C. BOUCHER
Institut Géographique National
2, Avenue Pasteur, F–94160 Saint–Mandé, France
K. PODER
Geodetic Institute,
Gamlehave Alle 22, DK–2920 Charlottenlund, Denmark
C.R. SCHWARZ
National Geodetic Survey,
NOAA / NOS, Rockville, Md. 20852, USA
C.C. TSCHERNING
Geodetic Institute,
Gamlehave Alle 22, DK–2920 Charlottenlund, Denmark

INTRODUCTION TO AND REPORT FROM THE SYMPOSIUM ON MANAGEMENT OF GEODETIC DATA, AUGUST, 1981.

Abstract

The Symposium on the Management of Geodetic Data was held in Copenhagen in August, 1981. Many geodesists expect the issues of data management to be the next major problem area in geodesy. This Symposium was the first organized attempt to understand these issues and to share experiences.

The main conclusions reached by the symposium are reported. These conclusions concern topics such as (1) the role and function of a geodetic data base, (2) the use of commercial data base management system, (3) the validation of observation data, and (4) the use of abstract data types.

1. Introduction

Many scientists working in geodesy have developed an interest in data management. This has occurred almost by necessity since many research projects deal with large sets of data : The data must be put into computer readable form ; data obtained from other agencies must be put into a form suitable for one's own computer ; errors must be eliminated ; data selection must be done ; and the results must be presented in tabular or graphical form. In many cases the time spent on data management issues will dominate a project.

However, quite a few scientists have also realized that the time spent on these very time–consuming functions could be substantially decreased if they used (possibly after some modifications) data management tools provided by the computer manufacturer.

In geodetic production (of coordinates or of gravity field information) we know

* – Together the authors formed the symposium program committee.

very well that data management is a major activity : collecting data in the field or from other agencies, validating and adjusting the data, archiving and putting the results at the disposal of our customers. Here much progress has taken place in the past decade due to the effective utilization of computer technology and the development of data bases. However, the resources used on data management are still very large, as indicated in *Table 1.*

Table 1

Resources spent on data management

(Rough estimates)

Area	Main Functions	Manpower %	Funds
Research	Placing data into computer readable form. Getting data from / to other agencies, Elimination of errors, Data selection, Display of results,	25	10
Production	Collecting data in the field or from other agencies, Validating and adjusting data, Archiving, Distribution of results to customers.	50	75

Further progress (i.e., reduction of costs) can be made but this requires research and development. Leaving this research to be done only by specialists in computer science will not be sufficient ; geodesists must play a major role in this development. We are in fact in the same situation as many times before in geodesy when new tools were presented to us. Our task is to find out (1) how to use the tools, (2) which tools to use, (3) in which situations to use the tools, (4) whether we can improve the tools, and (5) which modifications are necessary in order to adapt the tools to a geodetic environment.

In order to initiate cooperation among geodesists when dealing with these questions, Special Study Group (SSG) 4.66 "Management of Geodetic Data" was formed in 1979, and it was decided to organize a symposium on the topic of the study group. The Symposium took place in Copenhagen, Denmark, *August 24 – 26, 1981,* and SSG 4.66 had its first meeting in connection with the Symposium. The participation at the Symposium included 70 geodesists or computer specialists from 20 countries, and 25 papers were presented or distributed. The proceedings have been published by The Danish Geodetic Institute.

The purposes of this paper are to give an introduction to geodetic data management and to give a report from the symposium. In section 2 the main data management concepts are introduced, and we describe the role of data management in modern geodesy. It has only been possible to report on some of the results presented at the symposium, so in section 3 we have concentrated on the function of the geodetic data base and its management system. We have not commented on the special problems facing geodetic organizations in developing countries or the problem of data collection. The interested reader must consult the published proceedings.

2. Main Data Management Concepts

The ***data base*** is the main depository for our data. Within the data base, a single item will exist in only one representation, but it may be used for several different applications. This means that different users may view the data from very different standpoints. For example, one user may want to see only gravity values from a base network, while another user will want to see all data from a given region.

The data base concept has many practical advantages. The most important is that updates or corrections are made in only one place, and automatically become effective for all applications against the data base. This ensures that all applications have a consistent view of the data, and eliminates all those problems that arise from different users processing similar but slightly different data sets. Other advantages come from the ability of a group or agency to concentrate its efforts on a single data set : a variety of access methods can be provided, and reporting programs providing management information can easily be constructed.

