THE SOUTH AFRICAN STANDARD FOR THE EXCHANGE OF DIGITAL GEO-REFERENCED INFORMATION

A K Cooper, Project Leader: GIS
Centre for Advanced Computing and Decision Support,
CSIR, P O Box 395, PRETORIA, 0001,
Republic of South Africa

BIOGRAPHICAL SKETCH

Antony Cooper holds the BSc (Information Processing) and BSc (Honours) in Computer Science degrees from Rhodes University, Grahamstown. He is a recipient of a RICS/Auto Carto London Education Trust Award and is a member of the Council of the South African Institute of Computer Scientists. His research interests include standards for the exchange of digital geographically referenced information, fundamental concepts of digital geographically referenced information, digital terrain models and applications thereof.

ABSTRACT

Geographically referenced (geo-referenced) information consists of all information that refers to the human-environment system and that can be localized in space and time. This includes cadastral, topographic, hydrographic and statistical information. The need for standards for the exchange of digital geo-referenced information is well known. The author was a member of the project team which drafted the South African standard and is a member of the committee charged with maintaining this standard.

This paper will provide a technical overview of the South African standard for the exchange of digital geo-referenced information. It will describe briefly our concepts of geo-referenced information and the relational model used, which makes the standard easy to use and update. A set of data being exchanged consists of the *File Identification* (a fixed length, fixed format file that identifies the data), the *Global Information Section* (giving general details about the data being exchanged, such as reference surface and coordinate offsets used) and the *Geo-referenced Information Relations* (containing the data being exchanged). This paper will describe these components, specifying how they cater for information on data quality, classification, non-spatial attributes, alternate spatial attributes, vector and raster data.

INTRODUCTION

Geographically referenced (geo-referenced) information consists of all information that refers to the human-environment system and that can be localized in space and time. Thus, geo-referenced information is of a diverse nature and includes cadastral, topographic, hydrographic, geological, remotely sensed and statistical information. In a digital form, geo-referenced information consists of vector, raster and alphanumeric data, as well as the inter-relationships between the various data. Standards for the exchange of digital geo-referenced information have to cater for the diversity in the nature of the digital data and the diversity in the nature of the geo-referenced information.

This paper describes the South African standard [Clarke *et al* 1987], which attempts to cater for all forms of digital geo-referenced information. The standard is based on a relational model, which makes it modular and thus flexible and relatively easy to use and update. A set of data being exchanged consists of a File Identification, a Global Information Section and a number of Geo-referenced Information Relations.

The standard has been reviewed by Lane [1988].

THE NATURE OF GEO-REFERENCED INFORMATION

Digital geo-referenced information is a representation of part of the real world and typically its location in space and time is recorded in two or even three spatial dimensions (typically the two planimetric dimensions and the vertical distance above, or below, some reference surface) — only rarely is its location recorded in the temporal dimension. The current version of the exchange standard caters for two and three dimensions. There are three forms of digital geo-referenced information, namely *vector*, *raster* and *alphanumeric*. In addition, there is information on the spatial relationships inherent in the data, namely the *topology*. The exchange standard provides for the above, as well as mechanisms for exchanging information on the quality of the digital data and *alternate spatial attributes* — multiple versions of the digital representation of an entity.

Features

Features are the basic entities of digital geo-referenced information. A simple feature is a set of one or more uniquely identifiable objects in the real world where the defined characteristics of the objects are consistent throughout all the objects. Features can be man-made or natural, real or abstract. These defined characteristics are known as the *attributes* of the features, and can be *spatial* (that is, dependent on the feature's position in the n-dimensional space) or *non-spatial* (that is, independent of the feature's position — also known as the descriptive information of the feature). Thus, descriptive geo-referenced information is fixed in time and space through the features.

Classification is the arrangement of features into classes or groups and should be done on the basis of the *qualitative* characteristics of the objects, such as their function, and not on their *quantitative* characteristics. A feature's classification should be based on those of its characteristics that are least likely to change. There is a fine distinction between the non-spatial attributes of a feature and its classification because for different users, different criteria for classifying the information apply. One could even consider the classification itself to be a non-spatial attribute [Cooper 1987a].

While the exchange standard may be used with any classification scheme, the standard includes a skeleton classification scheme based on a variable-level hierarchical model for classification [Clarke *et al* 1987, Scheepers *et al* 1986].

