Image Cover Sheet

TITLE

A MATCHED FIELD PROCESSING ENVIRONMENTAL DATABASE: DATABASE VS FLAT FILE IMPLEMENTATION

National Defence
Research and Development Branch

Défense nationale
Bureau de recherche et développement

DREA CR/96/404SPK

A MATCHED FIELD PROCESSING ENVIRONMENTAL DATABASE:

DATABASE vs FLAT FILE IMPLEMENTATION

prepared by
Ron Lake

MacDONALD DETTWILER and ASSOCIATES Limited
13800 Commerce Parkway
Richmond, British Columbia, Canada
V6V 2J3

Accepted by: Jon Thorliefson
Contract Scientific Authority

September 1995

W8477-2-RK03/01-SV
Contract Number

CONTRACTOR REPORT

Prepared for

Defence Research Establishment Atlantic

Centre de Recherches pour la Défense Atlantique

Canada

Abstract

The solution of acoustic wave propagation in the ocean is a well established component in the toolset for the detection and identification of underwater sound sources. However, in order to carry out such computations, detailed information is required on ocean bathymetry as well as a variety of acoustic properties of the ocean and of the sediments under the ocean bottom. In this report, we describe a prototype Environmental Database (EDB) that has been built to provide this information. Designed around a Relational Database Management System (RDBMS), this EDB is shown to be faster and more efficient than one based on a flat file system. The report concludes with some recommendations for the next phase of the EDB development.

Résumé

La solution de la propagation d'ondes acoustiques dans l'océan représente un composant bien établi des outils pour la détection et l'identification des sources de sons sous-marins. Cependant, l'exécution de tels calculs nécessite des informations détaillées à propos de la bathymétrie de l'océan, ainsi que d'une variété de propriétés acoustiques océaniques et des sédiments au-dessous du fond de l'océan. Dans ce rapport nous décrivons le prototype d'une Base de données environnementales (Environmental Database, donc EDB) créée pour fournir ces informations. Conéue autour d'un Système de gestion de base de données relationnelles (Relational Database Management System, donc RDBMS), cette EDB se montre plus rapide et plus efficace qu'une base de données fondée sur un système de tables à deux dimensions. Le rapport se termine avec quelques recommandations pour la phase suivante du développement de la EDB.

Table of Contents

1.0 Introduction

Solution of acoustic wave propagation in the ocean is a well established tool for the detection and identification of underwater sound sources. Such techniques are employed in both passive and active detection of underwater targets. In order to carry out such computations, detailed information is required on ocean bathymetry as well as a variety of acoustic properties of the ocean and of the sediments under the ocean bottom.

The solution of the acoustic wave equation is carried out by a computer program termed the acoustic modeller. Each such program contains and solves a particular form of the wave equation. One form might be a one-dimensional simplification of the full three-dimensional equation. Another might employ a normal mode simplification to reduce the partial differential equations to a system of ordinary differential equations.

Each acoustic modeller requires information on the acoustic parameters of the ocean and the sediments below the sea bottom. This information is provided to the acoustic modeller by the Environmental Database (EDB).

The logical relationship between the acoustic modeller and the EDB is simply that the latter provides various functions which yield acoustic parameters as a function of position etc. within the ocean and beneath the sea bottom. This relationship is shown in Figure 1.

2.0 Sources of the Environmental Data

Data supplied to the acoustic modeller may be synthesize from measurements acquired by a variety of techniques and using a variety of types of instrumentation. Gridded data for bathymetry, for example, is available on a relatively coarse grid for most of the Arctic region. This data is usually augmented by higher resolution data which is available only in limited areas. Such data may be obtained from random soundings, from measurements along ship tracks or by airborne lidar.

There is a great variety of accuracy and precision associated with these different types of measurements. It will thus not always be appropriate to include all of the data in a given model solution, nor to include all of the different elements with the same weight.

Depending on their purpose, and depending on the particular type of equation being solved, the model may wish to exclude certain measurements from the solution process. For this reason it must be possible to select more or less arbitrary subsets of the available measurements based on spatial and other constraints.