The users of the data base are or may be aided by standard programs or subroutines for accessing, retrieving, displaying or changing the data. This software is sometimes put together in what is called a data base management system (DBMS), but this term is only used if a certain minimum of services are provided. The DBMS frequently comprises a so–called data manipulation language, which supports the programs and a query language assisting the occasional or ad–hoc user.

A review of the various DBMS functions is given in Tscherning (1980). The speed with which these functions can be executed and the complexity of tasks such as establishing or changing the data base depends on the data model used. Three main models are in use : the network, the hierarchical and the relational models The three models used for geodetic data are shown in *Figure 1.*

In the network model data may be connected in many different ways. For example, an observation may be accessed through the station number of the standpoint, through the observation type and the time when the observation was carried out, or through the specification of the instrument used. In the hierarchical model data can only be accessed one way, e.g., through one of the ways mentioned for the network model. In the relational model everything is broken down to homogeneous information sets, like the tables we use ordinarily. Also the connections between data elements are described in tables.

Each model has advantages and disadvantages, but all may be useful in geodetic applications.

Data management is frequently and traditionally a supporting function in a production activity (for example, functioning as an inventory management system in the production of automobiles). In geodesy, however, the product *is* information (or its representation as data), so the data base and the DBMS must play an important role. This is illustrated in *Figure 2.*

Note that *Figure 2* incorporates nearly all functions executed by an operational geodetic agency. If we exclude research and development, then the functions of modern geodesy are close to 100 % data management functions. With the advent of satellite

