# A farm decision support GIS: interoperability problems and solutions

**Marion Garton, George Taylor**
School of computing, University of Glamorgan
Pontypridd, CF37 1DL, UK
e-mail: mgarton@uclan.ac.uk, getaylor@glam.ac.uk

## Introduction

This paper describes the interoperability issues to be addressed in establishing a GIS as the major component in a farm decision support system and as a teaching resource for agricultural staff and students. The diverse data sets containing descriptions of current and historic land use are considered, and suggestions are proposed for their successful incorporation through the GIS. Access to datasets from external sources is explored with an assessment of the difficulties encountered by novice users.

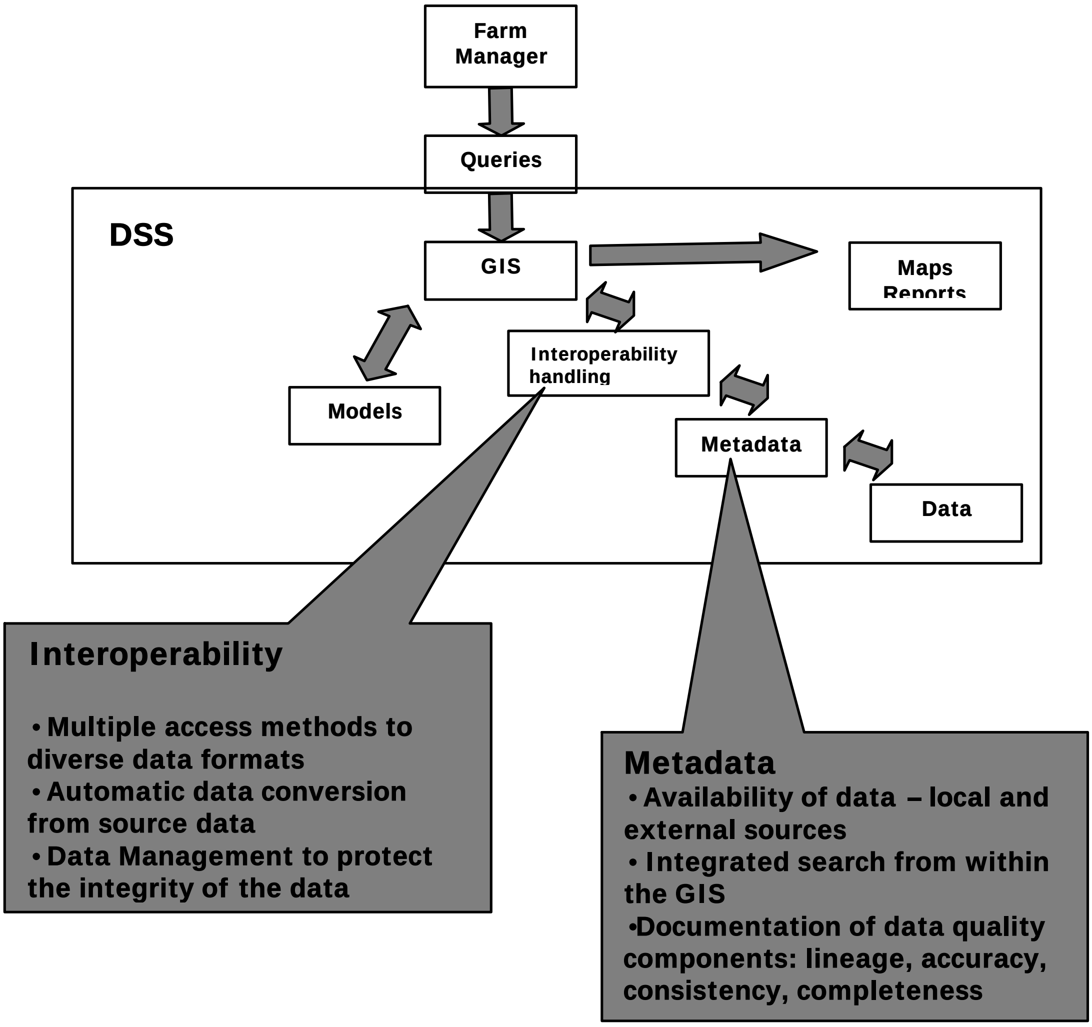



**Figure 1** Fundamental Data Issues within the DSS

The development of a prototype GIS for the farm manager at the University of Central Lancashire's farm in Cumbria, and as a teaching resource for staff and students on the campus, raises a number of data issues. This paper focuses on interoperability, an important consideration when users must access more than one dataset using distributed computing resources (Sondheim 1999). The farm manager, staff and students wish to explore different scenarios for land allocation. A GIS lying at the heart of a decision support system (DSS), is seen as the means of integrating data, linking to appropriate modelling packages and presenting results in a clear format. A high level of customisation will be required so that the use of the GIS is intuitive to novice users (Maguire 1999). Figure 1 shows a schematic representation of the proposed system, indicating the importance of data within the DSS.

## Interoperability

Within a single organisation most users access corporate datasets and software applications that meet the organisation's business needs. Such integrated systems are based on a limited set of data models and are implemented through common technology (Sondheim 1999). As computing environments have developed, users have become more sophisticated in their use of IT, increasing the demand for open systems, where it is possible to move easily between systems supplied by different vendors. Sondheim (1999) states that the two significant challenges that must be met to ensure basic interoperability are:

"

1. *the autonomous systems must be able to exchange data and to handle queries and other processing requests;*
2. *they must be able to make use of a common understanding of the data and requests."*

The drive for interoperability has concentrated on defining standards for network communications to meet the first requirement and standards for application programming interfaces and the transport of objects across networks to meet the second.

Users of the prototype GIS require access to datasets from a range of sources. Some of the digital datasets are stored on the campus network, others on standalone systems. The use of data locator services may also locate relevant datasets available from external data providers. Specialised environmental management systems and spreadsheets used for financial modelling are potential applications to be linked through the GIS. Interoperability is thus an important consideration in the development of the prototype.