We can express these requirements in a general form as follows:

acoustic property = funct(location constraint, measurement constraint)

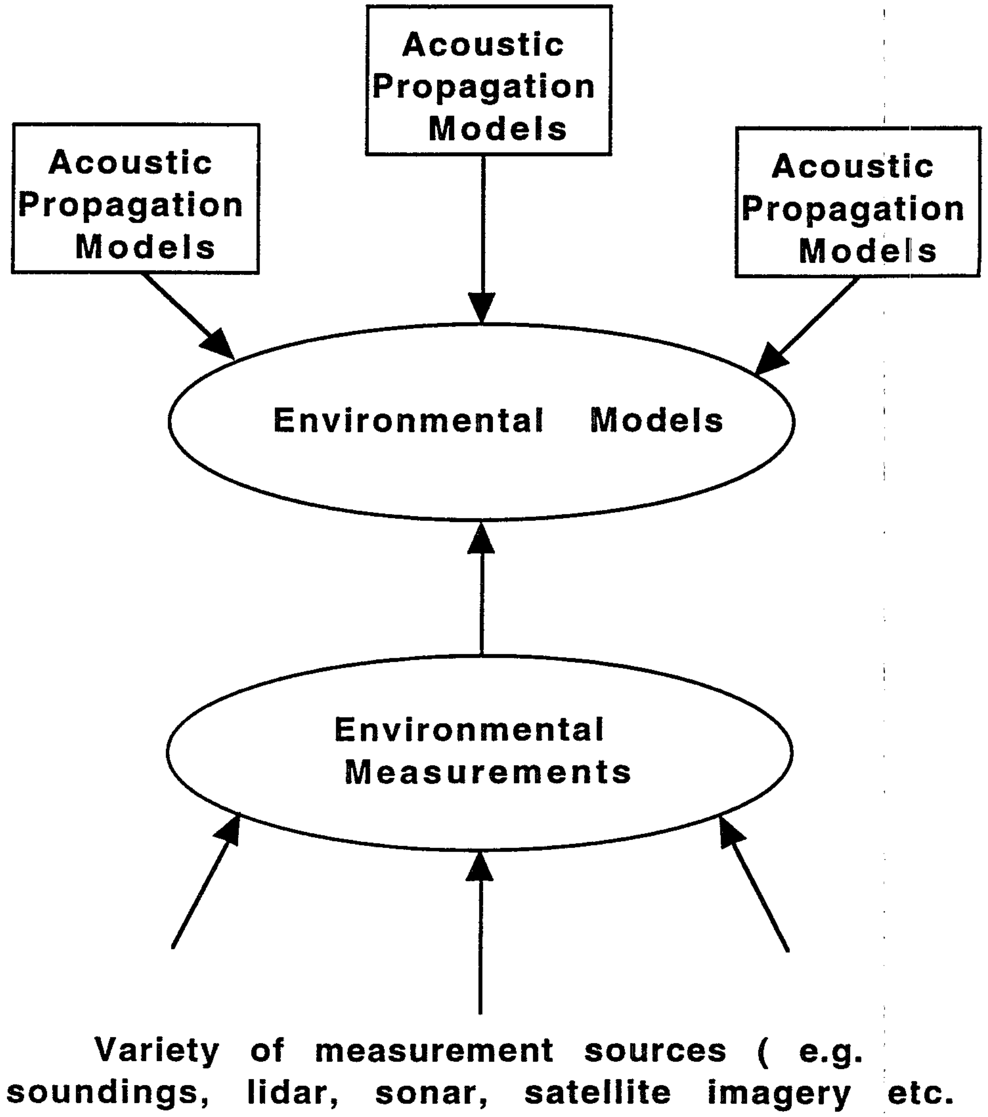

Figure 1. Acoustic modellers and the Environmental Database (EDB).

3.0 Spatial Searching

To return data satisfying the location constraint with acceptable performance requires that an efficient indexing technique be employed. Such techniques typically rely on some form of space partitioning. Spatial constraints are then examined using tree pruning techniques. This results in one or two orders of magnitude increase in speed relative to sequential application of the location constraint.

In Call-up #6, a point k-d tree was used as the spatial index. This tree divides the point space recursively based on the distribution of point objects within the space. More subdivisions are required where there are more points. The process is shown in Figure 2.