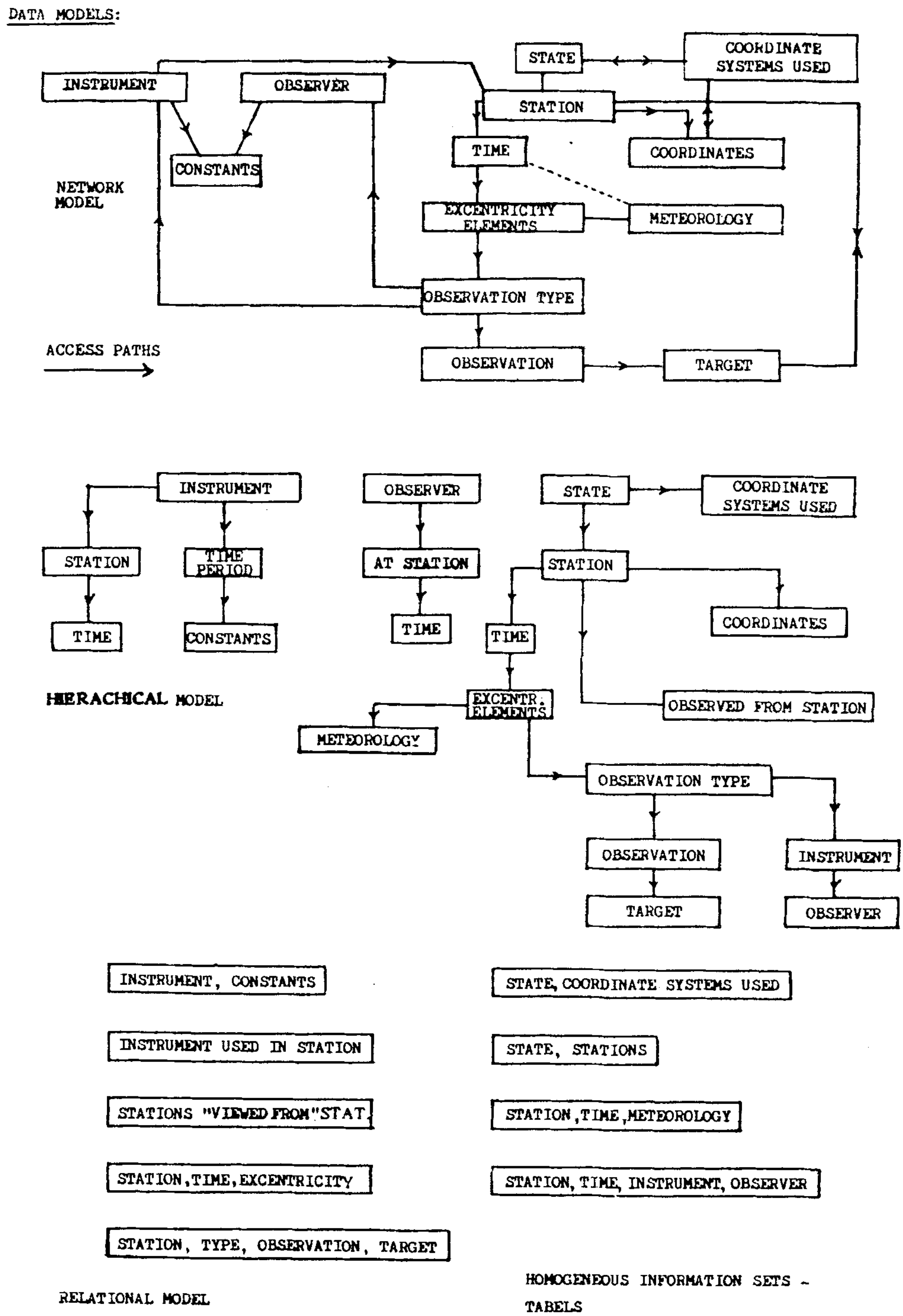

Fig. 1 — The three most important data models used for geodetic data

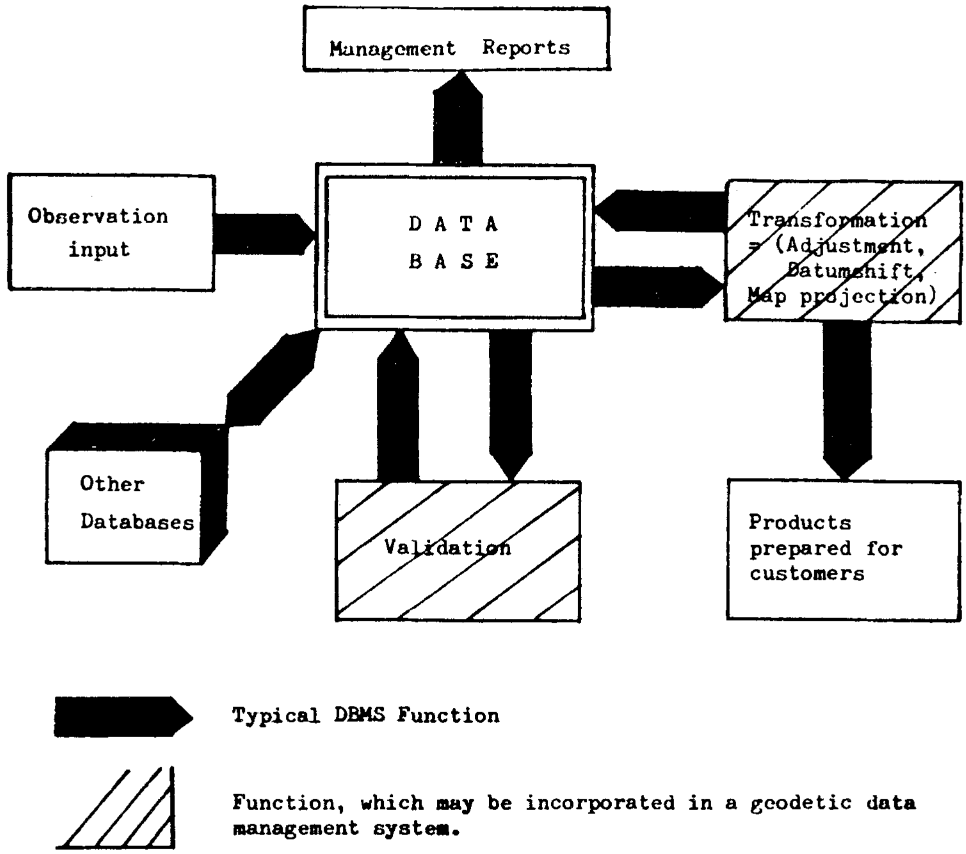

Fig. 2 — The central role of the data base and the functions of the DBMS are illustrated. In many cases it is reasonable to incorporate most data validation and transformation functions in the DBMS.

positioning system like the Global Positioning System, then also functions such as the establishment of station markers and mark maintenance may disappear.

