C.C. TSCHERNING
Geodaetisk Institut
Gamlehave Allé 22
DK–2920 Charlottenlund

DEFINING THE BASIC ENTITIES IN A GEODETIC DATA BASE *

Editor's Note .

The following article is not the type of scientific investigation which generally is published in the *Bulletin Géodésique*. It has been accepted for publication on the basis that it serves notice that dealing with questions of data definition and data management are important and appropriate activities for geodesists. The Editor encourages the submission of such descriptive manuscripts which push at the frontiers of geodetic activities.

Abstract

The term "entity" covers, when used in the field of electronic data processing, the meaning of words like "thing", "being", "event", or "concept". Each entity is characterized by a set of properties.

An information element is a triple consisting of an entity, a property and the value of a property. Geodetic information is sets of information elements with entities being related to geodesy. This information may be stored in the form of data *and is called a* geodetic data base *provided (1) it contains or may contain all data necessary for the operations of a particular geodetic organization, (2) the data is stored in a form suited for many different applications and (3) that unnecessary duplications of data have been avoided.*

The first step to be taken when establishing a geodetic data base is described, namely the definition of the basic entities of the data base (such as trigonometric stations, astronomical stations, gravity stations, geodetic reference–system parameters, etc...).

1. Introduction

The International Association of Geodesy (IAG) adopted at its XVI General Assembly in 1975 a resolution recommending exchange of information about the design principles and practical implications of geodetic data bases (Resolution No. 23, cf. Bulletin Géodésique No. 118 (1975), page 383). At the General Assembly it was decided by Special Study Group 4.38 to establish a working group on "Application of Computers for the Handing of Geodetic Data". The working group has five members : J. Isner, J. Gergen, C.R. Schwarz (all at NGS, USA), K. Poder and the author (both at Geodaetisk Institut, Danmark). The "Geodetic Data Base" will be one of the main subjects to be investigated by the working group and the purpose of this paper is to serve as an

* – Presented at the "International Symposium on Optimization of Design and Computation of Control Networks", Sopron, Hungary, July 1977.

Bull. Geod. 52 (1978) pp. 85–92.

introduction to the subject and to initiate a discussion of the subject in the general geodetic community.

Computers have in the past mainly served as very efficient computational tools in Geodesy, but it has become more and more evident that computers are extremely useful for data handling purposes as well. This is mainly due to the availability of fast peripheral storage units like drums and discs.

The application of computers for data handling has undergone a development from a stage where data does not reside permanently in the computer (or even are not available in computer readable form) to a stage where all kinds of geodetic data are directly available in the computer.

Simultaneously data formats and organizations have undergone standardizations as to make data available for many different purposes and avoid unnecessary duplications of data.

In computer science the term "data base" is used to describe the kind of data storage system which contains (or may contain) all data necessary for the operation of a particular organization. It must be suited for many different applications and must not contain unnecessary duplications of data. The system is operated by a so–called "data base management system" which takes care of the data storage, retrieval, checking and updating,

More details (also about the historical development) can be found in the referenced literature.

A data base and a data base management system can be thought of as the "utopia" of data management. Economical or practical obstacles may make it impossible to reach this utopia. But it is worth studying in order to be aware of all the possibilities which are provided by modern computer systems.

In this report are described the basic steps to be taken when establishing a geodetic data base. First some basic definitions will be given, which will lead to a systematic description of the content of a geodetic data base.

2. Representation of information

The information in a Geodetic Data Base can be looked upon as a picture – a map of the part of the real word relevant to Geodesy. In this section certain elements of the mapping function will be described following Engles (1972) and Lindgreen (1974).

The domain of the map can be considered as a set of interrelated *entities,* each characterized by a set of *properties.* Some of the properties reflect attributes of the entities (the station no. 10 is marked by a granite pillar) and others constitute relations between the entities (the station no. 10 is situated 2132,02 m from station no. 11).

The term entity covers the meaning of words like "thing", "being", "event" or "concept". In geodesy e.g. the trigonometric stations, the observation instruments, or a $1^\circ \times 1^\circ$ degree block on the Earth's surface.