Spatial attributes

A *spatial attribute* is an attribute whose value is a subset of any n-dimensional space — this version of the exchange standard caters for only two and three dimensions as they are the most commonly used. Should further dimensions become widely used, the standard will be expanded to cater for them, which should not prove difficult. Note that in the current version of the exchange standard, temporal values may still be recorded as non-spatial attributes. Spatial attributes may be *vector* (that is, positional data recorded as coordinate tuples forming nodes, chains, etc) or *raster* (that is, data expressed as a tesselation of cells, with spatial position implicit in the ordering of the cells).

The four fundamental types of two-dimensional vector spatial attributes are *nodes*, *chains*, *arcs* and *regions*, while the fundamental raster spatial attribute is the *matrix*.

A *node* is a 0-dimensional object with an n-tuple of coordinates specifying its position in n-dimensional space. The position of a *point feature* is described by a single node.

A *chain* is an ordered undirected sequence of n-tuples of coordinates with a node at each end. An *arc* is any continuous part of the circumference of a circle with a node at each end. The position of a *line feature* is described by a set of one or more chains and/or arcs, which do not necessarily form a continuous object.

A *region* is the interior of a continuous and closed sequence of one or more chains and/or arcs, known as the region's outer boundary. The position of an *area feature* is described by a set of one or more regions, which do not necessarily form a continuous object.

A *matrix* consists of an n-tuple of coordinates, specifying its origin, and an m-dimensional rectangular tesselation of data values encoded in a pre-defined format. The position of a *grid feature* is described by a set of one or more matrices, which do not necessarily form a continuous object.

Compound features are those which consist of one or more other features. This allows the user to build a hierarchy of features, for those occasions when the individual constituent features have their own non-spatial attributes (and classification), but together they have other additional non-spatial attributes and a classification.

Topology

The exchange standard caters for two topological relationships, namely *coincidence* and *exclusion*. Coincidence refers to the sharing of common sets of coordinate tuples, and is modelled by having more than one feature share the same spatial attributes. Exclusion refers to area features that consist of regions that wholly contain other regions that do not form a part of the area feature. Exclusion is catered for explicitly through two relations in the exchange standard.

Alternate spatial attributes

A feature has *alternate spatial attributes* when it is represented by a number of different sets of spatial attributes, where each set defines fully the location of the feature. An *alternate spatial attribute scheme* determines the manner in which the different alternate spatial attributes are related to their features. There are two main reasons as to why a feature would have alternate spatial attributes.

Firstly, in an area with a high density of features, the graphical representation of the area (be it on a computer screen or hard copy) would be messy, unless the display of some of the features could be suppressed, or unless some of them could be represented in a simplified manner. However, for analysis on the spatial attributes of the features, one would prefer to retain the spatial attributes of all the features and to as much detail as possible. Alternate spatial attributes allow one to keep different versions of the spatial attributes for the features to solve this problem — at one level, the alternate spatial attributes are for display, while at another level they are for analysis.

Secondly, if one deals with data at greatly disparate scales, one would like to retain different, scale dependent, versions of the spatial attributes of those features which appear at both small and large scales — automatic generalization of spatial data from a large scale to a small scale is still an interesting research area, and it is not possible to create large scale spatial data from small scale data! Again, alternate spatial attributes allow one to keep more than one set of spatial attributes for a feature.

In the exchange standard, an entry in the Global Information Section determines whether alternate spatial attributes are used in the data set being exchanged, and if so, which scheme is used. If they are used, then in the Geo-referenced Information Relations, the field *Alternate spatial attribute* is used in every relation between features and spatial attributes, as well as in the two relations which define the type of the feature (point, line, etc) and its planimetric spatial domain. If alternate spatial attributes are not used, then the field is ignored completely.

There is a relation in the Geo-referenced Information Relations for exchanging, with the data set, an alternate spatial attribute scheme — no such scheme is defined in the current version of the exchange standard.

Information on the quality of the digital data

The American National Committee for Digital Cartographic Data Standards (NCDCDS) identified the nature of information on the quality of digital geo-referenced information, and which information should be recorded [Moellering 1986, Chrisman 1986].

Although some exchange standards allow for the encoding of some forms of information on the quality the digital data, such as the British standard [Sowton & Haywood 1987], we have followed the lead of the NCDCDS and allow the information on quality to be exchanged as free text only. A relation is used which may be included as often as necessary in amongst the Geo-referenced Information Relations. The granularity of the information on quality can thus vary from coarse (referring to the whole data set) to fine (referring to a section containing only one instance of a particular relation) [Cooper 1987a].