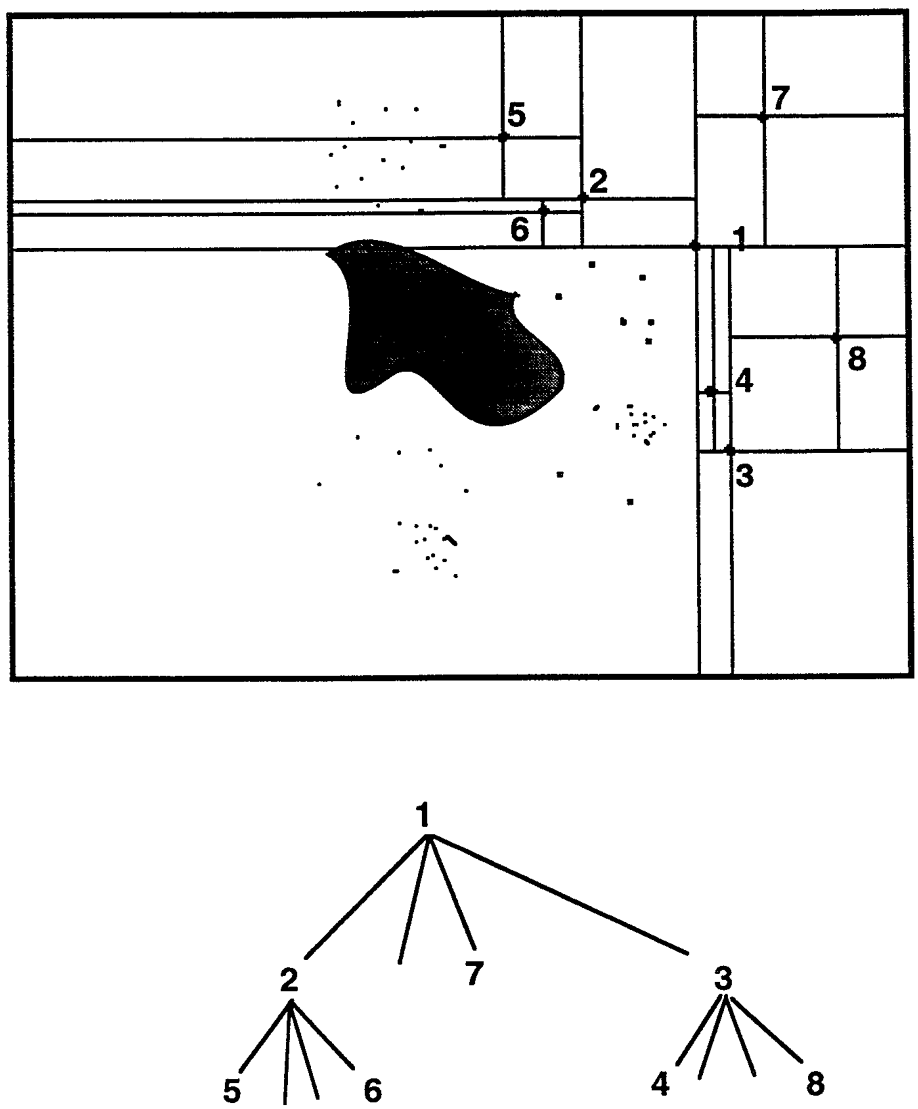

Figure 2. Spatial searching with a k-d tree.

Subdivisions of the space are arranged into a tree structure also shown in Figure 2. Spatial constraints (e.g. "find points nearest to this line") are examined by searching this tree structure and applying the exact spatial criteria only to the points that are candidates after pruning the tree.

Given a location constraint (e.g. nearest neighbour), the spatial index can provide an efficient means of selecting the points which satisfy this constraint. It returns simply a list of point identifiers (PIDs or Point IDs) and the point coordinates which satisfy the location constraint.

4.0 Metadata - Describing the Measurements

Measurements of acoustic and geometric properties assigns values of the measured quantity to a point beneath the surface of the ocean, or beneath the sea bed. The location of such points can also be considered another aspect of the measurement thus indirectly establishing a link between a location in the sea and specific values of acoustic parameters. Description of the non-locational metadata will require some form of 'data model' be developed for the measurement data. To date this has not been done. Such a model would describe measurements in terms of such parameters as acquisition date, acquisition time, instrument type, calibration date etc. In addition to such generic parameters it is also quite possible that instrument or measurement specific type parameters will be required.

Given a metadata model we will then be able to make data requests based on location (via the spatial index) and these metadata attributes. We might for example ask for all bathymetry measurements within a selected region which were obtained after June 15th, 1995, and which were obtained via airborne lidar.

5.0 System Architectures Compared

5.1 Flat File System

In a flat file system we would employ an architecture as shown in Figure 3. Note that the spatial index provides a mapping from a location constraint to a list of PIDs. These PIDs are then used to map through a file header to determine the measured quantities and the associated measurement metadata.

The measured data files could be stored using a file index structure as shown in Figure 4. This structure has a single file for each type of measured value. Records in this file, indexed by PID, contain the measured value and a second pointer to additional metadata. Possible data structures are shown in Figure 4.