A trigonometric station may have different types of "station markers". It is a property that something *has* a station marker, but a *general property.* When we say that a station has a station marker we still expect to be told which kind of station marker. "Granite pillar" is an example of a so–called *definitive property* or descriptor. General properties may be used as descriptors when we want to express that a property is relevant for a given entity, but normally a general property will have an associated *value* (e.g. granite pillar).

Basically values are of two kinds : ***data sets*** or ***numerical values.*** A data set is the result of a representation process, while numerical values are the results of a measurement, a counting or a calculation.

The values associated with a given property form a so–called ***property value set*** which serves to characterize the property in the same way as does the name of the property, its meaning and the set of entities to which it is relevant.

A property may be ***entity relating,*** in which case its property value set will consist of reference to a certain class of entities. For these entities the ***entity identifying property*** is required, namely an existing one to one correspondance between every used value from the property value set and every entity of the class. The values of entity identifying and entity relating properties are sometimes called "primary keys" and "foreign keys", respectively. In geodesy, trigonometric stations may have the entity identifying property through their associated station numbers while gravimetric stations, which are not included in a standardization network, frequently will not have this property. An example of an entity relating property is the property that a trigonometric station belongs to a specific Helmert block used in adjustment computations. The value of the property is a set of references to the other stations within the block.

We will now make a distinction between ***information*** and ***data.*** An ***information element*** is a triple, which consists of an entity, a property and a value of the property. (The station has a station marker which is a granite pillar). Any representation of the information element is called data. An example is the characters in the brackets. And data is what will be ***stored*** in the data base.

This distinction may appear arbitrary, but is made in order to underline that we, on this occasion, will not discuss the different ways of representing information as data, nor discuss data storage structures. These two processes may be quite dependent on the actual computer used and thereby involve technical matters which are out of scope of this report.

In order to handle the information elements, it is necessary to regard the elements as organized in ***information sets.***

There may be many possible organizations of the elements but generally only a few are useful. Information elements are frequently organized according to a common property (trigonometric stations, astronomical stations, etc...) and are then called ***isotypic information sets.*** All the entities in the information set form the ***entity spectrum*** of the set.

Another type of organization occurs when all elements refer to the same entity. Such a set is called an ***isonymous information set.*** All information regarding a trigonometric station may form such a set. The set of properties are called the ***characteristic*** of the information set.

Finally a set of isonymous information sets with the same entity spectrum is called a ***homogeneous information set*** and will, in geodesy for example, be the set of the geodetic coordinates of the trigonometric stations in an area.

(In order to create a suitably big homogeneous information set from two or more smaller sets it may be necessary to add two property values $\boldsymbol{null_0}$ and $\boldsymbol{null_1}$ to all value sets. These values describe that an entity does not have a certain property, that an entity does have a property, but that the value is unknown, respectively. Fictitious trigonometric stations may, for example, have a "$null_0$" station marker).

In traditional computer terminology we may talk of a homogeneous information set as being represented by a file consisting of a possibly ordered set of records each representing an isonymous information set. A record will consist of an ordered set of *fields* and may be interpreted either sequentially or by means of specific key values. In the last case the record will contain key fields representing information elements with entity identifying properties. A set of fields of a specific kind, but from different records in the file, constitute the representation of an isotypic information set.

In the following section we will discuss how geodetic information may be organized as homogeneous information sets.

3. The basic entities

We shall here consider "geodesy" as having the task of determining the position of points on the surface of the Earth and the time variations of the positions as well as the external gravitational potential. *"Geodetic information"* thereby, becomes information about the positions and about the geopotential, plus information about the activities which have lead to or are planned to lead to a determination of these quantities.

Let us suppose that we want to determine the position of some points in unsurveyed area, or in an area where we suppose or know that crustal movements take or have taken place. We then plan a geodetic (horizontal or vertical) network, which we investigate in order to assure that a required precision can be obtained, that the heights of observation towers are known etc... When the necessary observations (of distances, angles, deflections of the vertical, heights etc...) have been executed (and recorded), an adjustment computation will take place. This will (hopefully) give us the estimated positions and related standard deviations which finally must be made available to the users of geodetic information.