3. The Symposium

The Symposium heard reports on all the elements found in *Figure 2* : Automated data collection and preprocessing, data validation, data base content, data standardization, links to other data bases, and graphical display of results.

Several presentations described existing or planned geodetic data base systems. The use of commercial DBMS packages management of geodetic data did not seem to be widespread, although many agencies are now recognizing the potential advantages of such systems. One of the main potential advantages is that a geodetic organization would not have the burden of maintaining the DBMS software, which generally is very complicated and requires resources in terms of an in—house developed expertise.

Surprisingly, attemps to use DBMS packages which implement the network data model do not seem to have been successful in geodetic organizations. It is not clear why this is so, but it appears to be due more to the actual implementations than to the data model. The most successful efforts using a data base package seem to have been those which use the simple hierarchical data model, but implemented with multiple keyed access. Several papers mentioned the use of the SYSTEM 2000 DBMS, see, e.g., the paper by McConnel, this issue.

Other papers described geodetic data management systems that were constructed without the use of a DBMS. These systems are constructed entirely in algorithmic languages, usually FORTRAN. They often make extensive use of whatever keyed or indexed access method is available on the machine and for the language being used. For example, the Symposium heard about systems using the partitioned access method, the index sequential access method and hash table techniques.

The presentations at the Symposium clearly established that it is possible to have a geodetic data management system without using a commercial DBMS. The systems which were described as already operational generally fell into this class. Several of them had been built up over a number of years beginning at a time when the commercial packages were either not available or did not perform satisfactorily. The Symposium recognized that a new effort beginning now should try to take advantage of the power of a commercial DBMS, but this fact does not justify an effort to convert the older systems.

A number of papers discussed the content of various data bases. Many groups have begun with the idea that the data base should include only data which are thought to be nearly error free. In some cases only a part of the original archival data were put on computerized form. This approach treats the data base as a machine readable archive separated from the data validation process. During the Symposium it became clear that the data base is the proper tool to support data validation. Thus the data base should hold the original, unreduced data, and also contain as far as possible all available data elements. This would permit the DBMS to write reports on the geodetic information which would be useful tools for geodetic and surveying organization management and control decisions.

It was also stressed that the data should be organized so that the most general view of the data would be permitted, making the data base useful also for outside users. In this connection it is important that the data base contain information about the quality of the data, not only in terms of standard deviations of estimated quantities, but also in terms of the number of significant digits. Such information may otherwise be difficult to recover in case a datum transformation has taken place.

The existence of unique station identifiers was stressed throughout. Much effort has been spent in the past to match observational data with the correct station. With proper data base management, the necessity to repeat these matches may be avoided.

The data base also serves as an important tool when validating the data. It aids preserving the completeness and integrity of the data. Adjustment of parts of the (geometrical) data was facilitated by having the possibility for establishing ad hoc data bases, as discussed in a paper by John Isner. It was stressed several times that it must be possible to access the data using a high level programming language like FORTRAN or Algol. Access only through an interactive query language is not sufficient for most geodetic applications.

The validation of gravity data was shown to be much facilitated when graphical software was available for plotting the data and for the drawing of contour maps. These maps were made using part of the data, which then were compared with the rest of data.

Such procedures often made the identification of outliers possible (see the paper by Isaac, this issue).

The use of a data base involves many elements of a new technology. Geodesists must find out which of the various possibilities are most effective when dealing with geodetic data. The possible uses of different access methods and data models were reported at the Symposium. These subjects are rather technical and call for the assistance of experts from the field of computer science. The tools which should make the data base as easy to use as an ordinary hard copy file on tabular form or a map showing the situation of a network needs to be developed. A very extensive query system, described by Dave Alger, contains most of the needed elements, and may serve as an excellent example for other organizations establishing similar systems.

A very special problem area is the solution of large systems of equations arising during the adjustment of geodetic networks. The computational problems here have largely been solved. A major remaining problem is to find efficient methods to get the needed parts of the normal equations into core for processing at the right time. The problem is naturally diminished when a profile reduction is performed, and the Symposium heard reports about the use of one of the popular techniques. A new method for Helmert blocking, which automatically creates the blocks, was described by K. Poder.

The geodetic data base and its DBMS permits us to view our networks (point systems) or gravity field models as a whole. John Isner (see the paper in this issue) here asked whether we should also be able to operate on a whole network or a gravity field model as a single abstract entity. He described a new abstract data type, with its specific operations, which would permit us to do so. If in the future we have access to computer languages supporting a higher degree of abstraction than at present, then we probably also will be able to manage our data in a more consistent way.

