batch and interactive access to the base, navigation aids to increase speed, easy access to records distributed across multiple storage units, facilities to provide data migration and data clustering, etc. The DBMS should have an approach flexible enough to allow the designer whatever logical data model he chooses, or even to mix models within the base. These requirements are within current commercial practice.

The three data models most commonly used by DB designers are

- relational,
- network,
- hierarchical.

A DB using the relational model is made up of flat files and a management system that recombines the data elements to form different files (Martin, 1977). This type of DBMS has the following strengths:

- user determines the view,
- ease of use,
- mathematically elegant,
- data model is simple.

Because only data is stored in this type of DB, the users can select how they wish to organize or view the data. Since the data records do not contain pointers that "point" to other records, the user does not have to know how to maneuver through the data base. The relational data model uses very few data structures, composition rules and attributes, making it an elegant model (as defined by McGee, 1976). The data is stored in only one type of record, making the data more easily understood.

Although a relational DBMS has the above advantages, it also has the following disadvantages:

- slow,
- larger storage requirements.

A relational DBMS is able to create additional files by finding the common data element in more than one file and combining the data from different files. This implies that the same data element is stored more than once, using additional storage. Although there is no theoretical reason why a relational DBMS should be significantly slower than systems using other data models, it is widely reported that this is the case (Martin, 1977; McGee, 1976; Larson, 1983).

A DB using the network data model has "child" records in a structure with more than one "parent" record. The term "child" means that the data stored in record A is referenced or pointed to from the data stored in record B, the parent. In the case of the network model, one child record may have many parent records. This type of data model has the following advantages:

- speed,
- well-documented,
- widely used,

and has the following disadvantages:

- complex,
- difficult to implement,
- difficult to use,
- difficult to update.

Because of the use of pointers in the network model the DBMS usually provides a high performance. This type of data model has been used for many years to achieve the performance needed for very large data bases; as a result, it is used widely and is well-documented. This approach is appropriate for systems with clearly defined data relationships and query requirements, and where changes to record structures are not anticipated.

The child/parent records in a network DB makes the base difficult to understand. If the data base is large, as in the case of the Geonames Data Base, then there may be many child/parent relationships that need to be tracked. This complexity makes the base both difficult to implement and difficult to use. Also, if the data base is moved from one storage device to another the pointers stored in the DB must be changed, making updates difficult.

The hierarchical data model is similar to the network model except that each child record is allowed only one parent. Although this reduces the complexity of the DB design, it also makes the DB more inflexible and more difficult to change.

The relative advantages and disadvantages among the three available data models do not immediately suggest one method over the others in the Geonames Data Base application. The conceptual (and practical) case of the relational data model is very appealing with respect to user access to the base, but the observed burden on system performance is problematic for the Geonames Data Base. Network systems resolve performance problems but are more complex both conceptually and in implementation. Some commercial DBMS vendors have resolved this conflict over ease-of-use versus performance by either applying performance enhancements to relational structures (e.g., INGRES) or imposing relational data models on essentially network DBMS (e.g., IDMS/R).

Some of the techniques used to enhance performance of relational bases include

- execute DB functions in hardware,
- B tree storage techniques,
- hashing storage techniques,
- use of inverted files,
- optimizing queries,
- use of virtual space.

Modifying a network data base to allow for the use of the relational data model is a complex and product-dependent issue. Primarily, it involves allowing the user to define the data relationships and have the system determine how to implement the relationships.

Another option to achieving both high performance and ease of use is to design parallel data bases that use two different DBMSs. A disadvantage to this approach is that the DB administrator must maintain updates to both data bases. IBM offers a DBMS using the hierarchical data model, IMS, and a DBMS using the relational data model, SQL. IBM also supports software to transfer data between each type of base. This allows for the use of a high-speed DB for production runs using the hierarchical model and a very flexible relational DB for all other needs.

# Commercial data base management systems

## DBMS for large DB/multiusers

Data base applications with the size requirements of the Geonames Data Base have been operationally successful since 1968 when POLAR, the Production Order Location and Reporting System, was implemented for the Apollo program using IMS. By 1969 this system consisted of 30 data bases spread over 32 disk packs and supported 130 terminals (Grafton, 1983). Since then Very Large Data Bases (VLDB) have been implemented to support many large-scale projects, and the technology to support the VLDB has been expanded to include distributed data bases (Rothnic and Goodman, 1977; Polk and Byrd, 1981) and data base machines (Hsiao, 1979). These advances, furthermore, are being translated into commercial products and novel applications, as witnessed by recent implementation by Products Diversified of Houston of an on-line real estate DBMS of $15 \times 10^6$ records using Britton Lee's IDM data base machine with Alpha Micro front end processors (News item, *Datamation*, March 1984).

As shown in Section 3 of this report several, though not all, of the commercial DBMSs that operate on hardware/OS that support large on-line storage capacity and multiple users have chosen also to support VLDB. These systems make up the set of feasible solutions to the Geonames Data Base.

## Constraints imposed by the Geonames Data Base typical of other DB

The constraints imposed on the Geonames DBMS by users, administrators, and designers are common to most other similar systems. Every DBMS must provide similar capabilities to its users.