It is obvious that there are a huge number of basic entities, of which information must be recorded. And it is also obvious that some entities are important in one situation and only of historical interest in another.

A typical feature in geodetic operations is that the position of a point is determined by measuring distance to and angles or height differences between other objects (points). *Table 1* gives examples of basic entities all related to "points". They will then have, in common, the property of having associated a set of coordinates. They are distinguished by the type of observation performed at or to the point.

In order to perform the observations, instruments, observers and logistic facilities must be available. For the computations the instrument characteristics, geodetic reference system parameters, etc... are needed besides the manually or automatically recorded readings of scales. *Table 2* gives examples of these types of entities and related properties.

Finally, the result of the adjustment process may be either coordinates of physically well–defined points on the surface of the Earth (at a given epoch), the value of a gravity field dependent quantity in a point (geopotential, gravity or similar quantity) or numerical representations of such quantities (a terrain function for topographic heights, a potential coefficient set).

They may have been produced by executing a specific computer program in a specific run which then may be considered another type of entity or a property of the points or functions.

The reader may add other entities to these mentioned above. We will not go in

Table 1

Examples of Geodetic Point Entities
(1) Trigonometric station (with or without station marker)
(2) Precise levelling station
(3) Astronomical station
(4) Doppler, Laser or Optical–satellite station
(5) Gravimetric station
(6) Gravity gradient station
(7) Seismological or Earth tide station
(8) Pass–points on photogrammetric plates
(9) Balloons, satellites (geodetic)
(10) Recognizable points on the moon, the sun or other stars at specific epoch
(11) Airplane or ship (performing geodetic, photogrammetric or gravity field related observations)
(12) Center or corners of regulary shaped area on the Earth where mean value of topographic heights or gravity anomalies have been computed

more detail here with respect to other possible classifications of the basic entities but conclude, that it seems likely that homogeneous information sets may be formed of

(a) each of the entities (1) – (9) in *Table 1* with their related adjusted coordinates and contingently related gravity field dependent quantity,

(b) observation instrument and observer constants,

(c) observations (meteorological conditions, observer, instruments, etc...).

Note, that e.g. Schwarz (1975) regards (a) and (c) for trigonometric stations as one homogeneous information set.

Table 2

Examples of entities and properties needed when executing observations and performing adjustment computations

Entity	Property (examples)
Instrument (EDM, Doppler, Cameras, Barometers, Theodolites, Gravimeters, Torsion balance instr., levelling instr., Altimeters)	Instrument constants Instrument identification number
Observer	Name or other identification. Personal equation as used in astronomical position determination
Building and transport equipment	Heights of observation towers
Geodetic reference systems	Name of reference system, defining parameters (a, 1/f), datum transformation parameters
Map projections	Names of projections. Parameters defining the projections
Time corrections, Polar motion, Earth Tide	The value of these corrections or quantities at a specific time or location
Observations	Type of observation, value of observation (scale reading), meteorological conditions, instrument used, observer, time of observation, a priori standard deviation, (refraction), contingent excentricity elements.

It should be emphasized that *Table 1* and *2* contain examples and are not exhaustive.

4. Properties and values

Examples of recorded properties for trigonometric stations can be found e.g. in Porter (1975) and for gravimetric stations in Lepretre (1976). We shall here shortly discuss the important *entity relating* property and the properties associated with numerical values.

That a set of entities have the identifying property does, as mentioned above, imply the existence of a unique mapping between names (e.g. station numbers) and entities (trigonometric stations). The importance of and difficulties encountered when establishing such mapping has been described in Schwarz (1975), who describes a mapping based on the knowledge of the geographical coordinates of a point. Such a procedure may be problematic in an unsurveyed area where it often is needed to have unique identification of stations even before the first observations have been executed.