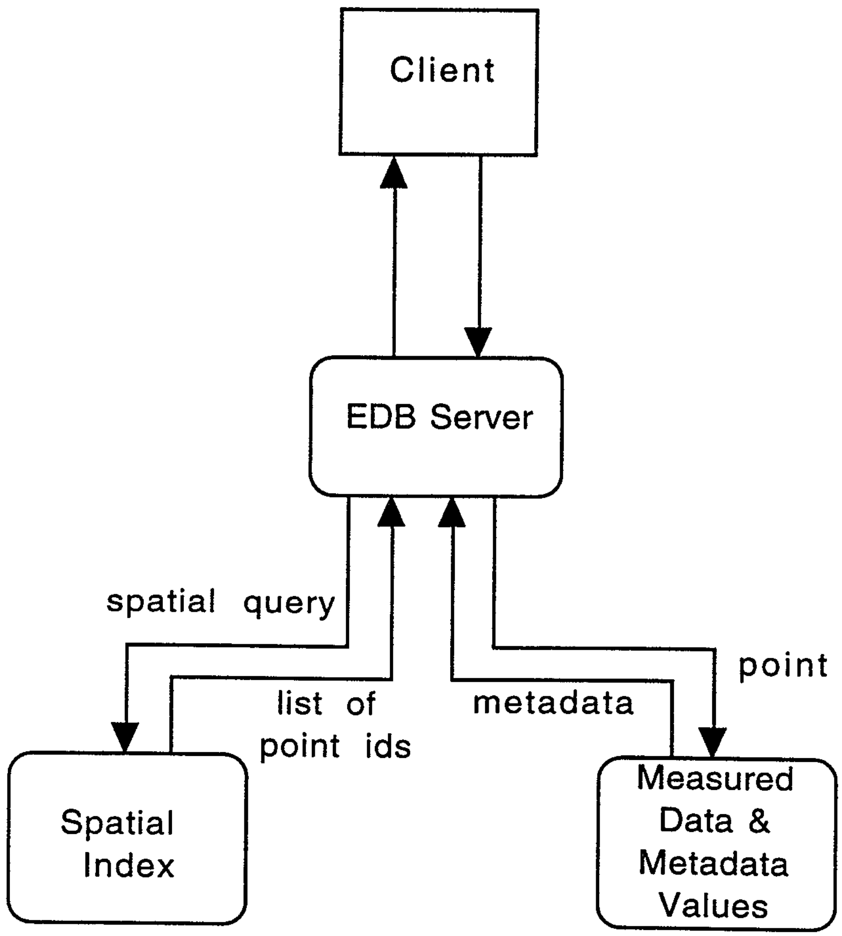

Figure 3. Flat file system architecture.