## Commercial market for DBMS

The commercial market for DBMS has been efficient in turning new academic ideas into improved products. The successful vendors in this market are primarily small, young companies that concentrate their efforts into a limited range of software products. These vendors tend to be responsive to market pressures for new capabilities (McClellan, 1984). Data base products introduced using network data models, e.g., IDMS, have been modified to allow relational data models while still supporting network and hierarchical structures. Mixed data models are not common. Every product has implemented one or more theoretical approaches to increasing performance in terms of speed. This strong market benefits all potential users and the Geonames Data Base application in particular.

# 3. Selecting candidate DBMS

## Approach

We now must assemble a list of products from this marketplace that match the Geonames Data Base application. To do so we need both a list of products and their technical descriptions as well as a set of discriminators to disqualify the products which are not likely to serve the task.

The list of products is easily made from proprietary software reviews, e.g., Datapro, Auerbach, Data Sources, etc., and trade journals. Technical data from these product reviews were supplemented by our own phone survey of vendors (Appendix A).

Since no specific design for Geonames Data Base is available, we look for discriminators to disqualify commercial DBMS from contention for the Geonames DBMS. Taken together, size and number of users provide the best discriminator. Although there are hundreds of DBMSs on the market, only a handful operate

- on hardware powerful enough to support $10^{10}$ bytes of on-line, direct access storage;
- under operating systems capable of reaching extended addresses:
- on systems (hardware and OS) capable of simultaneously serving 25 users regardless of their abilities to meet other DBMS goals.

Much of the market growth of the past few years has supported single-user operating systems (CP/M, MS/DOS, etc.) accessing on-line storage less than $10 \times 10^6$ bytes. Likewise, systems targeted to small- and medium-sized businesses, scientific or numerical data bases, or specialized applications can be dismissed. Systems implemented exclusively on obsolete hardware (e.g., DEC System 10), systems implemented on older minicomputers with limited interfaces, or inadequate operating systems (e.g., HP-3000) can also be dismissed without being troubled by the details of their implementation or the quality of user satisfaction; they simply will not do the job.

## First level of discrimination: Eliminate DBMS whose host systems do not support Geonames Data Base requirements

The first level discriminators were applied to the set of all commercial DBMS products identified in the literature search. When in doubt at this level of discrimination, the DBMS was passed into the list. Table 1 shows a list of 43 systems passed into a closer analysis as initial candidates for Geonames DBMS. These are systems that run on host systems thought capable of supporting Geonames Data Base. When these vendors were contacted by Planning Systems, Inc. (PSI), 11 disqualified their products for reasons listed in Table 2.

Table 1. Systems passing the first level of discrimination.

| Product Name | Company Name |
|---|---|
| ADABAS | Software AG |
| AIM/RDB | Fujitsu |
| ARC | Data Point |
| BASIS-DM | Battele Columbus Labs |
| CLIO | United System Software and Services |
| DATA COM/DB | Applied Data Research |
| DBMS | ISE |
| DB Machine | Maga/Net |
| DM IV | Honeywell |
| DMS II | Burroughs Corporation |
| ENCOMPAS | Tandem |
| EXPRESS | Management Decisions |
| FOCUS | Information Builders |
| FULL-RDM | International Tech |
| GYPSY | University of Oklahoma |
| IDM | Britton Lee |
| IDMS | Cullinet |
| IMAGE | HP |
| INFO | HENCO |
| INGRES | Relational Technology |
| INQUIRE | Infodata |
| ITX | NCR |
| MAXXIMUM | California Software Products, Inc. |
| MODEL 204 DBMS | Computer Corporation of America |
| ORACLE | Relational Software |
| PAC III | AGS Management Systems, Inc. |
| PLUS/4 | Century Analysis, Inc. |
| RAMIS II/RELATE | MPG |
| RAPPORT | Logica |
| RD4 | Hitachi |
| RELIANCE PLUS | Perkin Elmer |
| RIM | Boeing Computer Services |
| SEED | Seed Software |
| SIR/DBMS | Scientific Information Retrieval |
| SQL, DB2, DLI, IMS | IBM |
| SQL/UNIVERSE | INCO |
| SUPERSETUP | The Automated Quill, Inc. |
| SYSTEM 2000 | INTEL |
| TOTAL | CINCOM |
| VAX II DBMS | DEC |

Table 2. Products excluded on vendors' recommendations.

| Product Name | Reason for exclusion |
|---|---|
| AIM/RDB | Not currently for sale |
| BASIS-DM | Not currently for sale |
| EXPRESS | Not full DBMS |
| FOCUS | Base too large for product |
| FULL-RDM | Base too large for product |
| GYPSY | Not full DBMS |
| MAXXIMUM | Base too large for product |
| PAC III | Not full DBMS |
| PLUS/4 | Base too large for product |
| RD4 | Not currently for sale |
| SUPERSETUP | Base too large for product |

## Second level of discrimination: Eliminate DBMS that do not support Geonames Data Base requirements

The second level of discriminators were derived from specific constraints already imposed on the Geonames DBMS by the functional design specifications. These constraints include the following:

- Size,
  - —$10^{10}$ bytes
  - —60 x $10^6$ records
  - —multiple on-line disk packs
- multiple Data Base (150).
- automatic backup,
- security,
- relational data model with performance enhancement.