What has proved to be practical is the assignment of a certain interval of integer numbers to a surveying party in a specific period. The numbers will then generally be associated with points in the same geographical area and thus facilitate sorting of information. This may naturally lead to non–unique station numbers when e.g. several networks established by different local or national agencies have to be adjusted simultaneously. The problem is generally overcome by adding different sufficiently big numbers to the original numbers or by using prefix characters like the ones used to identify automobiles internationally.

(Maybe it would be worth while if an international agreement could be reached for the association of identification numbers to geodetic stations of international importance, such as satellite tracking stations).

A related problem is that the different identifiers may describe the same station. In Denmark, for example, trigonometric stations in the primary network has two sets of numbers, the one conforming to a cadastral station identification system, the other used internally by the Geodetic Institute. Similary stations on the ISGN 71 network will have both numbers assigned by Bureau Gravimetrique, Defense Mapping Agency (USA) and contingently national identification numbers. Here a tremendous confusion is possible.

A similar problem is related to the numerical values recorded either when reading instrument scales or originating from computations. The values will often carry an unknown number of significant decimals and the units used may vary. Each numerical value should therefore have associated a unit everywhere it occurs and a number of significant digits. (It should, when possible, also have an associated standard deviation, but this is another problem). Also a station may have several recorded values for its coordinates or observed (adjusted) values of directions and distances to other points. These values should always have an associated source identification (with possible time of origin).

These mentioned problems may seem unimportant, but as the number of recorded data increase and computers and geodesists change, previously well defined and established conventions will turn into confusion if the necessary precautions have not been taken.

5. Conclusion

When the basic entities and their properties have been defined and the homogeneous information sets have been characterized the next step is to find out how the information can be represented as data and how this data can be stored in a specific computer system.

Data base management systems are developed by most computer manufacturers and software producers, which may take care of the information when it first has been transformed to data in a computer readable form. A number of such systems have been investigated at the National Geodetic Survey (USA), but "no one system was found that could solve the problem" (**Alger** and **Gurley** (1975, page 3)). So the transition from information to data and from data to storage is not straightforward.

Acknowledgment

The members of the working group on "Application of Computers for the Handling of Geodetic Data" and my colleagues at Geodaetisk Institut are thanked for their valuable critical comments.

LITERATURE

(A) Geodetic Literature on or related to Data Bases and Data Base Management :

David E. ALGER, and Joseph C. GURLEY : A Design for a Geodetic Data Base Management System, in Proceedings of the American Society of Photogrammetry, 41st Annual Meeting, March 9–14, 1975, ASP, Falls Church, Virginia, USA.

R.J. BUCK, and J.G. TANNER : Storage and Retrieval of Gravity Data, Bulletin Géodésique, No. 103, 1972, pp. 63–84.

J.P. LEPRETRE : Progress Report for the Creation of a Worldwide Gravimetric Data Bank, Bureau Gravimétrique International, Bulletin d'Information, No. 39, Nov. 1976.

T.A. PORTER : Canadian National Positional Control Survey Date File, A Geodetic Data Bank for Canada, The Canadian Surveyor, Vol. 28, No. 5, (1974), pp. 724–731.

Charles R. SCHWARZ : The Geodetic Data Base at NGS, presented at the XVI General Assembly of IUGG, Grenoble 1975.

James E. STEM : The NGS Data Base as a Central Depository of Geodetic Control Data, ACSM, 1976, pp. 49–57.

(B) General Literature on Data Bases and Data Base Management :

C.J. DATE : An Introduction to Data–Base System, Addison–Wesley, Reading, Mass., 1975.

R.W. ENGLES : A Tutorial on Data–Base Organization, Annual Review in Automatic Programming, Vol. 7, part 1, 1972, pp. 1–64.

James P. FRY : Evolution of Data–Base Management Systems, Computing Surveys, Vol. 8, No. 1 (1976), pp. 34–42. (This volume of Computing Surveys contains several other very interesting papers on Data Base Management with extensive literature lists).

Paul LINDGREEN : Basic Operations on Information as a Basis for Data Base Design, Information Processing 74, (1974), North–Holland Publishing Company, pp. 993–997.

Received : 13.06.1977
Accepted : 10.08.1977