## Data Analysis

Data analysis, undertaken as part of the development process, identified datasets that already exist on campus and that will prove useful in supporting the farm manager in arriving at a well-reasoned decision for future land-use options (Table 1). The datasets form the basis for tasks such as the production of maps of current and past land usage so it is essential that the GIS integrate them.

| Dataset | Format | Computer Platform |
|---|---|---|
| Optimix: historical field records | dBase (.dbf, .cdx) | Standalone PC, single licence |
| Agridata: current records relating to stock and land parcels | Paradox (.dbf, .cxx) | Standalone PC, single licence |
| Annotated farm maps | paper | |
| Ordnance Survey tiles | MapInfo tables | NT network |
| Field boundaries | MapInfo tables | NT network |
| Soil analyses | ArcView | NT network |
| Aerial photographs | ArcView image | NT network |
| Oblique photographs | paper | |
| Pollution map | paper | |
| Habitat surveys | paper | |
| Accounts | Excel spreadsheet | Standalone PC |

**Table 1** Existing datasets on campus

Although individuals may develop their own datasets, the corporate datasets such as the field boundaries must be maintained centrally within the organisation, ensuring that everyone has access to up to date datasets.

### Access to External Datasets

Users of the proposed GIS have indicated they may wish to search for datasets, other than their own, that may prove useful to their studies. Though some of the datasets will be maintained on campus it is highly probable that useful data will exist in external organisations. To locate and access a relevant dataset a user needs (Longley et al. 2001):

?? facilities for search and discovery;
?? information to be able to “assess the fitness” of a dataset;
?? instructions for handling the dataset effectively;
?? information on the contents of the dataset.

This data about data is known as metadata and comprises summary information about a dataset describing, for example:

?? *what* the dataset contains;
?? *who* owns the dataset and *who* is allowed to access or update the data;
?? *where* the geographic objects covered by the dataset are located in space;
?? *how* the data was collected and stored and *how* it can be accessed;
?? *when* the data was collected (Guptill 1999).

This summary information implies three levels of metadata: discovery metadata, exploration metadata and exploitation metadata (National Geospatial Data Framework (NGDF) 2000**).**

Discovery metadata provides the user with sufficient information to discern the content, format and scope of a dataset. Many digital libraries (geolibraries), including Microsoft’s Terraserver and NGDF’s AskGiraffe service (FitzGibbon 2000) already exist and can be searched over the Internet (Longley et al. 2001).