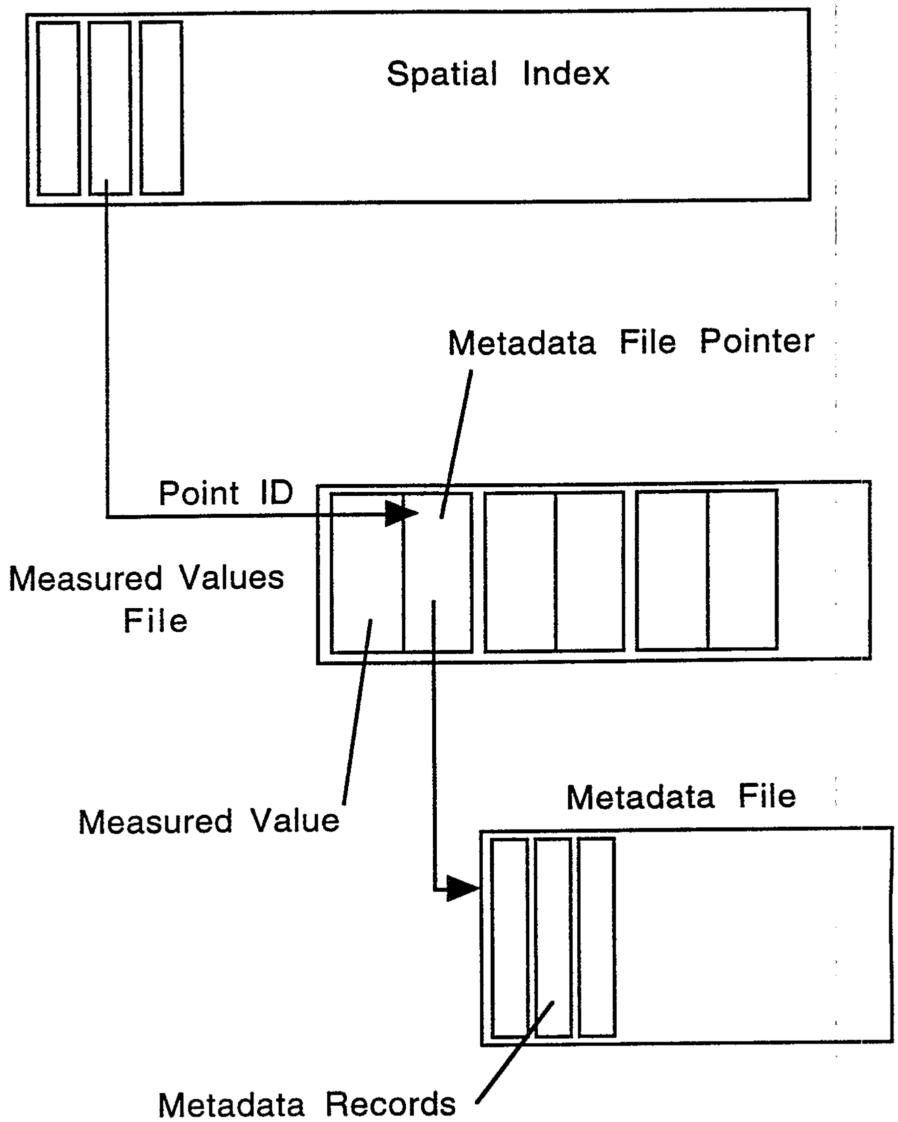

Figure 4. Flat file system file structures.

5.2 Using a Relational Database

Current relational database products do not support the search and retrieval of spatial data. Such databases employ one-dimensional indexing techniques only and two- or more-dimensional searching is required in the case of spatial data. To use a relational database it is thus necessary to implement the spatial indexing in some fashion external to the database. This leads to an architecture as shown in Figure 5.

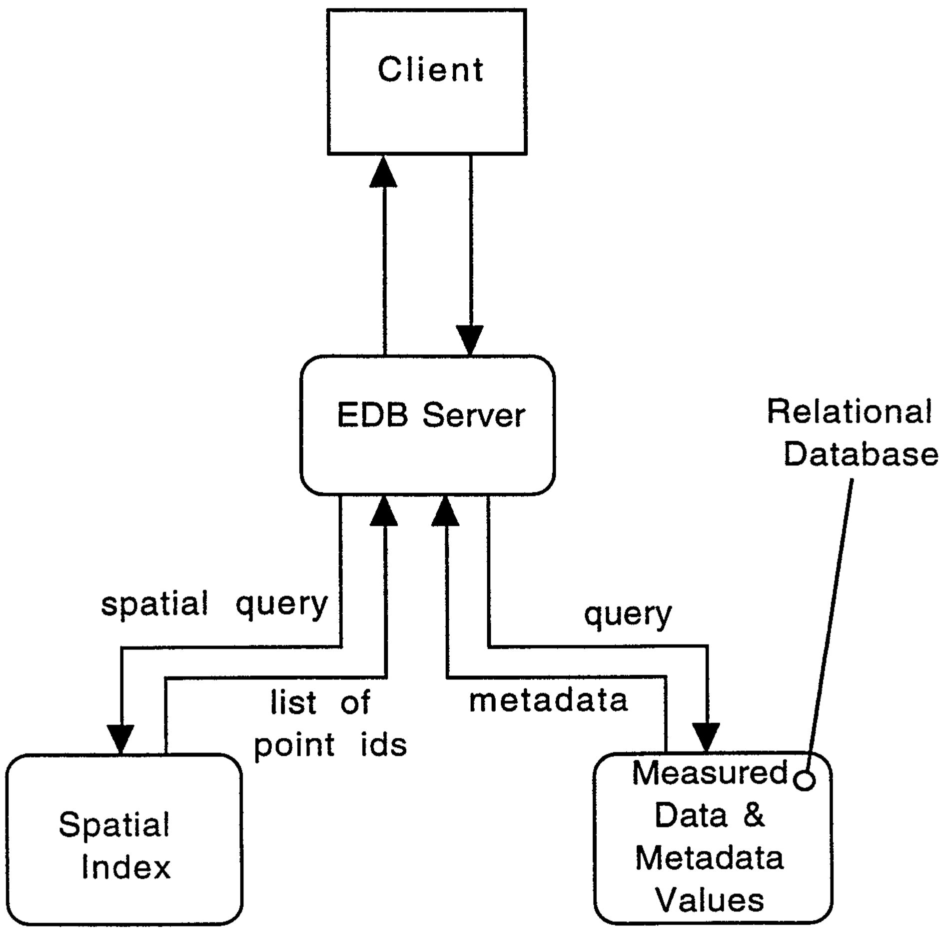

Figure 5. Relational database based architecture.