This last criterion is PSI's reconciliation of the conflicting requirements for ease of use and high speed.

Table 3 establishes the relationship of the remaining 30 systems with the criteria established in the second level of discrimination. A review of Table 3 shows that nine DBMS products should be considered as candidates for the Geonames DBMS. These systems are listed.

| Product Name | Company Name |
|---|---|
| ADABAS | Software A.G. of North America |
| CLIO | United Software Systems and Services |
| IDM | Britton Lee |
| IDMS/R | Cullinet |
| IMS/DL1 | IBM |
| INGRES | Relational Technology |
| Model 204 DBMS | Computer Corp. of America |
| SEED | Seed Software |
| TOTAL | Cincom |

Table 3. Results of the second level of discrimination.

| Product Name | Size | Multiple Data Bases | Automatic Backup | Security | Relational with Speed Enhancements | Data Integrity | Data Independence |
|---|---|---|---|---|---|---|---|
| ADABAS * | ● | ● | ● | ● | ● | ● | ● |
| ARC | | ● | ● | ● | | ● | ● |
| CLIO | ● | ● | ● | ● | ● | ● | ● |
| DATA COM/DB | | ● | ● | ● | | ● | ● |
| DB Machine | | ● | ● | ● | | ● | ● |
| DBMS | | ● | ● | ● | | ● | ● |
| DL1, IMS * | ● | ● | ● | ● | | ● | ● |
| DMS II | ● | ● | ● | ● | | ● | ● |
| DM IV | ● | ● | ● | ● | | ● | ● |
| ENCOMPAS | | | ● | ● | | ● | ● |
| FOCUS | | ● | ● | | | ● | ● |
| IDM * | ● | | ● | ● | ● | ● | ● |
| IDMS/R * | ● | ● | ● | ● | ● | ● | ● |
| INFO | | ● | ● | ● | | ● | ● |
| INGRES | ● | ● | ● | ● | ● | ● | ● |
| INQUIRE | ● | | | | | ● | ● |
| ITX | ● | ● | ● | ● | | ● | ● |
| Model 204 DBMS | ● | ● | ● | ● | ● | ● | ● |
| ORACLE | ● | | ● | ● | | ● | ● |
| RAMIS II * | | ● | ● | ● | | ● | ● |
| RAPPORT * | | ● | ● | ● | | ● | ● |
| RELIANCE PLUS | | ● | ● | ● | | ● | ● |
| RIM | | ● | ● | | | ● | ● |
| SEED | ● | ● | ● | ● | ● | ● | ● |
| SIR DBMS | ● | ● | ● | ● | | ● | ● |
| SQL DB2 * | ● | ● | ● | ● | | ● | ● |
| SQL UNIVERSE | | ● | ● | | | ● | ● |
| SYSTEM 2000 | ● | ● | ● | ● | | ● | ● |
| TOTAL | ● | ● | ● | ● | ● | ● | ● |
| VAX II DBMS | ● | | ● | ● | | ● | ● |

# 4. Performance measures

## Objective performance measures

Having established a list of nine candidate systems we would like to establish a set of objective measures of performance to judge the best for the Geonames DBMS. Unfortunately, no such objective measure can be determined. To do so would require the following:

- The hardware/OS must be determined.
- The data base design must be completed.
- A realization of the DB and the DBMS must at least be modeled and perhaps implemented for benchmark.

Other objective experiences in benchmarks for other system selections should be consulted when available, but caution must be exercised. All such benchmarks are conducted against some specific data base realization. How these data bases are constructed are usually beyond the scrutiny of secondary users of the benchmark, and so the results cannot be judged objectively.

Objective measures of the Geonames DBMS will be made only when the data base design is completed.

## Subjective performance measures

Without an objective measure of performance our final level of discrimination must be subjective. We choose to look at the candidate systems and the companies offering them through the collective eyes of their customers. Two recent surveys of DBMS users, one by Data Decisions for *Datamation* and the other by Datapro, show similar results: Cullinet and its IDMS product, and Software AG and its ADABAS product, are regarded highly by users. This result is reflected in the market share captured by these two companies over the past seven years (McClellan, 1984).

In December 1983, *Datamation* published the results of a software survey conducted by Data Decisions in July 1983. Shown are the customer satisfaction scores for six of the products that passed the previous screening. The scores shown are relative to a maximum of 10.

| | | | |
|---|---|---|---|
| IDMS | 7.7 | IMS | 6.5 |
| ADABAS | 7.7 | SQL | 6.1 |
| TOTAL | 7.1 | DL1 | 5.7 |

Only four DBMS products received higher scores.

| | | | |
|---|---|---|---|
| SAS | 9.0 | by | SAS |
| SYSTEM 1022 | 8.0 | by | Software House |
| DMSII | 7.9 | by | Burroughs |
| IMAGE | 7.8 | by | Hewlett Packard |

One DBMS received a score equal to IDMS and ADABAS:

FOCUS 7.7 by Information Builders

SAS by SAS, Inc., according to the technical description in Datapro, is not a complete DBMS but is a report writer with IMS doing the DBMS functions. SYSTEM 1022 by Software House runs only on the DEC System 10 computer for which DEC is reducing support. DMSII by Burroughs did not pass the previous screening discussed in this report because it is not a relational DBMS. IMAGE by Hewlett Packard runs on the HP-3000, which can support only 24 users. FOCUS by Information Builders was disqualified by the vendor because FOCUS is not designed for a data base of the expected size of the Geonames Data Base.