Once a dataset has been discovered, the user must assess its suitability to satisfy the user's requirements. This "fitness for purpose" is referred to as data quality. The Association for Geographic Information (AGI) has produced a set of guidelines for describing geographic information content and quality (Parker et al. 1996, section B) and identify five aspects of data quality:

- completeness;
- thematic accuracy;
- temporal accuracy;
- positional accuracy;
- logical consistency.

Exploration metadata should contain sufficiently detailed descriptions of the five aspects of data quality to enable the user to assess the suitability of a dataset.

The third level of metadata relates to the process of obtaining and using a dataset. This may contain information relating to the source of the data and restrictions on use (Parker et al. 1996, section C). Technical details such as the format of the data will guide the user in selecting datasets that are compatible with the user's current system.

## Implications for the Design of the Prototype

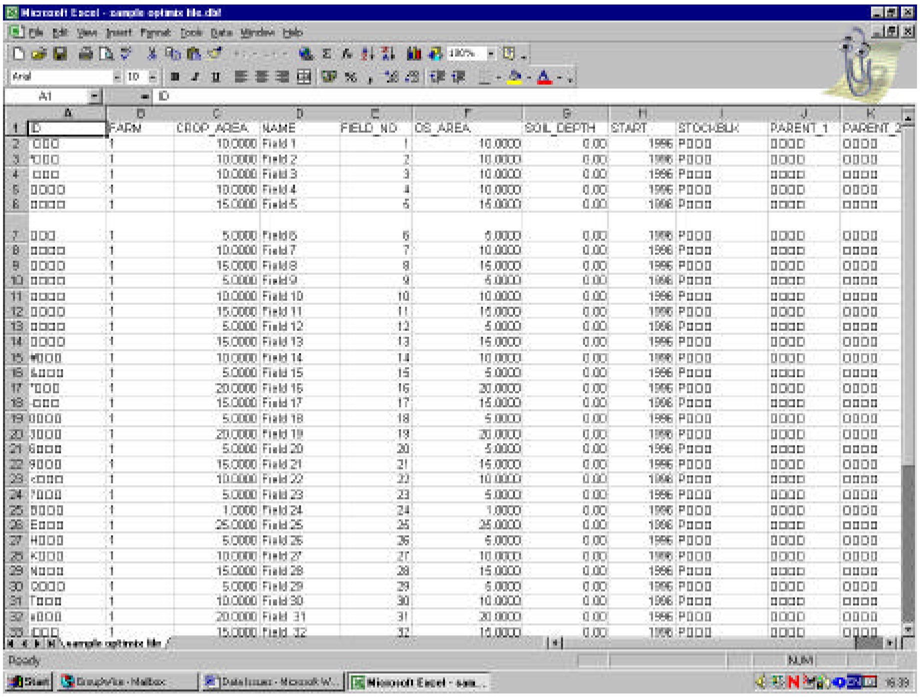
To locate external datasets it will be necessary to access metadata catalogues via the web. To meet this requirement the prototype must offer a facility to access the web from the GIS, search metadata catalogues and possibly to download datasets. This will be an important area whose potential will be explored in the development of the prototype.

The operating systems found on campus include Novell Netware for the network and Windows NT on workstations. Windows NT is also installed on the standalone PCs in the farm office. GIS packages on campus include MapInfo, ArcView 3.2 and Idrisi. The level of GIS expertise is limited, though some staff have developed basic skills in MapInfo. Although initial work in digitising field boundaries was undertaken in MapInfo, the prototype GIS will be developed in ArcView.

## Specific Interoperability Problems

For many users, the prime function of the prototype will be to provide historical and current information relating to land usage on the farm. Most of the important datasets will be local, collected by the farm manager and maintained in the farm office. Some national statistics from the Department for Environment, Food and Rural Affairs (DEFRA), formerly the Ministry of Agriculture, Food and Fisheries (MAFF), describing, for example, land usage and cattle and sheep holdings, will be of relevance to agricultural students, particularly in the wake of Foot and Mouth. It is anticipated that other useful datasets will be identified through data locator services via the Internet and that these will raise yet more integration requirements. This section introduces specific interoperability problems that must be addressed during the development of the prototype with areas of concern including data integration and data access.

