Data sharing using the X.500 Directory

Dr Richard T. Pascoe

Spatial Information Research Centre (SIRC)
Computer & Information Science, PO Box 56, University of Otago, Dunedin, New Zealand.
Telephone: +64 3 479 8321. Fax: +64 3 479 8311. Email: rpascoe@commerce.otago.ac.nz

Abstract: *Sharing geographical data sets is highly desirable for economical and technical reasons. In this paper the author describes the development of an agency for sharing geographical data which is based on the use of the ISODE implementation of the X.500 Directory Service and a collection of software agents which collaborate with each other to perform the various task associated with sharing data.*

1 Introduction

Sharing geographical data sets is highly desirable for economical and technical reasons. Economically, the cost of acquiring digital data can be distributed across different organisations resulting in either lower costs for each organisation or greater funding for capturing higher quality data. Technically, sharing data sets is made easier if the underlying framework on which these sets are captured is the same. That is, the underlying framework is itself a shared data set. Consider, for example, if utility organisations providing telecommunication, power, and water services all used the same cadastral data set as the basis upon which they added their information. Individual organisations could more easily share their own information with others because these sets have a common basis. In this paper is described the author's development of an agency for sharing geographical data in an effort to gain the advantages outlined above.

Traditionally, the function of an interface was restricted to translating data from one file format to another. If the user of a Geographical Information System (GIS) is to see, as a single geographical database, collections of data that may be stored in different representations at various locations, then interfaces must be constructed to transfer data between many different sites across a communications network. A communicating interface is defined to be an interface that sends data to, or receives data from, some other interface through a communications network.

In earlier research oriented towards simplifying the construction of interfaces, Pascoe & Penny 1993 describe a distributed application (called gds/gdc) which comprises several communicating interfaces. This application, shown in Figure 1, searches for data sets with a spatial extent overlapping that defined by a user, and transfers any such data sets back to the user. The search was performed on a collection of descriptions stored within a text file local to the user. Each description is for a data set which has an associated communicating interface within the application and consists of a filename and the spatial extent of the data contained within this file. Linking the communicating interfaces needed to transfer the data is dealt with using the quipu X.500 Directory Service (ISODE Volume 8 User's Guide Directory Services 1994).

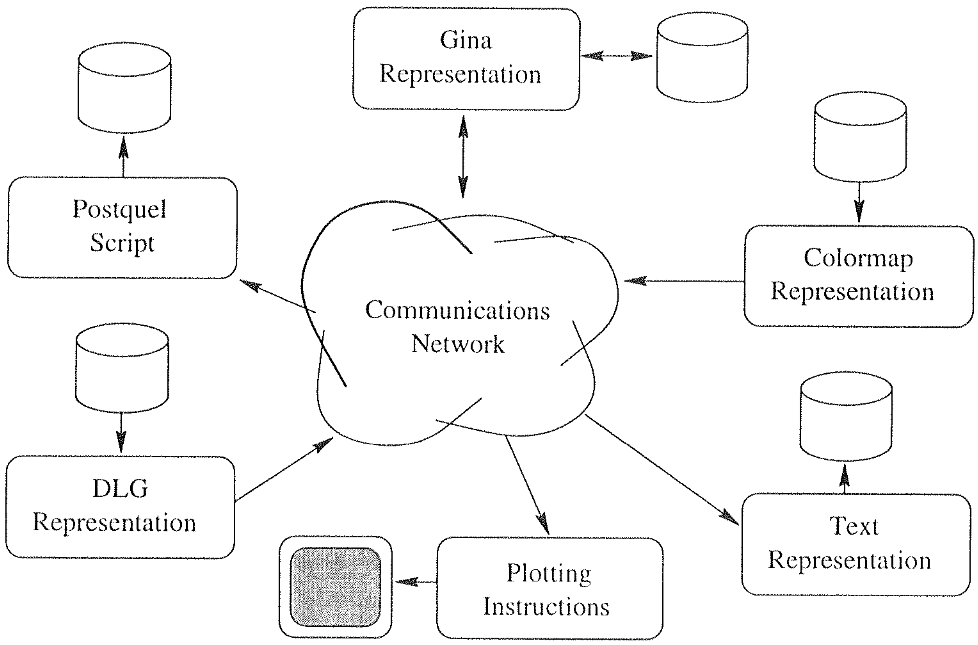

Figure 1: The communicating interfaces forming the gds/gdc application

Searching for data sets and interfaces in general would be facilitated by a directory in which are described publicly available data sets, interfaces, and any other software associated with the data. Ideally, entries within the directory should be maintained by the custodians of the data sets. Therefore, the directory needs to be distributed across many different computer systems with directory entries on each computer system being controlled by the custodians of the locally stored data. Design and use of geographical data directories are briefly reviewed in Section 2.