The vendors surveyed by Data Decisions were selected from a list supplied by International Computer Programs. This list contained software vendors with gross sales of more than $5 million and hardware manufacturers with at least 30 user sites. These vendors supplied Data Decisions with lists of the most recent 125 customers using their product. Vendors were requested not to contact the users they supplied, and Data Decisions verified this. Questionnaires were sent to the key vendor contacts, usually the data processing managers. If necessary, telephone interviews were conducted to insure there were at least 15 response/packages. *Ratings were based on the following scheme:*

| | | | |
|---|---|---|---|
| 9-10 | Superior | 3-5 | Acceptable |
| 6-8 | Very Good | 1-2 | Inadequate |

Tables 4 and 5 show further details reported by Data Decisions in *Datamation*. Overall satisfaction of product includes such factors as package features, capabilities, and utility and frequency of failure requiring extra effort for recovery. Overall satisfaction of support reflects the user's appraisal of installation, documentation, modification, and training. Performance economy/efficiency includes such factors as hardware resource utilization, ease of use, freedom from bugs and errors, and time required for initial installation. Vendor support gauges the vendor's responsiveness to user needs, effectiveness of training, and quality of documentation. Operation is a measure of the package's ability to handle expanding processing volumes, backup/recovery, and security.

Datapro research conducted a software survey during 1982 with the cooperation of *Computerworld* and the assistance of McGraw-Hill Research. Listed below are the four products that have passed the previous screenings and were included in the Datapro survey. The scores shown are the average and overall satisfaction ratings on the scale of 4 used by Datapro, and translated to a scale of 10 for comparison to Data Decision scores.

| | | |
|---|---|---|
| IDMS | 3.4 out of 4.0 | 8.5 out of 10.0 |
| ADABAS | 3.1 out of 4.0 | 7.8 out of 10.0 |
| IMS | 3.0 out of 4.0 | 7.5 out of 10.0 |
| TOTAL | 2.1 out of 4.0 | 5.3 out of 10.0 |

Table 6 gives the details of the results of the survey for these four products. Table 7 gives a summary of the users' opinions concerning the product/vendor advantages or disadvantages. The scores shown in Table 7 are based on the following:

| | |
|---|---|
| 4—Excellent | 2—Fair |
| 3—Good | 1—Poor |

The Datapro survey was conducted by McGraw-Hill Research using a questionnaire designed by Datapro. The questionnaire was mailed to users in May 1982, with a second mailing to nonrespondents in June 1982. Telephone calls were made in July and August to those users who did not respond to the mail survey in order to achieve at least a 50% response rate. The users were selected from a list of subscribers to *Computerworld* with the following job titles and functions:

- Director, Manager, Supervisor of Data Processing Services;
- Systems Manager and Systems Analyst;
- Manager or Supervisor of Programming.

User surveys were not available for objective determination of the following four systems:

- INGRES,
- CLIO,
- SEED,
- Model 204.

Subjective criteria were used in evaluating these systems.

Because of the lack of user information concerning Model 204 by Computer Corporation of America and CLIO by United Systems Software and Services, these two products are not recommended. Datapro reports the first installation of CLIO in 1982. CLIO is not recommended because of the lack of demonstrated use of CLIO on very large data bases. Although Model 204 was first installed in 1969, there are only 125 installations reported in Datapro. Because of the lack of user reports on Model 204 it is considered a technical risk to recommend this system.

INGRES by Relational Technology is currently being used by NORDA on their DEC VAX computer. The users of this DBMS at NORDA claim that for very small data bases (less than 1 Mbyte) INGRES took what they considered an excessively long time to respond to requests.

MITRE Corporation recently tested SEED by Seed Software on the DEC VAX computer. Two PSI employees observed the testing and were included in the analysis of the results. MITRE found that because of the journaling of the entries SEED was very slow. The journaling may be disabled but there would be no lock-out protection against concurrent users reading and writing to the same record. This test was also done on a very small data base.

In reviewing the surveys of both Datapro and *Datamation* two products constantly were rated high by their users: IDMS and ADABAS. While other DBMS products were also well-regarded by their users, these two products stood out. Because of their high user rating these two products passed the final level of discrimination.