Between 1996 and 1999 Optimix, (Farmplan Computer Systems, 1996), was used on a standalone PC in the farm office to hold records of local data relating to soil analyses, applied nutrients, fertilisers, crop varieties and yields per field. An investigation of the file directories indicated that the Optimix database created dBase files (.dbf and .cdx). However, these files do not import into packages such as Excel, MapInfo or ArcView with complete success. Figure 2 shows that the contents of some data fields were not converted correctly and are unreadable.



| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ID | FARM | CROP_AREA | NAME | FIELD_NO | OS_AREA | SOIL_DEPTH | START | STOCKBLK | PARENT_1 | PARENT_2 |
| 2 | □□□ | 1 | 10.0000 | Field 1 | 1 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 3 | '□□□ | 1 | 10.0000 | Field 2 | 2 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 4 | □□□ | 1 | 10.0000 | Field 3 | 3 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 5 | □□□□ | 1 | 10.0000 | Field 4 | 4 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 6 | □□□□ | 1 | 15.0000 | Field 5 | 5 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 7 | □□□ | 1 | 5.0000 | Field6 | 6 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 8 | □□□□ | 1 | 10.0000 | Field 7 | 7 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 9 | □□□□ | 1 | 15.0000 | Field8 | 8 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 10 | □□□□ | 1 | 5.0000 | Field9 | 9 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 11 | □□□□ | 1 | 10.0000 | Field 10 | 10 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 12 | □□□□ | 1 | 15.0000 | Field 11 | 11 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 13 | □□□□ | 1 | 5.0000 | Field 12 | 12 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 14 | □□□□ | 1 | 15.0000 | Field 13 | 13 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 15 | #□□□ | 1 | 10.0000 | Field 14 | 14 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 16 | &□□□ | 1 | 5.0000 | Field 15 | 15 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 17 | '□□□ | 1 | 20.0000 | Field 16 | 16 | 20.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 18 | -□□□ | 1 | 15.0000 | Field 17 | 17 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 19 | 0□□□ | 1 | 5.0000 | Field 18 | 18 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 20 | 3□□□ | 1 | 20.0000 | Field 19 | 19 | 20.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 21 | 6□□□ | 1 | 5.0000 | Field 20 | 20 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 22 | 9□□□ | 1 | 15.0000 | Field 21 | 21 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 23 | <□□□ | 1 | 10.0000 | Field 22 | 22 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 24 | ?□□□ | 1 | 5.0000 | Field 23 | 23 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 25 | B□□□ | 1 | 1.0000 | Field 24 | 24 | 1.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 26 | E□□□ | 1 | 25.0000 | Field 25 | 25 | 25.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 27 | H□□□ | 1 | 5.0000 | Field 26 | 26 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 28 | K□□□ | 1 | 10.0000 | Field 27 | 27 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 29 | N□□□ | 1 | 15.0000 | Field 28 | 28 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 30 | Q□□□ | 1 | 5.0000 | Field 29 | 29 | 5.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 31 | T□□□ | 1 | 10.0000 | Field 30 | 30 | 10.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 32 | «□□□ | 1 | 20.0000 | Field 31 | 31 | 20.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |
| 33 | Z□□□ | 1 | 15.0000 | Field 32 | 32 | 15.0000 | 0.00 | 1996 | P□□□ | □□□□ | □□□□ |

**Figure 2:** Optimix data fields

Agridata, a database package that stores records relating to cattle, sheep and land parcels, has now replaced Optimix for recording local data. Discussions with a representative from Agridata confirmed that the data is held in Paradox database format (.dbf, .cxx). As with Optimix, the files were not successfully imported by other packages. Figure 3 shows the result of importing one file within the database into Excel.

Agridata is installed on a standalone machine with a single user licence, leading to a number of specific problems that must be addressed by the prototype.