Note that at a high level this is very similar to the architecture shown in Figure 4. A significant difference is that the PID list returned from the spatial index is used in conjunction with an attribute (metadata) qualifying query. The PID list is then used to match records retrieved from the relational database which satisfy the attribute constraint.

The table structures which correspond to Figure 5 containing the point value data and the metadata are shown in Figure 6.

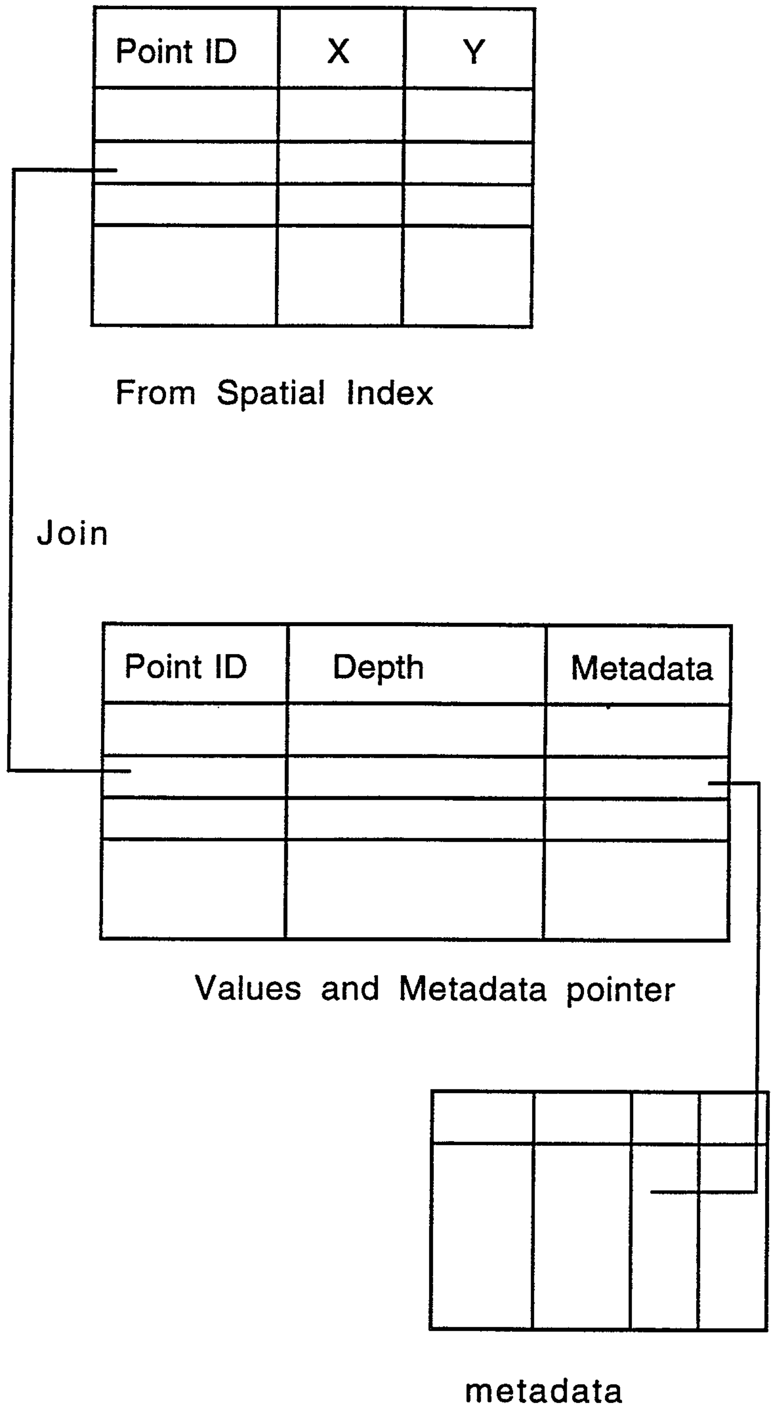

Figure 6. Tables for relational database.

5.3 Comparison of the Relational and Flat File Approach

This section provides a discussion of the differences in performance and flexibility between the architecture based on flat files and that based on a relational database.