Table 4. User opinion scores of products included in Data Decision survey. Maximum score—10.0.

| | Number of responses | overall product | satisfaction support | performance | vendor support | operational |
|---|---|---|---|---|---|---|
| All packages | 1089 | 7.1 | 6.3 | 6.7 | 6.2 | 6.4 |
| ADABAS | 65 | 7.7 | 6.3 | 7.5 | 6.2 | 7.2 |
| DL1 | 48 | 5.7 | 5.7 | 5.4 | 5.4 | 5.7 |
| IDMS | 61 | 7.7 | 6.9 | 6.9 | 7.3 | 7.2 |
| IMS | 35 | 6.5 | 6.2 | 6.1 | 6.1 | 6.5 |
| TOTAL | 53 | 7.1 | 6.0 | 7.4 | 5.9 | 6.5 |
| SQL | 37 | 6.1 | 5.7 | 5.7 | 5.4 | 5.9 |

Table 5. Product scores taken from Data Decision survey.

| | Number of responses | % rated DBMS outstanding | % rated vendor outstanding | % users seeking to replace | % not satisfied w/package |
|---|---|---|---|---|---|
| All packages | 1089 | 77 | 64 | 17 | 3 |
| ADABAS | 65 | 91 | 63 | 2 | 0 |
| DL1 | 48 | 56 | 58 | 27 | 10 |
| IDMS | 61 | 90 | 79 | 2 | 0 |
| IMS | 35 | 69 | 54 | 11 | 0 |
| TOTAL | 53 | 79 | 58 | 38 | 4 |
| SQL | 37 | 64 | 60 | 16 | 8 |

Table 6. Product scores taken from Datapro survey.

| | | IDMS | ADABAS | TOTAL | IMS |
|---|---|---|---|---|---|
| Date of initial testing/use | before 1977 | 7 | -- | 7 | 20 |
| | 1978—1979 | 9 | 5 | 8 | 7 |
| | 1980—1982 | 38 | 27 | 10 | 23 |
| Modification required | Yes by vendor | 7 | 3 | 3 | 6 |
| | Yes by user | 4 | 6 | 3 | 8 |
| | No | 46 | 2 | 21 | 37 |
| Significant advantages noted by users | Simple | 18 | 21 | 14 | 7 |
| | Flexible | 48 | 29 | 13 | 39 |
| | Inexpensive | 5 | 2 | 1 | 6 |
| | Save system resources | 19 | 9 | 12 | 9 |
| | Save human resources | 41 | 23 | 18 | 37 |
| | Compatible | 30 | 11 | 16 | 33 |
| Significant disadvantages noted by users | Inflexible | 1 | 2 | 5 | 9 |
| | Costly | 18 | 11 | 6 | 24 |
| | Complex | 13 | 8 | 1 | 33 |
| | Slow | 5 | 6 | 2 | 11 |
| | Use excessive resources | 8 | 6 | 5 | 28 |
| | Lack key capabilities | 1 | 2 | 8 | 5 |
| Did package perform as required | Yes, immediately | 41 | 20 | 19 | 34 |
| | Yes, eventually | 14 | 10 | 6 | 13 |
| | Never | 1 | 1 | -- | 1 |

# 5. Recommendations

Two DBMS products, IDMS/R by Cullinet and ADABAS by Software AG, passed the objective levels of discrimination and received high scores in the subjective levels of discrimination. These two products are candidates to be used with the Geonames Data Base. The exact hardware/operating system on which the DBMS will reside may change the candidate systems because of compatibility problems.

The next step is to benchmark test candidate products. Benchmarks will allow actual performance standards to be measured. Selected queries of the type expected will be written, a data base generated, and the systems compared. Benchmarking will also allow the products to be compared to each use and user friendliness.

There are two possible sources of data that could be used in the benchmark. The U.S. Geological Survey (USGS) Geographic Names Information System could be accessed and several thousand names along with their associated data would be used. This data base would be quite small by comparison to the entire Geonames Data Base but should provide a very similar type of data base. Multiset III data tapes could also be used for benchmarking of DBMS products. These tapes contain over 100,000 names per tape and represent a large data set for use. However, these tapes contain little additional data besides the name, making accessing on other entries greatly reduced in comparison to the data from USGS. In either of

Table 7. User opinion scores of products included in Datapro survey. Maximum score—4.0.

| Average User Rating | IDMS | ADABAS | TOTAL | IMS |
|---|---|---|---|---|
| Reliability | 3.6 | 3.5 | 3.3 | 3.4 |
| Efficiency | 3.3 | 3.2 | 2.8 | 3.6 |
| Ease of installation | 3.2 | 3.3 | 3.1 | 2.6 |
| Ease of use | 3.2 | 3.4 | 3.0 | 2.6 |
| Troubleshooting | 3.1 | 2.6 | 2.6 | 3.1 |
| Documentation | 3.0 | 2.6 | 2.7 | 2.9 |
| User education | 3.0 | 2.6 | 2.7 | 2.8 |
| Vendor maintenance | 3.1 | 2.9 | 2.7 | 3.4 |

the above cases the data base can be expanded with fictional data through randomization or other programmatic processes.

# 6. Conclusions

There are commercial DBMSs available that are capable of managing the Geonames Data Base. ADABAS by Software AG and IDMS/R by Cullinet are especially good candidates due to a high level of use and of user satisfaction. The decision of which system to purchase should be made at a higher technical level after additional DB details have been designed.

# 7. References


Brown, R., A. Zied, A. Barnes, and E. Gough (1983). *Advanced Type Placement and Geonames Database: Comprehensive Coordination Plan*. Naval Ocean Research and Development Activity, NSTL, Mississippi, Tech. Note 189.

Data Decisions, System Software Survey (1983). *Datamation*, December.

DATAPRO (1983). *Datapro Directory of Software*. Debreaux, New Jersey, v. 2.

