BDGEOS : The geophysical data base at SHOM (Service Hydrographique et Océanographique de la Marine)

M F Le Quentrec-Lalancette, Vincent Ribaud

BDGEOS : The geophysical data base at SHOM (Service Hydrographique et Océanographique ·de la Marine)

M.F. Le Quentrec-Lalancette
Section Géodésie-Géophysique
Etablissement Principal du Service Hydrographique
et Océanographique de la Marine
29275 Brest Cedex, France
e-mail : mflqc@shom.fr

V. Ribaud
SYSECA
29200 Brest, France
e-mail : ribaud@syseca.fr

Abstract - **A Geophysical Data Base (BDGEOS) has been developped at EPSHOM to come up to the military need in mapping the oceanic parameters. Data are measurements from geophysical surveys or are interpreted and modelled values. The BDGEOS runs efficiently to product maps of geophysical parameters, to validate data and to model behaviours as gravimetric markers of seamount or magnetic tracks of wrecks. In the future, it would be integrated in the Geographical Information System of SHOM.**

Abstract - **Une Base de Données Géophysiques (BDGEOS) a été développé à l'EPSHOM pour répondre aux besoins militaires en cartographie des paramètres océaniques. Les données proviennent directement des levés géophysiques ou sont des résultats de modélisation. La BDGEOS permet de manière efficace de produire des cartes géophysiques, de valider des données et d'effectuer des modélisations (detection de mont sous-marin, recherche d'épaves, ...). La base de données sera intégrée au Système d'Information Géographique du SHOM.**

I. INTRODUCTION

The Geophysical Data Base of the SHOM (BDGEOS) has been created because of the necessity to integrate a large number of spatially distributed geophysical data of various origin, to elaborate and to produce synthetic maps for the French Navy. The BDGEOS represents the state of the knowledge of geophysical data referenced in the time.

The BDGEOS is developped with the RDBMS ORACLE (V.7.0) in order to communicate with the other dedicated data bases of the EPSHOM.

II. THE DATA

The data are potential field data, i.e. derived from the earth magnetic or gravity field (declination, free air anomalies, geoid height...). They are issued from measurements or from numerical computations. The platform of measurements can be the satellites (altimetry), the hydrographic vessels (gravity and magnetic data) or terrestrial (stations). Two groups are distinguished : the marine data (main source of informations), and a generic one, named model.

A. Marine Data

The marine data comes from external laboratories and international data bases (National Geophysical Data Center, International Gravimetric Office..) or are acquired by the french hydrographic vessels of the SHOM.

Description of a typical geophysical survey

The pattern of the acquisition off shore is often made of four main divisions related to geographical and temporal constraints. An oceanic area (survey) is defined for a special study (geographical limits), see Fig. 1. Depending on the disponibility of the vessels in the time, some sub-survey areas are defined. Different geophysical measurements are acquired along tracks. The precise instructions about the way of acquiring (spacing of the tracks, timing of measurements, ...) depend of the main goals of the survey.
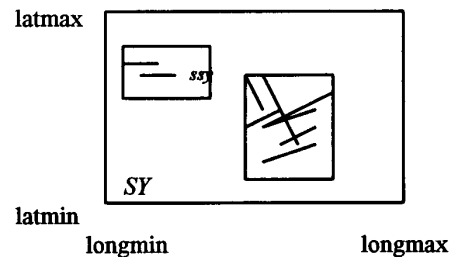
Figure 1: Description of a survey

_____ : track line
SY   : map of the survey area
ssy   : map of the sub-survey area

Geophysical parameters acquired at sea are : the intensity of the earth magnetic field and of the gravity field

The anomalies in relation to suitable models are then computed depending on the object of the survey.

The informations on these data are divided into global targets identified by their geographical and temporal properties :

the survey, sub-survey, tracks and measurements.

In external data, the sub-survey division does not always exist : in this case, sub-survey is supposed to be identical to the survey area.

B. The models

The models represent some geophysical values computed for three main reasons :
- the model is the result of some mathematical transformation of the marine data (derivation, upward or downward continuation, ...);
- the model is a mix of data from various platform of measurement. For instance, in oceanic area it is sometimes useful to put together, gravity anomalies of a marine survey and those derived from an altimetric measurement with geodesic satellites;
- the model is a numerical map for integrating cartographic or computing tools.

The models are described in the same way as the marine survey. Punctual elements are scattered in a geographical area.