Since both architectures employ the same spatial index we can assume that performance and functionality respecting spatial only searches are more or less identical. The real differences appear in the case of queries which provide both spatial and attribute constraints. The latter arises quite naturally in association with constraints on measurement metadata such as the instrument type, data resolution, and various data quality indexes. For example, we might wish to do a run using only the most recent data. This might require a query like: (expressed in pseudo Structure Query Language)

```
select x,y,depth from measured_data
where window.encloses(x,y) and
where date > 10/02/96;
```

In the case of the flat file system, we would have to write a piece of code which explicitly conducts the search. Assuming that the set of points from the spatial search is given as a file, loctbl, we would need a code fragment like: (PID = Point ID, fields in the file are read via filename.value("ValueNameString")

```
for i = 1 to loctbl.noRecords()
    PID = loctbl.value("PID");
    for j = 1 to datatbl.noRecords()
        if datatbl.value("PID") = PID
        then
            if datatbl.value("date") > 10/02/96
            then
                writeRecord(PID, valtbl.value());
                writeRecord(PID, datatbl.value());
            endif
        endif
    endfor
endfor
```

Clearly such an exhaustive search is not very efficient. More efficient searches could be achieved by building an index on the PID for each of datatbl and loctbl. Code would then have to be written to construct this index after each spatial search construction of loctbl. This would provide a reasonably fast means of retrieving the points satisfying the metadata constraint. With more complex metadata we would need multiple metadata files, in which case the datatbl would also contain a file pointer.

In some cases it will be more efficient to do the attribute portion of the query in advance of the spatial part. In the above case, for example, there may be comparatively few recent data points, in which case conducting the attribute part of the search first would be much more efficient. To accommodate this alternate type of search in the file based system would require a complete new piece of code. In addition we would need to conduct either exhaustive searches or build indexes for each new parameter which is selected as the basis of the attribute constraint (date in the above example). Efficient searches can be achieved in each case, but only at the expense of additional coding for each type of search.

The relational database case is decidedly simpler, although otherwise similar in appearance to the flat file case. We can conduct a spatial search using the spatial index and return a table of PIDs. These we simply join with the table of values and any additional metadata tables, e.g.:

```
select x,y,depth from loctbl,valtbl,datatbl
where loctbl.PID = valtbl.PID and
where loctbl.PID = datatbl.PID and
where datatbl.date > 10/10/96
```

Construction of indices on PID for the tables is usually handled automatically by the Database Management System (DBMS). Using an INSERT query these results can be saved directly into a database table. Note that we can readily do the search attribute query first, then build the spatial index on the resultant answer table and complete the query by an additional join. This would go something like:

```
insert into pointtbl values(
        select PID,x,y,depth from datatbl,valtbl
        where datatbl.date > 10/10/96 and
        where valtbl.PID = datatbl.PID)
```

Load the spatial index from the result table pointtbl. Execute the spatial query. This yields the desired result.

It should be clear from these few examples that use of the relational database provides considerably greater flexibility than can be achieved with a file based system. Furthermore, this flexibility is achieved with relatively little impact on the code base. This difference becomes much greater as the amount and complexity of the metadata describing the measured data values increases.

The relational database approach also makes it easy to sort and rearrange datasets prior to execution of an analytical model. We might, for example, have raw measurement values along a uniform grid and at a series of randomly distributed points. On the other hand, we may, for some situations, wish to evaluate the acoustic model along radial lines (bearing, azimuth) from some selected point. Using the EDB we could extract the appropriate subset of the raw data points and then interpolate to create a table indexed by range and azimuth. This might, for example, have the schema:

```
data( PID, azimuth, range, depth, tag_pointer)
```

This data can then be very efficiently read in sequential order along a particular azimuth value.

6.0 Next Steps

Call-up #6 demonstrated a prototype system for the retrieval of environmental information for acoustic simulation. This prototype system employed a memory resident spatial index, together with a relational database. This prototype system demonstrated that such an approach is a viable means of providing information to the acoustic modeller program. This section examines possible next steps in the program.

6.1 Requirements Analysis Review

It is strongly suggested that the requirements for the EDB, especially with respect to the acoustic modeller interface be reviewed. The current interface is unclear in a number of areas, most especially with respect to the anticipated computational modes of the modeller. This review should examine in particular:

1. The dimensionally of interpolation required (one-dimensional, two-dimensional, etc.),
2. The need for true on-the-fly interpolation (i.e. variable step size),
3. Performance requirements (time, size), and,
4. Metadata requirements - anticipated query constraints.