Grafton, W. (1983). IMS: Past, Present, Future. *Datamation*, September, p. 158.

Hsiao, D. (editor) (1979). Special Issue on Data Base Machines. *Computer IEEE*, v. 12, n. 3, March.

Langran, G. (1984). *Geonames Processing System Functional Design Specification, Volume 5: Performance Specification* (draft). Naval Ocean Research and Development Activity, NSTL, Mississippi, Tech. Note 259.

Langran, G., A. Barnes, and S. Miller (1984). *Geonames Processing System Functional Design Specification, Volume 2: Geonames Data Base* (draft). Naval Ocean Research and Development Activity, NSTL, Mississippi, Tech. Note 256.

Larson, J. D. (1983). Bridging the Gap Between Network and Relational Database Managements Systems. *Computer*, v. 16, n. 9.

Martin, J. (1977). *Computer Data-Base Organization*. Prentice-Hall, Englewood Cliffs, New Jersey.

McGee, W. C. (1976). On User Criteria for Data Model Evaluation. *ACM Transactions on Database Systems*, v. 1, n. 4.

McClellan, S. (1984). The End of the Hardware ERA. *Datamation*, v. 30, n. 6.

Polk, W. and K. Byrd (1981). Managing the Very Large Database. *Datamation*, September.

Rothnic, J. and N. Goodman (1977). A Survey of Research and Development in Distributed Database Management. *Proceedings of the Third International Conference on Very Large Data Bases, IEEE*.

# Appendix A: Questionnaire for Planning Systems, Inc. phone survey of DBMS vendors

## Information for Geonames Data Base DBMS selection

Company Name:

Company Rep/title:

Product Name:

Machines/OS Product is used with:

Type of DBMS (Relational/Hierarchical/Hybrid):

Can Product handle:

(a) 10 Gigabyte size base:

(b) 60,000,000 records:

Other items to be considered:

1. Does product handle variable length fields? If so how?
2. Does product support multiuser (25), multibases (150), multidisks? How?
3. Does product support aliasing and name variant?
4. Does product support interactive and batch modes?
5. Where may application programs be used (front, back, front and back, interior)? In what languages?
6. What are the maximum number/size of the fields, records, and files?
7. Does product support data dictionary?
8. Does product support data security at record, file, and field levels?
9. Does product support "read only" data security?
10. Does product support data independence?
11. Does product have an audit trail? At what level?
12. How does product support blank fields?
13. Does product give MIS? Of what type and level?
14. What are the backup/restart facilities?
15. Does product support priorities of users?
16. Does product support common working area?
17. Does product support intermediate messages and cancellation procedures for long running request?
18. What type/cost of vendor support and/or training is available?
19. Cost:

    a. Purchase (GSA Rate)
    b. Lease (GSA Rate)

20. Is source language available? (type, cost)
21. Are all DBM system languages included in cost? (DDL, DML)

PREVIOUS PAGE IS BLANK

# Appendix B: Additional notes on selected DBMS products

This appendix contains comments of DBMS products that were selected because of the features the products have or will presently have. These products are

| | | |
|---|---|---|
| SQL/DB2 and IMS/DL1 | by | IBM |
| RAMIS II | by | Mathematics Products |
| IDM | by | Britton Lee |
| RAPPORT | by | Logics |
| GYPSY | by | University of Oklahoma |

IBM currently has four DBMS products for use of their mini and main-frame computers. Two of these products, DB2 and SQL, use the relational data model, and two of these products, IMS and DL1, use hierarchical or network data model. The primary difference between the DB2 and SQL is the hardware/operating system for whch the product is written. The same is true for IMS and DL1. Because the hardware/operating system difference was not part of the selection/evaluation criteria, the two relational products were evaluated as one product and the two nonrelational products were evaluated as one product.

Both RAMIS II from Mathematical Products Group and RAPPORT from Logica are to be upgraded in the future. The representatives of these companies indicated the upgrades would be such that the products should be able to overcome the current limiting factor of the products.

The Britton Lee IDM-500 is a data base machine rather than a software DBMS, achieving good performance through their hardware implementation of relational data model. The IDM product is limited to a maximum of 50 data bases. Although this eliminates the IDM from further consideration based on the proposed data partitioning into 150 bases corresponding to gazetteers, it should be noted that the IDM is successfully being used with a data base of 15 million names. The users of that data base report excellent response time. If alternate data partitioning designs are considered the IDM should be reconsidered as a possible candidate system for the Geonames Data Base.

The DBMS GYPSY from the University of Oklahoma is used by the USGS with their names data base. This product is not a DBMS but a file management system and is used by the USGS as such. However, it does have a proven history of being used on a large data base of very similar nature.

## IDMS*

| | |
|---|---|
| Company: | Cullinet Software, 400 Blue Hill Drive, Westwood, Maine 02090. Telephone (617) 329-7700. |
| Functions: | Data base management system. |
| Hardware Systems: | IBM: System/370, 3000, 4300. |
| Minimum Memory Requirements: | 500K. |
| Operating System: | OS, DOS, VS(E) counterparts, VM. |
| Source Language: | Assembler. |
| Pricing: | Contact vendor. |
| Options: | Central Version, CMS, Distributed Data-base System, DMS interface, Escape/DL/I, Escape/DBOMP, Escape/Total. |
| Maintenance: | First year free; 10 percent of license fee annually thereafter. |
| Documentation: | Included in price. |
| Training: | Included in price, plus expenses. |
| Number of Current Users: | 800. |
| Date of First Installation: | May 1973. |

IDMS/R is the current release of Cullinet's DBMS. It contains within it the previously released DBMS, IDMS.