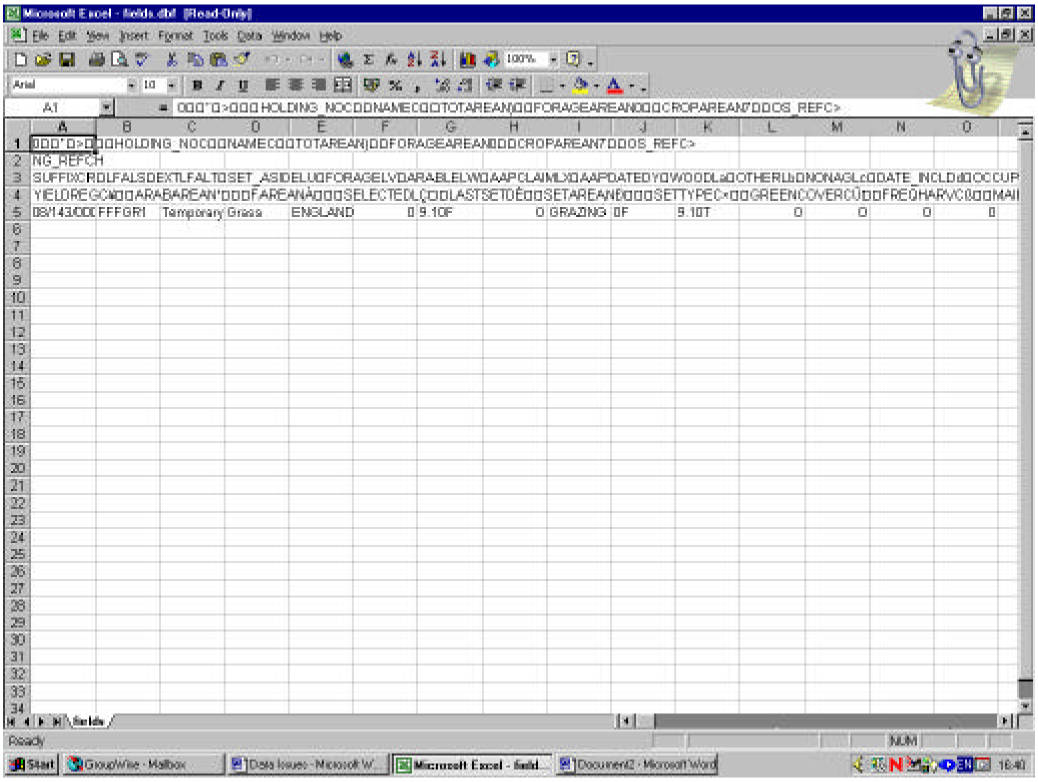


**Figure 3:** Agridata Field Data

When submitting grant applications, for organic accreditation for example, the farm manager must provide an up to date map with associated field records for the current year and also the previous five years (MAFF 1999). As indicated, these records are held in Optimix and Agridata and, to assist in this task, the GIS must combine data from the two applications and link to polygons that were digitised by tracing the field boundaries in MapInfo. The field boundaries themselves may have changed over the period and so a separate map will be required for each year of historical data. Academic users will have similar requirements when exploring the impact over time of changing farming practices.

## Data Access

The prototype GIS will be available from the campus network for the farm manager, students and teaching staff. Where possible the datasets will be stored on network drives, offering security against unauthorised amendments by users. A major problem to address however is the access to datasets such as Agridata on standalone machines.

The current implementation of Windows NT on campus allows data to be shared through "Fetch boxes". Users connect to a Fetch box belonging to a second user and through this connection have access to files the second user has made publicly available. This situation is not ideal as users can only connect to one Fetch box at a time. If datasets are not on a central network drive there will be a considerable management overhead on the GIS ensuring users connect to the correct Fetch box holding the required data source.

In addition, the GIS may require access to relevant data sets and models from other sources, including Rights of Way data for Cumbria, results of habitat surveys, pollution levels and other sources that data locator services may raise. The seamless integration of these disparate sources is a further challenge to address.