4. Conclusion

DBMS's make the implicit assumption that data is almost correct, and when once and for all corrected, then remains correct indefinitely. This clearly is not realistic for geodesy so we must try to work towards systems where errors may occur, but they will only have a finite lifetime. This means that we should use the data base systems as validation tools and not just as the safe harbor for data after validation.

The geodetic community has been reasonably successful in adapting its mathematical algorithms to computer processing. The problems of modifying, extending, and implementing these algorithms have now been studied for 10 – 20 years . We are only now discovering the problems of adapting our data management practices to the use of computers. For many geodesists, data management seems to be the next major problem area. We should not be surprised if it takes another 10 years and several symposia for us to understand and learn how to deal with these problems.

Much progress has been made, and more may be made if the research in data bases and their applications is studied and followed by the geodetic community. We must all be grateful for what has been achieved by some of our colleagues, and ask them to continue their work for the benefit of the whole geodetic community.

o

o o

REFERENCES

C.C. TSCHERNING : "Management of Geodetic Data". Proceedings of Symposium "Automated Processing of Surveying Data", pp. 221–237, Varna, November, 1980.

C.C. TSCHERNING, (Editor) : Proceedings of the International Symposium, Management of Geodetic Data, Kobenhavn 24 – 26 August, 1981. Meddelelse nr. 55, Geodaetisk Institut, 1981.

o

o o

APPENDIX

Papers presented at Symposium Management of Geodetic Data

S.I. AGAJELU : On the Establishment of a Geodetic Data Base in Nigeria.

David E. ALGER : The Design, Development, and Implementation of the National Geodetic Survey Data Base Query System.

Battista BENCIOLINI and Luigi MUSSIO : Test of a reordering Algorithm for Geodetic Network and Photogrammetric Block Adjustment.

C. BOUCHER : Impact of Standardization on the Management of Geodetic Data.

Kjell DEGERSTEDT : Handing of Geodetic and Cartographic data at the National Land Survey of Sweden.

W. EHRNSPERGER and R. KELM : The new Data Base System Established in the German Geodetic Research Institute, Dep. I (DGFI).

F.A. FAJEMIROKUN : Establishing an African Geodetic Data Bank.

Gordon M. FRANK and James P. MORAN : Bathymetric Data Reduction Subsystem (BDRS).

A. FRANK : The Role of Geodetic Data in a Land Information System.

D. FRITSCH and H. WILMES : The use of Microcomputers for Data Acquisition and Data Preprocessing.

J. Rudolf FURY : Data Bank Techniques for the Management of Large–Volume Geodetic and Geophysical Data at the National Geodetic Survey.

L. GRÜNDIG and W. STARK : One–Way Dissection and Sparse Matrix Algorithms : A Strategy for the Adjustment of Large Geodetic Networks.

J.F. ISAAC : Gravimetric Data Management at BGI.

John ISNER : A Programming Methodology Based on Data Abstraction.

John ISNER : The Redesign of the NAD Adjustment System.

Tom JACOBSEN : Trends in the Evolution of Communication.

Horst KREMERS : Data Handling Capabilities of FORTRAN Standards.

Horst KREMERS : Data Management in the Curricula for Geodesists and Surveyors.

Andrade, Bartolo LARA : At the Commencement of the Mexican Geodetic Data Book.

C. LE COCQ and C. BOUCHER : The I.G.N. Geodetic Data Base – Results of the Feasibility Study.

Ken McCONNEL : The Organization and Management of Canada's National Gravity Data Base.

J. Dennis MILBERT : Validation of Horizontal Observations for the National Geodetic Survey Data Base.

K. PODER : Top–down Strategy for Helmert Blocking.

C. POITEVIN : Gravity Data Management : From Field to Publication.

Klaus SCHÜLLER : Graphic Interactive Computer Systems – Design and Application with respect to Geodetic Data and Mapping.

Charles R. SCHWARZ : The Geodetic Requirements for Commercial Data Base Management Systems.

H. Hiram SKAGGS, Jr. : Doppler Geodetic Point Positioning Data Base Used at DMA.

Received : 15.09.1981
Accepted : 22.01.1982