IDMS (Integrated Database Management System) is a data base management system designed to conform with the CODASYL Data Base Task Group Language specifications. It includes a schema data description language (schema DDL), a subschema data description language (subschema DDL), a device/media control language (DMCL), and a data manipulation language (DML), as well as the data base management (DBMS) modules themselves. IDMS also includes a data dictionary system, which operates from the user-established schema definition of data and a series of data administrator utility programs. The data description language (DDL) is stand-alone and its records descriptions are comparable to those of Cobol. The schema DDL input provided by the data manager completely

*Taken from Datapro Directory of Software.

PREVIOUS PAGE IS BLANK

defines the data base. The data base description is composed of areas, files, records, and logical relationships. The schema compiler validates and stores the schema DDL information in the data dictionary. The subschema compiler generates a series of tables that are maintained in a catalogued file for later interpretation by the data management routine. The subschema is that portion of the data base known to a particular applications program.

Special features include: path calls, multiple dictionary support, high speed terminal response time, re-entrance, record journaling, concurrent update prevention, and automatic recovery with warm start. This last feature allows all unaffected programs to continue normal processing when an individual program fails. The teleprocessing monitor, IDMS-DC, provides a data communications capability integrated with IDMS. IDMS-DC is a dictionary driven and designed specifically for use in the DBMS environment. The Distributed Database system is an option that allows application programs running on multiple CPUs to access and update a common data base. Data integrity is guaranteed, and there is full recovery capability at each machine for application program failure and for machine failure. The DMS Interface, Escape/DL/1, and Escape/DBOMP are a set of interfaces that allow DMS, DL/1, and DBOMP users to access and update an IDMS data base. All of the flexibility, integrity, and security features of IDMS are available to these users.

| | |
|---|---|
| No. of Programs in Package: | 100. |
| Source Language: | ALC. |
| Source Listings: | Are not available. |
| Pricing: | Purchase—$162,000 (MVS), $132,000 (OS), $99,000 (DOS). |
| Maintenance: | Provided for first year; 10 percent of the then current price, thereafter. |
| Documentation: | Provided with purchase or lease. |
| Training: | Included with purchase; available with lease. |
| Number of Current Users: | 750. |
| Date of First Installation: | July 1972. |

## ADABAS*

| | |
|---|---|
| Company: | Software AG of North America, Inc., Reston International Center, 11800 Sunrise Valley Drive, Suite 917, Reston, Virginia 22091. Telephone (703) 860-5050. |
| Functions: | Data base management system. |
| Hardware Systems: | IBM: System/370, 303X, 308X, 4300; Siemens: 4004. |
| Minimum Memory Requirements: | 160K. |
| Operating System: | IBM: OS, DOS, VS, DOS/VS(E), MVS, VM/370-CMS; Siemens: PPS, BS1000. |
| Time-Sharing Service: | Software AG, PRC Computer Center, Inc. |

* Taken from Datapro Directory of Software.

ADABAS (Adaptable Data Base System) is a data base management system with a number of utility programs used under DOS or OS with BDMA for data base generation and access. The system uses a variety of high-efficiency data management techniques and provides a generalized file-coupling capability. The ADABAS nucleus supports concurrent batch and on-line processing. Included with ADABAS is ADASCRIPT, an on-line query language with English-like syntax. Interfaces are provided for popular TP monitors such as COM-PLETE, CICS, TSO, and Taskmaster. ADACOM, a report generator, is also included. A data compression algorithm to load into the data base is an integral function of the system. Also featured is the separation of physical data storage from the representation of logical relationships in the data base. ADAMINT is used to generate high-level interface routines for applications programs. ADABAS also includes an integrated Data Dictionary system and full restart/recovery capabilities.

ADABAS/VM is the version of the product that operates under the CMS component of VM/370. ADABAS/VTAM is the DDP product that allows applications running in one processor to access data in one or more secondary processors connected via a channel-to-channel or VTAM network line.

# Appendix C: Methods of speed enhancement claimed by vendors

| | dB Function Hardware | B tree | Hashing | Link List | Inverted Files | Query Optimizer | Virtual Space |
|---|---|---|---|---|---|---|---|
| IDM * | ● | ● | ● | | ● | | |
| IDMS | | ● | ● | | ● | | |
| ADABAS | | ● | | | ● | | |
| MODEL 204 * | | | ● | | ● | | |
| TOTAL | | ● | | | ● | | |
| CLIO | | | | | ● | | ● |
| SEED * | | | ● | ● | ● | ● | |
| INGRES * | | | ● | | ● | ● | |
| IMS/DLI | | | | ● | | | |

Methods of speed enhancement claimed by vendors.