Only once the quantification of information on the quality of digital geo-referenced information is well understood, will the exchange standard address the encoding of such information on the quality of the digital data.

THE RELATIONAL MODEL OF THE EXCHANGE STANDARD

A data set in the format of an exchange standard is not a data base — it is merely a set of data that has been extracted from one data base with the purpose of being incorporated into another data base. To be successful, an exchange standard must be independent of the data bases that might be interfaced to it.

There are three common models for data structures, namely the hierarchical, the network and the relational. This exchange standard uses a relational model because it is inherently modular and more flexible than the hierarchical or network models. In a relational structure, the data are represented in a single uniform manner, and thus operations on the data are robust and simple to implement.

When creating a data set in the format of the exchange standard, one merely omits those relations for which one has no data. It is easy to add new relations to the exchange standard — in fact, data that can be exchanged through the relational structure of the exchange standard should always be able to be exchanged through the exchange standard, no matter how many new relations are added to cater for new concepts or types of data [Cooper 1987b]. This is achieved by adding new relations and leaving the existing ones as they are, rather than modifying the existing relations.

It is desirable to have a degree of normalization in data in a relational form [Van Roessel 1987]. There are some relations in the exchange standard for which normalization was not really feasible due to the excessive storage and processing overheads that would be introduced. For example, the records in the relation containing the internal coordinates of chains have variable numbers of fields (one field for each coordinate). For the rest of the relations, an attempt was made to normalize the relations to the third normal form. This required the introduction of a *Sequence number* field to the keys of those relations where the keys were not unique, for example the relation relating classification to feature — any feature class may have many features with that classification. However, the sequence number appears only in the document describing the standard and not in the data being exchanged. As the data in the data set have an inherent ordering, the sequence number is implied by the record's position in the data set.

As an example, the following are the relations which relate an area feature to its classification and its spatial attributes:

1. *Feature/classification* which relates:
   Feature ID $\Longleftrightarrow$ *Classification*

2. *Feature/feature type* which relates:
   Feature ID $\Longleftrightarrow$ *Feature type*

3. *Area feature/included regions* which relates:
   Feature ID $\Longleftrightarrow$ *Region ID*

4. *Region/chains & arcs & direction* which relates:
   Region ID $\Longleftrightarrow$ *Indication of chain or arc* $\cup$ *Chain ID* $\cup$ *Direction indicator*

5. *Chain/nodes & coordinate tuples* which relates:
   Chain ID $\Longleftrightarrow$ *Node ID* $\cup$ *Node ID* $\cup$ *Length of chain* $\cup$ *Data ID*

6. *Node/coordinate tuple* which relates:
   Node ID $\Longleftrightarrow$ *Coordinate tuple*

7. *Chain data* which relates:
   Data ID $\Longleftrightarrow$ *Coordinate tuples*

Relation 1 classifies the feature, relation 2 identifies the feature as an area feature, relation 3 connects the feature to its region spatial attribute, relation 4 performs the topological link between the region and the chains and arcs which form its boundary (specifying whether the chains and arcs are used forwards or backwards), relation 5 links the chains to their start and end nodes and to their internal coordinate tuples, relation 6 specifies the coordinate tuples identifying the locations of the nodes and relation 7 contains all the internal coordinate tuples for the chains.

FILE IDENTIFICATION

The *File Identification* is a fixed format file for identifying the set of data being exchanged. It is 2048 bytes long and consists of standard 7-bit ASCII characters. The fixed format facilitates the extraction of the various fields, both by computers and humans! Most of the information in the File Identification is in a free text, human-readable form (for example, the *Data identification*, *Source* and *Maintenance organizations*, *Copyright statement* and *Comments*), while some is in a formatted, computer-readable form, yet still intelligible to a human (for example, the *Volume number*, *Time* and *Date stamps*, *Physical record size* and *Blocking factor*.

The purpose of the File Identification is to allow the recipient of the data set to identify the data set, its currency and its relevance to his geographical information system, without having to do involved interpretation of the data set. The volumes of digital geo-referenced information that any user might receive, and thus the volumes of various physical media containing such information that might reside in the user's storage, are potentially enormous. The File Identification is there to provide identification of the data should the physical label on the media prove to be missing, illegible or cryptic.

In addition, the File Identification provides some information to the interface program attempting to interpret the data set — for example, the *Physical record size* and *Blocking factor* indicate the manner in which the data are stored on the physical exchange medium, and the *ASCII/Binary* and *Explicit lengths/Delimiters* flags indicate whether the data are stored using 7-bit ASCII characters or in binary, and whether the fields are separated by delimiters of whether the lengths of the fields are determined by explicit length fields appearing before each field.

The File Identification forms the *first physical file* of a data set being exchanged. The rest of the data forms the *second physical file.* On a magnetic tape, these two files are separated by two end-of-file markers. The first version of the exchange standard describes the use of only magnetic tape as the physical exchange medium, as very few users in South Africa use anything else at this stage. This does not preclude the use of any other exchange medium, however.

GLOBAL INFORMATION SECTION

The *Global Information Section* provides details of the data being exchanged, such as the *Projection or coordinate system* and the *Reference surface* used. Some consider this information to be information on the quality of the data being exchanged — we consider the information to be critical for the correct interpretation of the data being exchanged.

The entries in the Global Information Section consist of variable length fields and records with either delimiters between the fields and records, or with explicit lengths at the beginning of each field, as indicated in the File Identification. However, the use of delimiters is recommended as they are conceptually easier to understand and implement, both when creating and interpreting the data set.

Most of the entries have default values and are thus optional. Those that do not have defaults are essential, for example the *Standard meridians & parallels & scale factor.*

Other entries in the Global Information Section include the *Units* and *Increment* of the *Planimetric* and *Vertical coordinate resolutions*, the *Bounding planimetric quadrilateral coordinate tuples* and the *Data quality*, *Feature classification*, *Attribute* and *Alternate spatial attribute schemes* and *release numbers.*

GEO-REFERENCED INFORMATION RELATIONS

The *Geo-referenced Information Relations* contain the actual data being exchanged. Each section, which corresponds to a table in a relational data base, contains a sequence of instances of a particular relation.

As in the Global Information Section, the sections in the Geo-referenced Information Relations consist of variable length fields and records with either delimiters between the fields and records (and sections), or with explicit lengths at the beginning of each field, as indicated in the File Identification. In addition, there is a relation, namely *TEMPLATE*, which allows the creator of the data set the option of using explicit lengths to set up templates for the fields, and hence make the fields fixed length fields. However, the use of delimiters is recommended.

The relation for exchanging information on the quality of the digital data, namely *DATAQUAL*, consists of free text which describes the quality of the data. In addition, the *Description* fields in the relations for exchanging the classification, namely *EXCHCLAS*, the non-spatial attribute scheme, namely *EXCHATTR*, and the alternate spatial attribute scheme, namely *EXCHASAS*, also contain free text. All other fields and relations contain information in a format encoded explicitly for automatic interpretation by the interface program of the recipient.

EXCHCLAS and *EXCHATTR* provide the user with a data dictionary facility for exchanging the definitions of the classification and attribute schemes with the data set.

A number of relations have inverse relations. For example, there is the relation *FEATCLAS* relating a feature to its classification, and the inverse relation *CLASFEAT*, relating a feature class to features with that classification. All the relations between features and spatial attributes have inverses, for example *FEATNODE* (relating point features to nodes) and *NODEFEAT* (relating nodes to point features), and all the topological relations amongst the spatial attributes, for example *REGICHAI* (relating regions to chains and arcs) and *CHAIREGI* (relating chains and arcs to regions).

Finally, there are the geometric data relations which contain the coordinate tuples and constitute the bulk of the data set — especially *CHAIDATA*, which contains the internal coordinate tuples of the chains.

DIFFERENCES WITH RESPECT TO OTHER STANDARDS

The designers of the South African national standard for the exchange of digital geo-referenced information had the benefit of drawing on the work performed in other countries on similar exchange standards, as well as the opportunity of holding discussions with some of the designers of these standards — in particular, the standards of Australia [SAA 1981], North America [DCDSTF 1988], United Kingdom [Sowton & Haywood 1987] and the International Hydrographic Organization (IHO) [CEDD 1986].

We believe that the South African standard is the first to attempt to cater for all forms of digital geo-referenced information — the other standards are generally targeted at either cartographic or hydrographic information.

The Australian standard uses a hierarchical structure, the British standard uses a combination of network and relational models, the American standard allows the use of either a hierarchical or a relational model and the IHO standard uses a network model. The South African standard uses a relational model.

All four of the abovementioned standards include a full classification scheme and the American, British and IHO's standards include comprehensive non-spatial attribute schemes. The American, British and IHO's standards allow the use of any classification and non-spatial attribute schemes. The South African standard includes a skeleton classification and non-spatial attribute scheme, and allows the use of any such scheme.

CONCLUSIONS

While the community in South Africa supports the national exchange standard in principle, few attempts have been made to implement it. The Institute for Natural Resources at the University of Natal in Pietermaritzburg have implemented a significant subset of the interface in both directions between their home-grown geographical information system (GIS) and the exchange standard, and other organizations are at the design stage of the implementation. For their tender for a GIS, the Department of Water Affairs distributed benchmark data in the format of the exchange standard [Olivier *et al* 1989]. These efforts have shown that the basic concept of the exchange standard is sound. They have also highlighted a few problems with some of the relations in the exchange standard. None of these problems is critical and they will be addressed in the next edition of the exchange standard, due to be published in the first half of 1989. In addition, they have unearthed some interesting problems concerning the fundamental nature of digital geo-referenced information [Greewood 1988].

In addition to maintaining the exchange standard, the National Exchange Standard Committee will keep a record of digital geo-referenced information available in South Africa. To this end, a questionnaire was distributed [NESC 1988].

We believe that the process of developing this standard has made a significant contribution to creating more awareness among the South African GIS community of the fundamental concepts of geo-referenced information [Cooper 1988].

ACKNOWLEDGEMENTS

I should like to thank the Trustees of the Education Trust of The Royal Institution of Chartered Surveyors (RICS) for awarding me a RICS/Auto Carto London Educational Trust Award, which made presentation of this paper possible.

I should like to thank the other members of the project team, Derek Clarke, Hester van Rooyen and Elri Liebenberg, my colleagues at the Centre for Advanced Computing and Decision Support and the GIS community, both in South Africa and abroad, for all their help, criticisms and advice.

The opinions expressed in this paper are those of the author and not necessarily of the CSIR.

REFERENCES

Chrisman NR, September 1986, *Obtaining information on quality of digital data*, Proceedings: Auto Carto London, Vol 1, pp 350–358.

Clarke DG, Cooper AK, Liebenberg EC & Van Rooyen MH, September 1987, *A national standard for the exchange of digital geo-referenced information*, NRIMS CSIR Special Report SWISK 45, 201 pp.

Committee on the Exchange of Digital Data, November 1986, *Format for the exchange of digital hydrographic data*, International Hydrographic Organization, 350 pp.

Cooper AK, March 1987, *Thoughts on exchanging geographical information*, Proceedings: 1987 ASPRS-ACSM Annual Convention, Vol 5, pp 1–9.

Cooper AK, June 1987, *A data structure for exchanging geographical information*, Proceedings: 4th South African Computer Symposium, pp 267–277.

Cooper AK, September 1988, *Exchanging geographically referenced information — a status report*, Proceedings: Computer Graphics '88, pp B1-6 – B1-20.

Digital Cartographic Data Standards Task Force, January 1988, *The proposed standard for digital cartographic data*, American Cartographer, Vol 16, No 1.

Greenwood PH, September 1988, *Using the proposed national exchange standard for GIS data*, Proceedings: Computer Graphics '88, pp B1-21 – B1-29.

Lane A, 1988, ***Review:** a national standard for the exchange of digital referenced information*, International Journal of Geographical Information Systems, Vol 2, No 1, pp 81–82.

Moellering H, ed, January 1985, *Digital cartographic data standards: an interim proposed standard*, Report no 6, National Committee for Digital Cartographic Data Standards, 164 pp.

National Exchange Standard Committee, July 1988, *Questionnaire: National Exchange Standard*, Chief Directorate: Surveys and Mapping, 11 pp.

Olivier JJ, Greenwood PH, Cooper AK, McPherson DR & Engelbrecht R, April 1989, *Selecting a GIS*, Proceedings: Auto Carto 9.

Scheepers CF, Van Biljon WR & Cooper AK, September 1986, *Guidelines to set up a classification for geographical information*, NRIMS CSIR Internal Report I723, 12 pp.

Sowton M & Haywood P, chair, January 1987, *The national transfer format: release 1.0*, Ordnance Survey, Southampton.

Standards Association of Australia, August 1981, *Interchange of feature coded digital mapping data*, Australian Standard 2482–1981, 24 pp.

Van Roessel JW, 1987, *Design of a spatial data structure using the relational normal forms*, International Journal of Geographical Information Systems, Vol 1, No 1, pp 33–50.