The gds/gdc application enables a user to access spatial data stored on a remote computer system in one of the following formats: Colormap (CSIRONET 1986), GINA (Geo Vision 1986), and DLG (Geological Survey 1990). The user can display data or store data as either a GINA formatted data file, or as a postquel script for insertion into a postgres database (StoneBraker 1992). A request for data is given by specifying either the spatial area of interest (bottom right and top left corners) or the name of the file to be accessed.

This application was implemented to gain experience with the ISO Development Environment (ISODE), and to gain insight into the difficulties of accessing geographic data through a network. ISODE consists of an implementation of some network communication protocols defined by ISO[1], IEC[2], CCITT[3], and ECMA[4], and some software tools for developing applications that make use of these protocols. Of particular interest to the author was quipu, described in Section 2.1, which is an implementation of the X.500 directory service provided within ISODE.

Of the insights gained in developing the gds/gdc application two are of particular interest here: the importance of including meta-data, data about data, such as the coordinate system in which data sets are specified and the filename(s) in which the data sets are stored; and the much greater functionality that can be incorporated within a communicating interface.

The accessible datasets were given in a variety of coordinate systems, such as New Zealand Map Grid, Australian Map Grid, and latitude and longitude. The particular coordinate system used within a data set was not described in the directory (the text file). Consequently, the user was required to know what coordinate system was used for the desired data set, and to use it when specifying the region to be accessed. Specifying the data by filename was also potentially awkward because different operating systems have different conventions for the naming of files. Including meta-data such as the

---

[1] International Standards Organisation
[2] International Electrotechnical Commission
[3] Comite Consultatif International de Telegraphique et Telephonique
[4] European Computer Manufacturers Association

coordinate system or file name of a data set as part of the data descriptions within a directory is discussed further in Section 2.

Accessing geographical data sets through a communications network enables much more functionality to be provided by a communicating interface. Taken to an extreme, this functionality could include all that provided by a GIS. Initially, however, one might want the interface to search the directory for suitable data sets as well as to transfer these data sets when found. Furthermore, a user may also look for the interface(s) necessary to accomplish the data transfer. Where the user has a choice among different data sets, all of which meet the prescribed criteria, the chosen data set may be determined by the availability of interface(s) to transfer the data sets.

Rather than expand on the functionality of an interface, however, the author suggests the use of software agents, acting on behalf of users, to locate data and the associated interfaces for transferring this data. Ultimately, these agents could manage the entire transfer process including, when no interfaces are available, the construction of interfaces 'on the fly' using meta-data provided within the directory. Adoption of the software agent paradigm and the issue of interface functionality is discussed further in Section 3.

This paper is concluded by a description of a project in Section 4 that aims to develop an agency for sharing geographical data by inter alia significantly extending the author's earlier use of a geographical data directory. These extensions include developing: standard X.500 directory entries to describe geographical data sets, interfaces, and any other associated software to be used for processing geographical data sets; and specialised agents which collaborate to accomplish the goal of sharing geographical data.

2 Geographical Data Directories

Communication networks facilitate access to data sets provided the location of each data set is known or can be found. The notion of a geographical data directory has been discussed (Baker 1996, Newsome 1995, Anderson 1995, Pascoe & Penny 1993), although for different purposes. The New South Wales National Resources Data

Directory (NRDD) (Baker 1996) provides an electronic database in which there are entries describing a variety of data sets related to natural resources such as land, soil, water, catchments, estuaries, atmosphere, biodiversity and so on. Included in each description is information such as the custodian's name, an abstract of the data set's content, a description of the spatial extent, the representation in which the set is stored, and the name and contact point from whom the data can be supplied. The NRDD is similar to a directory of geographic databases provided by Manaaki Whenua - Landcare Research NZ Ltd (Newsome 1995).

Anderson 1995 discusses the development of a GIS directory for the purpose of organising the storage of resources on a computer system. This directory structure is required to: be easy to learn and use; support the needs of a multi-departmental user community; support incremental implementation and expansion; support database management and system administration functions; and reinforce the concept of a commonly shared and integrated system. In this instance the discussion is on the hierarchical (tree) structure of the directory, deciding what resources are stored in particular branches of the directory and what these branches and sub branches are called. Essentially, this is a disk directory structure and descriptions of individual resources are not of primary importance.