*taken from direct quotes

Distribution List

Department of the Navy
Asst Secretary of the Navy
(Research Engineering & System)
Washington DC 20350

Department of the Navy
Chief of Naval Operations
ATTN: OP-951
Washington DC 20350

Department of the Navy
Chief of Naval Operations
ATTN: OP-952
Washington DC 20350

Department of the Navy
Chief of Naval Operations
ATTN: OP-987
Washington DC 20350

Department of the Navy
Chief of Naval Material
Washington DC 20360

Commander
Naval Air Development Center
Warminster PA 18974

Commander
Naval Air Systems Command
Headquarters
Washington DC 20361

Commanding Officer
Naval Coastal Systems Center
Panama City FL 32407

Commander
Naval Electronic Systems Com
Headquarters
Washington DC 20360

Commanding Officer
Naval Environmental Prediction
Research Facility
Monterey CA 93940

Commander
Naval Facilities Eng Command
Headquarters
200 Stovall Street
Alexandria VA 22332

Commanding Officer
Naval Ocean R & D Activity
ATTN: Codes 100/111/112
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 113
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 125L
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 125ED
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 110
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 105
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 115
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 200
NSTL MS 39529

Commanding Officer
Naval Ocean R & D Activity
ATTN: Code 300
NSTL MS 39529

Commanding Officer
Naval Research Laboratory
Washington DC 20375

Commander
Naval Oceanography Command
NSTL MS 39529

Commanding Officer
Fleet Numerical Ocean Cen
Monterey CA 93940

Commanding Officer
Naval Oceanographic Office
NSTL MS 39522

Commander
Naval Ocean Systems Center
San Diego CA 92152

Commanding Officer
ONR Branch Office LONDON
Box 39
FPO New York 09510

Officer in Charge
Office of Naval Research
Detachment, Pasadena
1030 E. Green Street
Pasadena CA 91106

Commander
Naval Sea System Command
Headquarters
Washington DC 20362

Commander
DW Taylor Naval Ship R&D Cen
Bethesda MD 20084

Commander
Naval Surface Weapons Center
Dahlgren VA 22448

Commanding Officer
Naval Underwater Systems Center
ATTN: NEW LONDON LAB
Newport RI 02841

Superintendent
Naval Postgraduate School
Monterey CA 93940

Project Manager
ASW Systems Project (PM-4)
Department of the Navy
Washington DC 20360

Department of the Navy
Deputy Chief of Naval Material
for Laboratories
Rm 866 Crystal Plaza Five
Washington DC 20360

Officer in Charge
Naval Underwater Sys Cen Det
New London Laboratory
New London CT 06320

Defense Technical Info Cen
Cameron Station
Alexandria VA 22314

Director
Chief of Naval Research
ONR Code 420
NSTL MS 39529

Director, Liaison Office
Naval Ocean R & D Activity
800 N. Quincy Street
Ballston Tower #1
Arlington VA 22217

Department of the Navy
Office of Naval Research
ATTN: Code 102
800 N. Quincy Street
Arlington VA 22217

Director
Woods Hole Oceanographic Inst
86-96 Water St.
Woods Hole MA 02543

Director
University of California
Scripps Institute of Oceanography
P. O. Box 6049
San Diego CA 92106

Working Collection
Texas A & M University
Department of Oceanography
College Station TX 77843

Director
Defense Mapping Agency
Washington, DC 20305

Director
Defense Mapping Agency
Hydrographic/Topographic Cen
6500 Brooke Lane
Washington, DC 20315

Director
Defense Mapping Agency
Aerospace Cen
St. Louis Air Force Station, MO 63118

Director
Defense Mapping Agency
Special Program Office of
Exploitation and Modernization
8301 Greensboro Drive, Suite 1100
McLean, VA 22102

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>NORDA Report 73 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Commercial Data Base Management System for Geonames Data Base | | 5. TYPE OF REPORT & PERIOD COVERED<br>Final |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Gail Langran | | 8. CONTRACT OR GRANT NUMBER(s) |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Naval Ocean Research & Development Activity<br>Ocean Science Directorate<br>NSTL, Mississippi 39529 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>P.E.64701B |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Naval Ocean Research & Development Activity<br>Ocean Science Directorate<br>NSTL, Mississippi 39529 | | 12. REPORT DATE<br>December 1984 |
| | | 13. NUMBER OF PAGES<br>21 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

very large data base
geographic names
cartography

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Defense Mapping Agency (DMA), with help of the Naval Ocean Research and Development Activity (NORDA), Code 351, is conducting a large effort to automate production of maps, gazetteers, charts, and similar products. One element of this effort is the creation of a single, controlled source for the placenames used in these products: the Geonames Data Base. Because of its size (up to $10^{10}$ bytes) and user requirements (up to 25 simultaneous users), the Geonames Data Base belongs to a special subset of current data base management practice, the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Very Large Data Base.

From requirements generated earlier by DMA, NORDA, and others, this task surveys commercial data base management systems (DBMS) to determine candidates for the Geonames DBMS. These systems are then ordered by objective and subjective measures of performance to help in selecting the Geonames DBMS. Among the systems which may satisfy the requirements of the Geonames Data Base, two are used widely today and are exceptionally well-regarded by current users. These systems are IDMS/R--Cullinet and ADABAS--Software AG of North America.

END

FILMED

3-85

DTIC