III. STRUCTURE OF THE DATA

The subdivision of the used data, induces the definition of two data types : the 'SURVEY' data type and the 'MODEL' data type (Fig. 2).
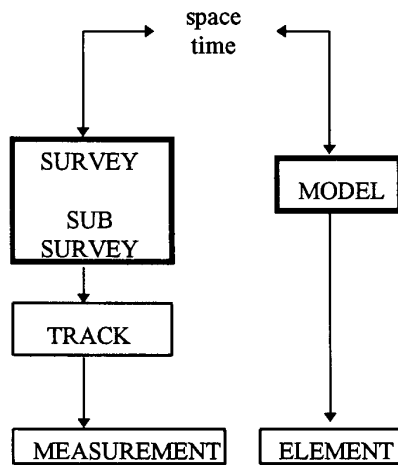

Fig. 2: The structure of the data

The 'SURVEY' data type is related to the 'SUB-SURVEY', 'TRACK' and 'MEASUREMENT' properties to restore the steps of the acquisition off shore. The

'MEASUREMENT' is in fact represented by the 'MAGNETISM' and 'GRAVIMETRY' properties.

Behind each properties, suitable informations are grouped at each step of the hierarchical tree. The position or time of a data is placed into the 'MAGNETISM' or 'GRAVIMETRY' boxes depending of its nature, but the geographical or temporal limits are in the superior boxes.

The 'MODEL' data type is related to the elements of the models simply called 'ELEMENT'. It has been viewed to integrate the already used models and to allow the future developments (see Table 1). Depending on the nature of geophysical models, some set of variables define groups as ANOMALY which can be of different nature or VECT-M which can obviously depend on magnetic data.

TABLE 1
DESCRIPTION OF MODELS

NATURE	GROUP		
GRAVIMETRY ALTIMETRY MAGNETISM	ANOMALY	GRADIENT	MISCEL
GRAVIMETRY ALTIMETRY	GEOIDE	MAPS	VECT-G
MAGNETISM	VECT-M		

VECT-M : magnetic vector
VECT-G : gravity vector
MISCEL : miscellaneous
GEOIDE: geoide vector
GRADIENT: first derivatives
MAPS: gravity model of the SHOM
ANOMALY: value computed from measurement with respect to a model

IV. THE BDGEOS SOFTWARE

A. Overview

Data base end-users interact with the geophysical data base through the BDGEOS software.

The data base is relational. This means that the basic information is structured and stored in tables.
A table can be viewed as a set of similar objects, each of them described by a collection of properties (e.g. each survey is a particular object, described by a date, geographical coordinates and so on; and the table SURVEY is the set of all the particular surveys).

The software achieves five main goals :
- confidentiality,
- loading,
- updating,
- querying and reporting,
- extracting.

I-640

B. Confidentiality

The confidentiality underlies the access to the data. Each object i.e. each row of each table owns a property called OBJ_PROTEC (for Object Protection). The value of this property is an integer (currently from 1 to 4) representing the increasing level of object confidentiality.

Each user of the BDGEOS software is associated with a level of authorisation, which represents the maximum level of the data accessible to this particular user.

The BDGEOS can be seen as the superposition of several layers, each level corresponding to a level of confidentiality.

The BDGEOS is intended primarily for non-restricted users, and the choice of implementation priviledges this type of users.

So, except for non-restricted users, the access to each object is controlled by a test on the OBJ_PROTEC property. Performance in this bottleneck test is strongly related to the physical implementation scheme.

As it will be discuss in section E, table clustering is not suited to BDGEOS's structure.

The opportunity occurs of including OBJ_PROTEC heading or following each index, but performance's evaluation proved it to be inefficient.

C. Loading

The BDGEOS has been designed to load data supported by ASCII files under multiple formats.

The data loading is accomplished in two steps : the preparation and the execution.

The preparation step is used by the operator to describe the information used for several tasks such as controlling the files to load or specifying missing and default values.

The preparation step is a tedious and detailed task. To help the operator, the preparation is based on the use of notes.

A note is a collection of the informations used for loading. Note management is similar to electronic document management (new, open, save, delete, ...).

Note filling is done through forms as is most human machine interface. However, information to fill in the note is dependent on the file's format and on the BDGEOS's structure and this could result in multiple different forms.

Rather than multiply individualised forms, the human machine interface use a single meta-form, based on the

BDGEOS's metabase. The metabase collects the descriptions of the data base's structure and of the file's format. These descriptions are used to build dynamically a form well-suited to a particular loading.