Another example of a geographical data directory is under the control of The National Resource Information Centre (NRIC), which was established in 1988 by the Australian government primarily to improve the information base on which policies and decisions are made for proper environmental management, land use conflicts, and issues surrounding ecologically sustainable development. The result is a geographical data directory called National Directory of Australian Resources (NDAR). The directory comprises two levels: state/territory, at which individual directories are constructed and maintained by each state/territory; and national, at which a single directory is constructed and maintained by NRIC.

Periodically, new and updated information stored in the state/territory directories is uploaded to the national level directory. This information is typically sent to the NRIC on magnetic tape or cartridge. The two level directory architecture was chosen for 3 reasons:

1. to allow data custodians in the state/territory agencies to maintain their own directory entries;
2. to allow state/territory agencies to fine tune their own directories for their own environment because not all information contained in the state/territory directories needs to be propagated up to the national level directory; and
3. a recognition that to expect the NRIC to maintain a centralised directory structure exceeded the available resources of NRIC.

National and state/territory directories are implemented using the Facility for Interrogating the National Directory of Australian Resources (FINDAR) directory software system (Johnson, Shelley, Taylor & Callahan 1996). The software comprises three main components: 'tables of attribute, keyword and spatial data for the directory entries; a gazetteer of geographic entities for spatial indexing and searching; and a thesaurus of standard terms for subject indexing and searching' (op cite). Attributes recorded for a standard data set are shown in Table 1 (op cite).

Increasingly, there is a move towards developing standards for geographical meta- data. An example is the 'ANZLIC Guidelines: Core Metadata Elements' (ANZLIC 1996), prepared by the Australia New Zealand Land Information Council (ANZLIC) Working Group on Metadata. These guidelines describe a multilevel distributed architecture, shown in Figure 2, for a geographical data directory analogous to that of the NDAR. The directory entries comprise meta-data which describe to varying levels of detail the different data sets available to users. Meta-data at a particular level of detail is contained within a page, with page 0 being the least detailed and higher numbered pages having greater levels of detail. The ANZLIC Guidelines define a core set of meta-data elements, listed in Table 2, which are believed to be needed for describing most data sets.

Although the author's implementation of a geographical data directory will be based on the ANZLIC core of meta-data elements, the directory will also describe the various software packages for transferring this data, processing this data, or both. Furthermore, the directory will be accessible for use by software packages, thereby providing the infrastructure to allow a variety of software packages on behalf of the user to locate data sets and other software packages for transferring and processing these data sets. Central to the implementation of such a directory is the X.500 Directory Service and in particular the quipu implementation of this service provided by ISODE.

Section	*Attributes*
Identification	Name, acronym, abstract, owner and other organisations associated with the dataset.
Data items	Name, description and, where applicable, spatial resolution of each item in the dataset. Items can be grouped where appropriate.
Spatial identification	Type of spatial referencing, projection, coordinate units and feature types.
Spatial coverage	General and detailed information on the area covered by the dataset.
Dataset information	Working form, working medium, size, applicable hardware and software, interchange format and supporting documentation.
Data currency	Custodian details, data collection start and end dates, dataset update frequency, future proposals, archive details.
Data lineage and quality	Data collection method, source material, data processing details, positional quality and attribute accuracy, consistency and completeness.
Ordering information	Access restrictions, output products and charges, supplier information and order procedure.
Keywords	Keywords describing the dataset suggested by person providing the entry.
Organisation/ position	Additional information about the custodian or supplier organisation/position information.

Table 1: Attributes recorded in the FINDAR directory for a standard data set

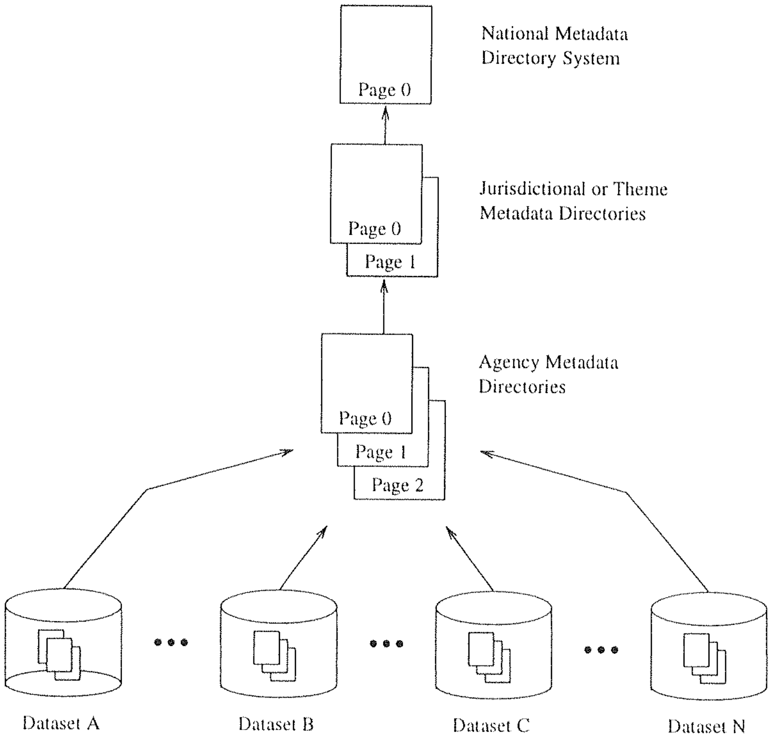

Figure 2: The ANZLIC directory architecture (ANZLIC 1996, Figure 2, page 11)

Category	Element	Comment
Dataset	Title	The ordinary name of the dataset.
	Custodian	The organisation responsible for the dataset.
	Jurisdiction	The state or country of the Custodian.
Description	Abstract	A short description of the contents of the dataset.
	Search Word(s)	Words likely to be used by a non expert to look for the dataset.
	Geographic Extent Name(s) OR	A picklist of pre defined geographic extents such as map sheets, local government areas, catchments, that reasonably indicate the spatial coverage of the dataset.
	Geographic Extent Polygon(s)	An alternate way of describing geographic extent if no pre-defined area is satisfactory.
Data Currency	Beginning date	Earliest date of data in the dataset.
	Ending date	Last date of information in the dataset.
Dataset Status	Progress	The status of the process of creation of the dataset.
	Maintenance and Update Frequency	Frequency of changes or additions made to the dataset.
Access	Stored Data Format	The format(s) in which the dataset is stored by the custodian.
	Available Format Type	The formats in which the dataset is available, showing at least, whether the dataset is available in digital or nondigital form.
	Access Constraint	Any restrictions or legal prerequisites applying to the use of the dataset, eg. licence.
Data Quality	Lineage	A brief history of the source and processing steps used to produce the dataset.
	Positional Accuracy	An assessment of the closeness of the location of spatial objects in the dataset in relation to their position on the Earth.
	Attribute Accuracy	An assessment of the reliability assigned to features in the dataset in relation to their real world values.
	Logical Consistency	An assessment of the logical relationships between data items.
	Completeness	An assessment of the completeness of coverage, classification and verification.
Contact Information	Contact Organisation	Ordinary name of the organisation from which the dataset may be obtained.
	Contact Position	The relevant position in the Contact Organisation.
	Mail Address 1	Postal address of the Contact Position.
	Mail Address 2	Aust and NZ: Optional extension of Mail Address 1.
	Suburb or Locality	Suburb of the Mail Address
	State or Locality	Aust: State of Mail Address. NZ: Opt. extension for Locality.
	Country	Country of the Mail Address.
	Postcode	Aust:Postcode of the Mail Address. NZ: Opt. postcode for sorting.
	Telephone	Telephone of the Contact Position.
	Facsimile	Facsimile of the Contact Position.
	Email Address	Electronic Mail Address of the Contact Position.
Metadata Date	Metadata Date	Date that the metadata record for the dataset was created.
Additional Metadata	Additional Metadata	Reference to other directories or systems containing further information about the dataset

Table 2: The ANZLIC core metadata elements (ANZLIC 1996, Table 1, page 5)

2.1 The X.500 Directory Service

The X.500 Directory Service (ISODE 1994) is a facility for storing and retrieving widely distributed information. The directory is hierarchically structured in the form of a Directory Information Tree (DIT) with branches of the tree stored on different computer systems and independently administered. An example of a DIT is shown in Figure 3.

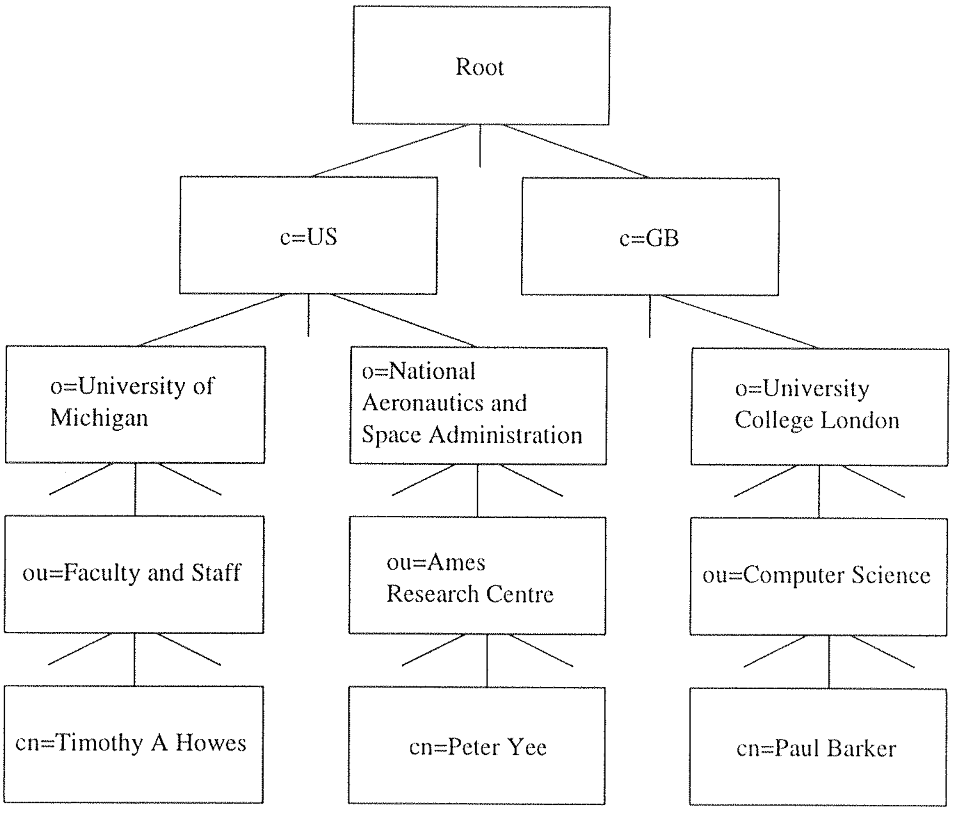

Figure 3: Example DIT (ISODE 1994, Figure 1.1, page 2)

A wide variety of information, including names, addresses, photographs, and pointers to software, can be stored within the directory. A typical example of an entry in the X.500 Directory is shown in Figure 4.

```
objectclass=top & person & 2.5.4.1 & thornObject
roomNumber = G24
2.5.4.20 = 453-5674
commonName = Timothy A Howes & Timothy Andrew Howes
photo = -ASN" 0308207b4001488001fd...
```

Figure 4: Example DIT Entry

This entry comprises five attributes, two of which, "objectClass" and "commonName", are multivalued with each value separated by the "&" Symbol.

The X.500 directory is object-oriented in the sense that entries are instances of object classes. Users define object classes by specifying a set of mandatory attributes and a set of optional attributes, perhaps using the inheritance mechanism provided by quipu. Attributes are defined by an object identifier and a syntax, which defines the types of values that an attribute can take. A person class may, for example, inherit a post office box number attribute from the class defining the organisation employing this person. In doing so, a single value of the organisation's post office box number can be shared by all employees.

The author's intention is to define an object class hierarchy for representing meta-data associated with geographical data sets. Within this hierarchy there will be instances of various application object classes corresponding to different types of applications such as agents, interfaces, and other data processing software packages. Attributes will be defined to correspond to the ANZLIC core meta-data elements listed in Table 2.

3 Interfaces or Agents?

The functionality of an interface has changed over time. In the author's earlier development of interfaces (Pascoe & Penny 1990), the function of these software packages was as 'a mechanism by which one data structure can be directly converted into another for the purpose of communication between systems or sub-systems' (van Roessel, Bankers, Connochioli, Doescher, Fosnight, Wehde & Tyler 1986). Consideration of software packages operating within a network environment lead to the definition of a communicating interface as being 'an interface that sends data to, or receives data from, some other interface through a communications network' (Pascoe & Penny 1995).

With the introduction of a network environment, the functionality of a communicating interface could be easily expanded to include other functions such as: searching for data sets, other interfaces, and more generally software packages for processing geographical data; pre-processing data to provide the user with 'thumbnail sketches' of data sets before the user commits themselves to transferring potentially large data sets; merging and filtering data during the transfer in an effort to reduce the cost and the time associated with a transfer; and so on. Given the increasing number of complex functions that could be performed by an interface, in this paper the author suggest a paradigm switch such that an interface becomes one or more specialised software agents.

Quipu, the ISODE implementation of the X.500 Directory, comprises two types of agents: a Directory User Agent (DUA), which helps users to formulate queries, wrap them in the required protocol, pass them on to directory system agents, and display the results obtained; and a Directory System Agent (DSA), which responds to DUA directory searches either directly by supplying information contained within the directory branch under the DSA's control, or by passing the search on to other DSAs managing other parts of the directory.

In the context of the research described here, a software agent is a package that acts on behalf of a user to achieve some task according to a priori knowledge of the user's preferences. For example, a user may send an agent 'to collect all telephone and power line data for the city of Dunedin, New Zealand, and to store this data in a text file conforming to the SDTS vector profile'. The author's expectation is that such a task will be accomplished by collaboration among several specialised agents including DUAs, DSAs, a variety of agents designed to transform geographical data among different representations, and so on. All these agents will be described in the geographical data directory and will collectively form an agency for sharing geographical data.

4 An Agency for Sharing Geographical Data

The primary objective of research described in this paper is to describe the author's approach to developing an agency for sharing geographical data. This agency will comprise a wide variety of software agents responsible for subtasks associated with collecting geographical data sets such as locating data sets, transforming data sets into the desired representation, and merging and filtering data sets during collection. Also included within the agency will be 'managerial' software agents who are responsible for locating other agents within the agency, and co-ordinating the interactions among the various agents and between the agency and the user.

References

Anderson, J. S. 1995, 'Building a Useful GIS Directory: Snohomish County, Washing ton', *URISA Journal* pp. 45-52.

ANZLIC 1996, Anzlic guidlines: Core metadata elements version 1, Technical report, Autralia and New Zealand Land Information Council: Working Group on Metadata.

Baker, H. 1996, 'Data Directory Blazes Trail for Metadata Standards', *The Australasian Geographic Information Systems Applications Journal, GIS User* (15).

CSIRONET 1986, *Command Driven Colourmap, User's Guide*, 1st edn, CSIRONET Graphics System Section, Canberra.

Geo Vision 1986, *Data Translation Guide*, GeoVision, Ottawa.

Geological Survey, U. S. 1990, Digital Line Graphs from 1:2,000,000-Scale Maps, in 'Data Users Guide', Vol. 3 of *National Mapping Program Technical Instructions*, Department of Interior, U. S. Geological Survey.

ISODE Volume 8 User's Guide Directory Services 1994, Technical report, The Dome The Square Richmond TW9 1DT UK.

Johnson, B. D., Shelley, E. P., Taylor, M. M. & Callahan, S. 1996, The findar directory system: a meta-model for meta-data.
*http://www.nric.gov.au/nric/publishing/papers/metadata.html

Newsome, P. F. J. 1995, Directory of Geographic Databases within Manaaki Whenua - Landcare Research, Technical report, Whenua - Landcare Research New Zealand Ltd., Private Bag 11052, Palmerston North New Zealand.

Pascoe, R. T. & Penny, J. P. 1990, 'Construction of interfaces for the exchange of geographic data', *International Journal of Geographical Information Systems* **4**(2), 147-156.

Pascoe, R. T. & Penny, J. P. 1993, Transforming geographic data between different concrete representations, *in* G. Gupta, G. Mohay & R. Topor, eds, 'Proceedings of the Sixteenth Australian Computer Science Conference', pp. 653-663.

Pascoe, R. T. & Penny, J. P. 1995, 'Constructing interfaces between (and within) geographical information systems', *International Journal of Geographical Information Systems* **9**(3) 275-291.

StoneBraker, M. 1992, *Postgres Reference Manual*, Version 4.0, University of California, Berkeley. This manual is distributed with the Postgres source code.

van Roessel, J., Bankers, D., Connochioli, V., Doescher, S., Fosnight, G., Wehde, M. & Tyler, D. 1986, vector data structure conversion at the EROS data center, final report, phase I, Technical report, EROS Data Center, Sioux Falls, South Dakota 57198.