6.2 Evaluation of Commercial Database Products

At the time that Call-up #6 was initiated, no suitable spatial database products were available in the commercial marketplace. To some extent this has changed. The following products are either now in the market, or will be in the marketplace within the next four months:

Oracle Multi-Dimension,
Spatial Database Engine (ESRI), and,
Illustra Lt.

1. Oracle Multi-Dimension.

This product or near product was developed as a joint project between the Canadian Hydrographic Service and Oracle's Hull PQ research department. The project provides a spatial index type, which they call HH Code, which can be used as the type of an attribute in a relational database. This allows spatial searches for points to be conducted within the Oracle 7.1 relational database. The actual order/shipping status was unknown at the time of this writing as was the pricing and platforms supported.

2. Spatial Database Engine (ESRI):

This is the former Spatial Engine of GTI (Geographic Technologies Ltd.). This is spatial server layered on a relational database, more or less in the conventional geo-relational model. It does, however, achieve fast retrieval of spatial information through use of a spatial index, use of a simplified spatial model, and intelligent buffering between the client and server. This product is anticipated to be shipping within the next four months.

3. Illustra Ltd.

Illustra is a next generation, extended relational database, which provides add-in components for spatial information. Illustra calls these add-ins DataBlades and provides DataBlades for two-dimensional and three-dimensional spatial data. Illustra is low cost relative to the ESRI Spatial Database Engine and Oracle Multi-Dimension.

In order to determine which, if any, of these products should be considered for the next phase of the EDB implementation, it would be necessary to:

1. Complete the requirements analysis discussed in Section 6.1,
2. Obtain further literature and evaluation materials respecting these products, and,
3. Where possible, obtain demonstration copies and experiment with the prototype dataset.

This evaluation would likely lead to the selection of one of these products and a corresponding modification of the architecture employed in Call-up #6.

7.0 Summary

In this report, we have described a prototype Environmental Database (EDB) that was built to provide acoustic propagation modellers with information on ocean bathymetry as well as a variety of acoustic properties of the ocean and of the sediments under the ocean bottom. To ensure acceptable performance, it was designed with an efficient spatial indexing scheme. Centered around a Relational Database Management System (RDBMS), this EDB was shown to be faster and more efficient than one based on a flat file system. For the next phase of the database development, this report suggested that the requirements for the EDB be reviewed, especially with respect to the acoustic modeller interface. In addition, it recommended the evaluation of some newer RDBMS products which could possibly enhance the overall performance of the EDB.

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM
(highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)
MacDonald, Dettwiler and Associates Limited
13800 Commerce Parkway, Richmond, B.C. , V6V 2J3

2. SECURITY CLASSIFICATION (Overall security of the document including special warning terms if applicable.)
UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title.)
A Matched Field Processing Environmental Database:
Database vs Flat File Implementation

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)
LAKE, Ron

5. DATE OF PUBLICATION (Month and year of publication of document.)
September 1995

6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.)
17

6b. NO. OF REFS. (Total cited in document.)
0

6. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)
Contractor Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the reseach and development. include the address.)
Defence Research Establishment Atlantic/EsquimaltDefence Research Detachment
P.O. Box 1012, Dartmouth, N.S. B2Y 3Z7

9a. PROJECT OR GRANT NUMBER (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)
Project Spinnaker

9b. CONTRACT NUMBER (If appropriate, the applicable number under which the document was written.)
W8477-2-RK03/01SV

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number mrst be unique to this document.)

10b. OTHER DOCUMENT NUMBERS (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DREA Conractor Report 96/404SPK

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; futher distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic annoucement of this document. This will normally correspond to the Document Availability (11). However, where futher distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)
Unlimited Announcement

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM

DCDO3 2/06/87

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

The solution of acoustic wave propagation in the ocean is a well established tool for the detection and identification of underwater sound sources. However, in order to carry out such computations, detailed information is required on ocean bathymetry as well as a variety of acoustic properties of the ocean and of the sediments under the ocean bottom. In this report, we describe a prototype Environmental Database (EDB) that has been built to provide this information. Designed around a Relational Database Management System (RDBMS), this EDB is shown to be faster and more efficient than one based on a flat file system. The report concludes with some recommendations for the next phase of the EDB development.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title).

Descriptors

Environmental Database
Relational Database Management System (RDBMS)
Underwater Acoustics

Keywords

Spatial Indexing
Metadata
Flat File System

500694

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM