SLAC-PUB-5260
June 1990
A

DATA FLOW INTEGRATION FOR GEODETIC APPLICATIONS*

DETLEV RULAND

Kiel University, Computer Science Department
D-2300 Kiel, West Germany

ROBERT RULAND*

Stanford Linear Accelerator Center,
Stanford University, Stanford, California, 94309

ABSTRACT

Geodetic applications even for a defined project consist of various different activities, and access a vast amount of heterogeneous data. Geodetic activities need a hybrid and heterogeneous hardware environment. This paper gives a brief introduction to the geodetic data flow using a sample application in survey engineering data. It states a general multi-level integration model providing an open system architecture. The model yields the GEOMANAGER project. Its data management aspect is addressed by this paper.

INTRODUCTION

- Preface -

Over the last decade data handling in-applied geodesy and surveying has changed dramatically. What use to be the fieldbook is now a portable computer and the fieldbook keeper has been substituted by a microprocessor and an interface. Further down the data processing line one can see the same changes, least-squares adjustments used to require the computational power of mainframes, now, there is a multitude of sophisticated program systems available which run on Personal Computers (PC) and provide an even more elegant human interface. Also, there are solutions available for the automated data preprocessing, i.e. for the data handling and preparation from the electronic fieldbook to the creation of input files for the least-squares adjustments, /FrPuRRu87/, /RRuFr86/. However, an equally important step has not found much consideration in geodetic discussions and publications, the integration of the geodetic data flow, i.e. the management of geodetic data in large projects. This paper will summarize the geodetic activities and the data flow shown at a sample and representative geodetic application. A two-level integration model is introduced, consisting of communication integration and information integration. Whereas an integrated communication system can be implemented using todays market standard components, an information integration requires a customization of new database management systems. The goals, requirements, and solutions for geodetic database management systems, especially for the GEOMANAGER of our sample application are emphasized.

Contributed to the Scientific Computing and
Automation Conference, Maastricht, The Netherlands, June 12-15, 1990

*Work supported by the Department of Energy, contract DE-AC03-76SF00515

- Data Flow -

The geodetic data flow is summarized in Figure 1. First, the readings are stored in measurement instruments or data collectors. The data then is up-loaded and prepared by DATA PREPARATION programs yielding a measurement data file for each considered observable. These measurement data are raw measurement data, which must be processed by PREPROCESSING programs yielding reduced data. To do so, the preprocessing programs need to access the calibration data. Furthermore, point identifiers are normalized to standard point identifiers, i.e., alias or synonyms are replaced by standard identifiers. The preprocessed measurement data form the input of various DATA ANALYSIS programs, which compute (new) coordinates for the considered points. -For- each point all sets of new and previously measured coordinates are stored. Each set of coordinates refers to a common or measurement specific underlying coordinate system.

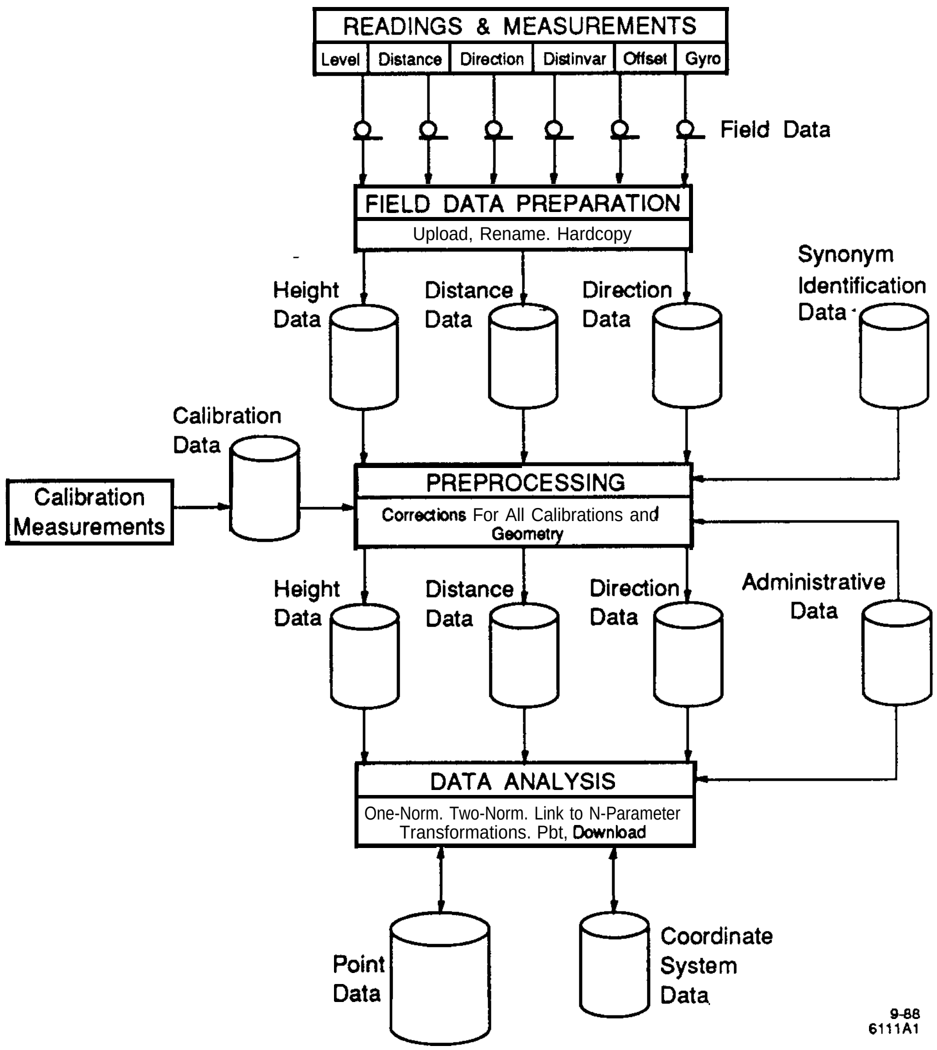

Fig. 1 - Geodetic-Data Flow

Thus, a huge amount of highly structured data is generated and accessed by various activities from data collection in the field to time consuming data analysis programs.

Due to the nature of geodetic applications, the geographical sites of the activities are widely spread. The observables are collected in the field using portable microcomputers (e.g. HP Portable Plus or 71 computers) running the specialized data collection programs. The observation data are either manually entered or the data collectors are interfaced with the survey instruments (e.g. KERN E2 or WILD T3000 theodolites) to transmit bi-directional signals. Preparation and preprocessing of the collected field data is executed on a de--p&mental cluster of workstation and personal computers, respectively. Hence, the field data collection is off-line connected to the cluster.

Other activities, like calibration of survey instruments, is performed in sites, located several miles away from the cluster. The data analysis programs run mainly on the cluster. Only some special analysis programs still need a mainframe computer. Summarizing, most of the geodetic activities are performed on the PC/WS cluster.

As shown, the different geodetic actions deal with various data which can be classified as follows (see Figure 2):

- Measurement Data (Data concerned with different observables)
  - Height Data
  - Distance Data
  - Direction Data
  - ...
- Calibration Data (Data about instruments)
  - Tape Data
  - Rod Data
  - Circle Data
  - ...
- Point Data
  - Point Identification Data (Synonym Identifiers)
  - Coordinate Data
  - Coordinate System Data
  - ...

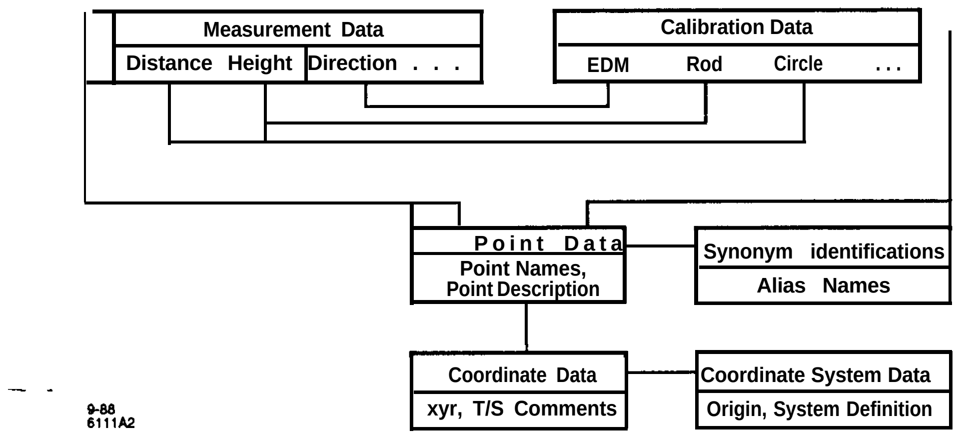

Fig. 2 - Data Classification

- Enhenced Data Integration -

The GEONET data management approach /FrPuRRu87/ was developped to handle the huge amount of geodetic data originated during the construction survey and subsequent realignment surveys of the Stanford Linear Collider (SLC), built by the Stanford Linear Accelerator Center (SLAC) /Er84/. The SLC is a high energy physics particle collider for the research into the behavior and properties of the smallest constituents of matter. During the construction alone, some 100,000 coordinates had to be--determined /OrRRu85, /Pi86/.

The GEONET approach was based on a hierarchical DBM concept which required the hardwiring of data relationships. Nevertheless, GEONET proved to be very successful and has found many applications in the high energy physics survey and alignment community. However, the concept does not provide flexibility of easy assimilation to changing requirements, of easy integration of new tools and of establishing new data relationships. Therefore, future projects like the Supraconducting Super Collider (SSC) which will produce an at least 20-fold increase in the amount of data and will show more complex data relationships due to an increase of observables and more sophisticated and complex mathematical modelling will require new concepts. This situation triggered the project GEOMANAGER.

INTEGRATION OF THE GEODETIC DATA FLOW

As pointed out in the introduction an integration of the data flow among the various geodetic activities is necessary. An integration must provide an open system architecture for an easy integration of new tools and instruments. The geodetic integration concept provided by the GEOMANAGER project consists of two levels:

- Communication integration
- Information integration

Communication integration emphasizes a fully interfacing of all used computers and instruments. The interfaces must be suitable for the required communication. The requirements of the main interfaces of the sample application are:

- Interface: survey instruments and data collection computers
  Special purpose low level signal transmission communication
- Interface: WS/PC cluster and data collector computers
  Transmission of small amounts of field data files
- Interface: WS/PC cluster
  High speed local area network
- Interface: WS/PC cluster and mainframes
  Transmission of large amounts of various data

Interfacing an hydride computer environment can use today will equipped communication standards. But in some cases (e.g. interfacing survey instruments) the customization of special interface boxes is required.

Information and data integration requires an integrated communication management. The major goals of a geodetic data integration result from the following characteristics:

- Geodetic software tools (e.g. data gathering, data analysis programs, etc.) use a huge amount of different data.
- Geodetic software tools share same data.
- Geodetic software tools run in different project environments
- New geodetic software tools must be easily integrated.
- Geodetic data are highly structured and need heterogeneous types.
- Geodetic data own various complex consistency constraints.

Thus, the major goal of geodetic data integration is a unified high level data management, such that all activities can access the data on a high level of abstraction and in a unified way.

These requirements are best fulfilled by a database approach, providing the following concepts:

- Conceptual data centralization
  - Data redundancy elimination
  - Data sharing
- Data independence
  - High level interfaces
- Open system architecture

The GEOMANAGERs database resides on the WS/PC cluster, because all major geodetic activities take place here (see Fig. 3).

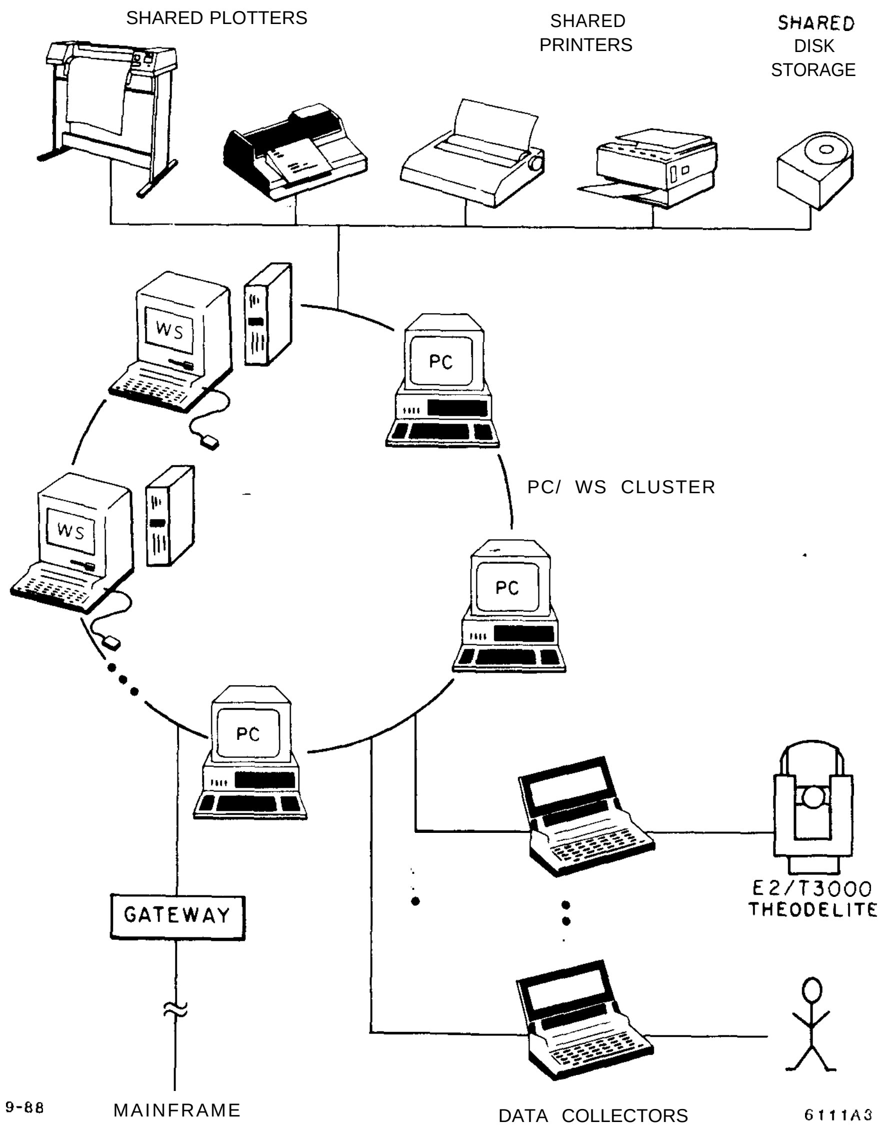

Fig. 3 - Hybride Computer Enviroment

Databases provide some further well-known functions and capabilities, which are also required by geodetic applications. They are not discussed here /DRuRRu87/.

There are various problems and aspects in applying a geodetic database system. Because of space limitations, we focused only on the following aspects:

- Data modelling
- Database interface
- Consistency constraints

Data Modelling

As already pointed out, geodetic data are highly structured and use heterogeneous data types. However, traditional data models do not support all relationship and data types as well as more sophisticated data abstraction concepts. These limited data modelling capabilities complicate the database design process and the database usage. The lack of semantics becomes more important the more complex the data structure of the application is (especially in more sophisticated "nonstandard" database applications, like engineering design, office automation, geographic applications, etc.) Furthermore, geodetic tools run in a wide range of project environments using database systems based on different data models. Thus, the same application data structure must be modelled in different data models, which causes redundant database design processes.

These gaps between applications and traditional data models are bridged by semantic data models. We use the Entity/Relationship model (ER models) extended by the data abstraction concepts of aggregation and generalization hierarchies.

Extended ER schemes are developed for the following major geodetic data classes:

- distance measurement data
- height measurement data
- point data.

ER SCHEMES FOR THE SAMPLE GEODETIC APPLICATION

In Figures 4, 5 and 6 ER diagrams are given for distance measurement data, height measurement data, calibration data, and point data. These ER schemes contain 17 entity types and relationship types, respectively. Because of space limitations, and since the ER diagrams are self-explaining, only a few aspects are pointed out in the following. The entity types DISTANCE- MEASUREMENT, TAPE- METHOD, EDM- METHOD, DISTINVAR- METHOD, and INTERFEROMETER- METHOD describe the distance measurement data.

Entities of the latter entity types specialize the distance measurement data by adding property properties of a specific method. A DISTANCE- MEASUREMENT entity describes the method-independent properties. It must be related to exactly one entity of exactly one METHOD entity type. Thus, DM_ METHOD represents a generalization among the generalized DISTANCE_ MEASUREMENT entity type and the 4 individual METHOD entity types.

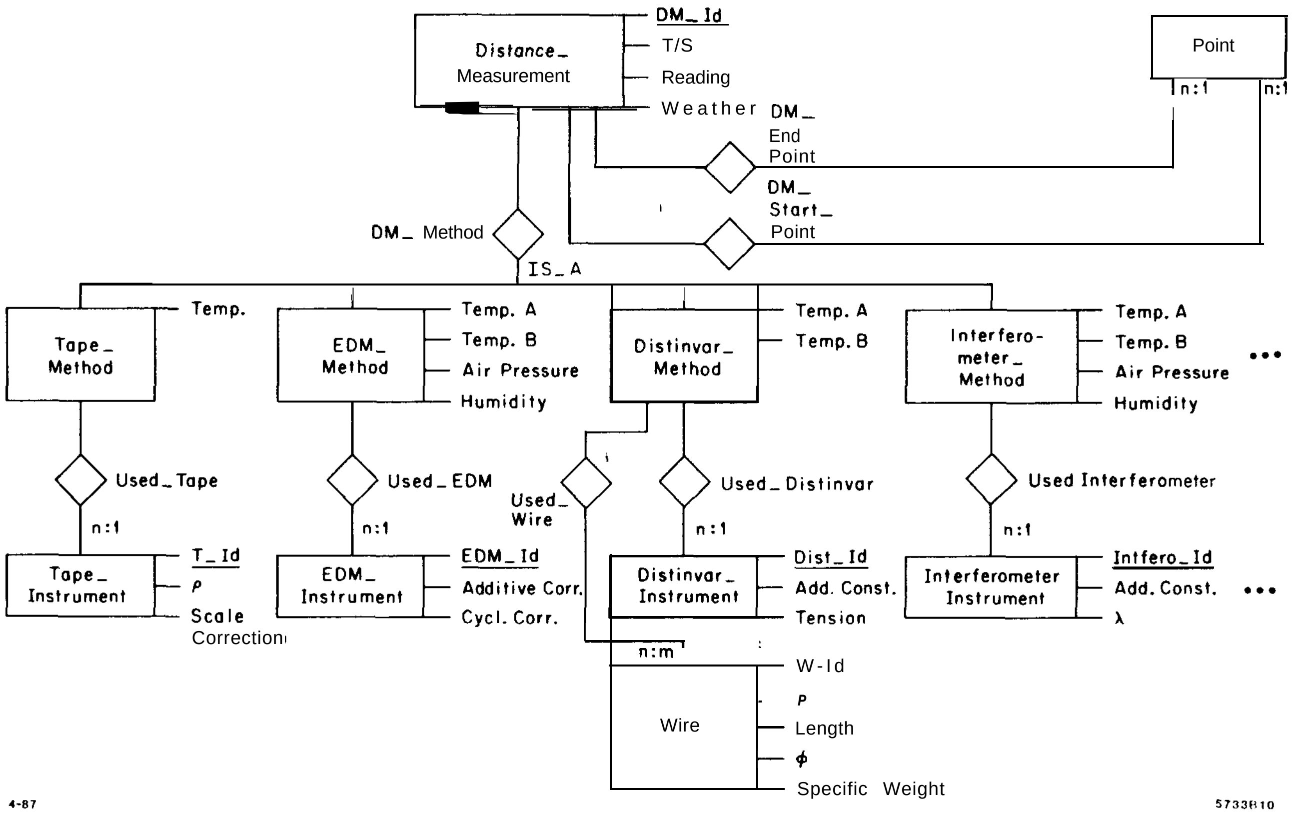

Fig. 3 - ER Diagram: Distance Measurement Data

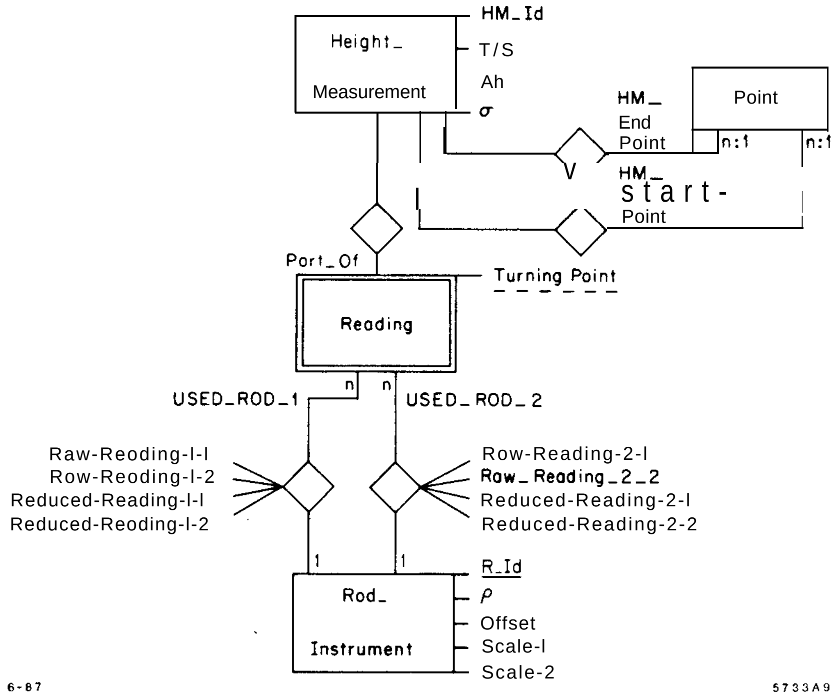

Fig. 4 - ER Diagram: Height Measurement Data

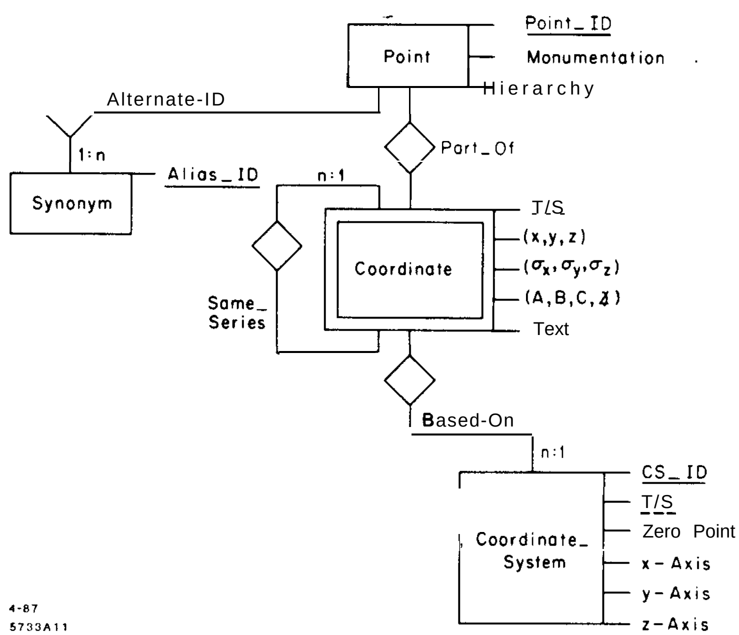

Fig. 5 - ER Diagram: Point Data

The entity types HEIGHT- MEASUREMENT and READING describe the height measurement data. Since a height measurement consists of several readings, which are existence and identification dependent, a PART-OF-relationship type represents these associations.

The entity types TAPE- INSTRUMENT, EDM_ INSTRUMENT, DISTINVAR- IN- STRUMENT- WIRE, and INTERFEROMETER- INSTRUMENT, as well as ROD- IN- STRUMENT describe the calibration data of the instrument used for distance and height measurements, respectively. These entity types are connected to the entity types describing the measurement data. Notice, that the relationship types USED- ROD- 1 and USED- ROD- 2 are the only relationship types with attributes. The attributes represent the raw and reduced readings on the two scales on each of the two rods used. Finally, the entity types POINT, SYNONYM, COORDINATE, and COORDINATE- SYSTEM describe the point data. Each point owns several coordinate data sets. Notice, that the relationship type SAME- SERIES is the only recursive relationship type. It relates coordinates, which result from the same measurement epoch.

DATABASE INTERFACE

The database interface is based on the used data model and must meet the data access and manipulation requirements of the geodetic tools. GEOMANAGER's interface is a hybrid data interface, combining descriptive and procedural elements. First of all, the interface supports elementary operation to access sets of entities or relationships of a single type. The entities or relationships must be qualified by their identifiers. Thus, the elementary operations support a procedural interface.

But, most geodetic tools need an access to aggregates of associated entities of several types. Thus, operations for accessing data aggregates must be supported by the interface. These aggregate operations define a descriptive interface. Its design is based on the following properties of geodetic applications. First, for each geodetic tool a set of generic data aggregate types accessed by this tool can be specified. Hence, the set of used data aggregates are pre-known. Second, some geodetic applications do not have any direct access to the database provided by the communication system (e.g. data analysis programs running on mainframe computers). Other existing geodetic tools do not yet support any database interface. They use their own dedicated file structures.

Thus, the interface supports a pre-defined set of the parametrized access modules for data aggregates. In a first step, the data aggregate type is specified. If data aggregates are retrieved or modified, their qualification is also given. The specification model for qualification statements is derived from predicate logic extended by concepts for handling hierarchies for object classes. This information is especially used by the transaction management for concurrency control and recovery. The second step depends on the communication mode. If direct access is possible, then the specified data aggregates can be retrieved, modified, or written using elementary operations. Thus, this second step access if a procedural one.

If there is no direct access possible, the retrieved data aggregates are downloaded from the database in'a data stream using standardized interchange format. The interchange format is derived from the database scheme. If data aggregates are entered, they must be given as datastream, which is uploaded to the database.

CONCLUSION

In this paper, the need of a data flow integration in geodetic applications is shown. The goal of this paper is:

- to provide some understanding of geodetic activities and of geodetic data flow,
- to evaluate the general potential of integration the geodetic data flow,
- to introduce a two-level integration model,
- to show the problems in applying software tools (i.e. DBMS) in todays market place for this "non-standard" application.

The proposed integration model provides an open system architecture and has two integration levels:

- Communication integration,
- Information and data integration,

This paper addresses the information and data integration level. The requirements are:

- Access of a huge amount of data by the tools,
- Tool migration among various projects,
- Open system architecture,
- Highly structured data,
- Complex consistency constraints.

The goals of the information and data integration are to provide:

- Conceptual data centralization,
    - Data redundancy elimination,
    - Data sharing,
- Data independence and high level database interfaces, using a database approach.

However, DBMs are commonly used in commercial applications, and not frequently in "non-standard" applications, like engineering and scientific applications.

The paper mentioned three problems in using geodetic DBMs in geodetic applications, i.e. GEOMANAGER project:

- Data modelling
- Entity/Relationship schemes, extended by aggregation and generalization hierarchies are developed for our sample application.

- Database interface A hybrid, i.e. procedural and descriptive database interface is developed for accessing simple entities/relationships as well as complex data aggregates. Furthermore, up-and-downloading of data is possible.
- Consistency constraints.

BIBLIOGRAPHY

/DRuRRu87/

D. Ruland, R. Ruland: *"Integrated Database Approach for Geodetic Applications"*, IV. International Working Conference on Statistical and Scientific Data Base Management, Rom, 1988.

/Er84/

Erickson, R. (Editor), *SLC Design Handbook*, Stanford Linear Accelerator Center, Stanford University, CA.

/FrPuRRu87/

H. Friedsam, R. Pushor, R. Ruland: *"The GEONET Approach - A Realization of an Automated Data Flow for Data Collecting, Processing, Storing and Retrieving"* ASPRS-ACSM Fall Convention, Reno, 1987.

/OrRRu85/

W. Oren, R. Ruland, *"Survey Computation Problems Associated with Multi-Planar Electron Positron Colliders"*, in: Proceedings of the ASPRS-ASCM Convention, Washington, 1985, pp. 338-347.

/Pi85/

M. Pietryka, H. Friedsam, W. Oren, R. Pitthan, R. Ruland: *The Alignment of Stanford's New Linear Electron Positron Collider"*, in: Proceedings of the ASPRS-ASCM Convention, Washington, 1985, pp. 321-329.

/RRuFr86/

R. Ruland, H. Friedsam: *"GEONET - A Realization of an Automated Data Flow for Collection, Processing, Storing and Retrieving Geodetic Data for Accelerator Alignment Applications"*, Invited Paper, XVIII Congress, Federation Internationale de Geodesie, Toronto, 1986.