The execution step is triggered off by the preparation step. Either the execution can run immediately, or in the background or in batch at a desired moment.

D. Updating

Updating data in the BDGEOS is of two types :
- update of particular objects in a conversiatonal mode,
- update of a set of objects.

Updating particular objects is done by the use of dedicated form, one for each type of object. These forms perform classical functions of data base management such as querying, inserting, deleting and updating. Although objects are visualized one at a time, the underlying power of SQL allows the simultaneous update of several objects by use of predicates.

A set of objects can be updated from a file in four ways :
- inserting new objects in a particular hierarchical object,
- substituting values of some properties of designated objects,
- superseding null values of some properties of designated objects,
- updating a geographical zone.
Sets of objects are designated according to spatio-temporal or other criteria.

As for to the loading, updating a set of objects is acomplished in two steps, preparation using a note and execution.

Each property of each object of the BDGEOS can be updated from various file format. Update notes are dependent on these formats and this magnifies the need for dynamically building the forms.

The execution step is then triggered off in one of the three ways described : immediately, in background or in batch.

E. Querying and reporting

Queries are of two types :
- usual, ready-to-use queries such as key criteria, geographical or temporal retrieval;
- unusual, complex queries based on the use of variables and expressions.

Queryng the data is organised in hierarchical levels. Each level is associated with an object type, and data can be accessed at every level. Once a level has been chosen, query criteria are entered by the operator and used to retrieve objects. Excepted for leaf objects, it is then possible to go down in the hierarchy by selecting one of the objects retrieved and entering new criteria.

Each level can be used to query objects along a transversal path rather than a hierarchical path.

Each query's results can be displayed on the monitor, printed on an output device, or stored in a file.

The hierarchy plays an important role in the querying's processus and hierarchy's implementation affects performances of the system.
There are two main directions of implementation : clustering and indexing.
Clusters group objects sharing common columns values and which are often used together. It can reduce disk input/output and improve access times for joins. Clustering accelerates some research and can slow others down. The choice of clustered columns has to be done carefully.
Indexing doesn't force physical placement and doesn't penalize queries. An index is merely a fast access path to the data; some queries can benefit from indexes, some not.

In all cases, the use of accelerator such as indexes or cluster improves data retrieval but slow down data update. Loading and updating will be mostly executed in a batch or background manner so this slowing effect should not affect the end-user.

The critical access time concerns the access to objects which are the leaves of the hierarchical tree, so performances are evaluated for these objects on different types of queries.

Performance's evaluation proved than clustering was inefficient for two reasons : the number of rows with the same cluster key is highly variable and the size of set of clustered rows exceeds more than one physical data base block.

F . Extracting

Currently, most of these geophysical data at the SHOM exists on their own as ASCII files and are used as input data by different software.

The BDGEOS is intended to integrate all these data and all the different software should input their data from the base.
Rather than update these software, data can be extracted from the BDGEOS under different formats to meet software input requirements. Some formats are well-known now and the corresponding extractions are stated. However, the perennity of the software depends on its ability to produce data in what are, for the moment, unknown formats.

Extracting data consists of querying the data base and formatting the results. Most extractions are generally reusable, differing only in a few research parameters. But, tuning an extraction needs a good knowledge of data base schema and comprehension of SQL features.
Extraction's design aims to separate complex and reusable operations from simple and unreusable tasks. Thus, data extraction is accomplished in two steps : editing and extracting.

The edition step is reserved to advanced operators and is intended to describe the work in a note.
Note's editing is in fact writing a complex SQL query through the help of the software.

The extraction step is controlled by the use of a note, issued from the edition step. Owing to the note, non-advanced users are guided through software menus and choices. The complex SQL query underlying to the note is thus completed by the criteria needed for a particular extraction. The extraction itself is then triggered off in one of the three ways described.

V. CONCLUSIONS

The BDGEOS has been created with the purpose of standardization and simplification of most used tools (extraction, querying, ...). The human machine interface is user-friendly and compatible with any kind of screen. Nevertheless, it's still a most performing tool (special tasks can be developed).

The relational data base does not allow to exploit the use of geographical constraints. The functions of the SQL language are poor in managing topological objects like the inclusion or intersection of areas. In the future, the BDGEOS will be integrated into a Geographical Information System to resolve problems related to spatial requests.