The Library and Learning Resource Service (LLRS) at the campus has recently subscribed to Digimap, a service that provides access for academic users to Ordnance Survey (OS) maps. Maps may be downloaded and used to enhance teaching, learning and research. Digimap was used to complete the digital map of the farm, which will form the base for the GIS. The process involved a number of stages that would prove beyond the skill of a novice user. As an example, the steps necessary to acquire a Landline tile included:

?? Identification of OS tiles. If the tiles are not part of the set currently available through Digimap, the user must request them through the site representative;
?? Downloading tiles. The files are zipped and downloaded by the user. Once unzipped, the OS tiles are in NTF format and must be translated into a format that can be read by the GIS, in this example, MapInfo;
?? Locating a translator. A translator was located on the Digimap site, downloaded and installed on a laptop. (An internal user policy prohibits users from installing software on the network);
?? Conversion of NTF tiles. The NTF tiles were converted into a MapInfo Interchange Format (MIF) file. This was finally imported into MapInfo and saved as a MapInfo table.

Users with little IT skills cannot tackle such procedures. The example does however highlight interoperability problems. Further problems acquiring data sets from external sources will be identified as the prototype is developed.

## Interoperability Strategies

The development of the GIS requires a network platform with common datasets stored on drives accessible to all users on a read-only basis. Other access rights are required for users who own and are responsible for maintaining datasets. The following sections consider how data can be accessed on the network.

Many database and GIS packages offer Open Database Connectivity (ODBC), where data in one application may be accessed by a second application (Lorents 1998). This connectivity allows data in the user's database to be updated and saved through commands in the second application. MapInfo offers ODBC, linking to databases such as Access and Oracle. Some processing is necessary initially to add spatial co-ordinates to a data table to make it "mappable", that is, to allow the table to be displayed in a Map window (MapInfo 1998). ArcView 3.2 also provides a database connection feature, which creates a table containing records from the database. Direct editing is not permitted however. The data can be added to a map by joining it to the attribute table of a theme (ESRI 1996).

MapInfo and ArcView both support data from outside sources such as dBASE (.dbf) and delimited ASCII text files. As shown earlier, data in Optimix and Agridata were not imported with a 100% success. An automatic mechanism must be developed to convert the data from the database packages into a format that can be read by the GIS software.

Use of a common format to transfer data between systems is a two-stage operation. Data from a file in a given format is converted into an intermediary file in a common format, and then translated to a third file in a compatible format (Sondheim 1999). This “bridge” however must be totally transparent to the user of the GIS and will require customisation (Maguire 1999).

The Optimix system offers reporting facilities and can export data in Comma Separated Variable (csv) format. As the data is now historical, it will be possible to create new datasets in the GIS that are based on the Optimix records, a conversion that would be undertaken only once. The translated files can be stored on the network and accessed by the GIS with the new datasets would available to all.

Live datasets such as Agridata, installed on standalone machines, pose the greater problems. Firstly, there is the management issue of who will upload data to the network, how often and when. To reduce inconvenience to the dataset owner, it would be desirable to connect the standalone PC to the network and offer a facility to upload data to the network. This would remove the need for finding a suitable medium for transferring data should the datasets prove large.

Once the dataset is stored on the network, the second difficulty of converting data into a format compatible with other software applications must be tackled. The development of a programming solution, invoked from within the GIS, will be one option explored in the development of the prototype. MapInfo and ArcView have their own associated programming languages. Later versions of the software allow direct data manipulation through languages such as Visual Basic.

## Summary

This paper has represented data as the major component of a prototype system designed to support users in the exploration of land use allocation. The system must allow easy access to data available on a range of computer platforms, both internal and external to the organisation. The GIS must permit the transfer of data and commands between applications through the design of interfaces acting as bridges (Djokic 1993). A careful consideration of the interoperability of systems will ensure data integrity by introducing automatic conversion routines for the successful transport of data between systems, access to common data sets and, with sound management of the selected computer platform, protection of data from accidental changes. The ability to access external datasets in the future has indicated the potential use of metadata catalogues in locating data sources and assessing the relevance and quality of datasets. The main challenge of this research topic is to develop a prototype capable of searching and integrating data between applications without direct intervention from users.

## References

Djokic D (1993), Towards General Purpose Spatial Decision Support System Using Existing Technologies, Centre for Research in Water Resources, Texas.

ESRI (1996), Using ArcView GIS, Environmental Systems Research Institute Inc., Redlands, USA.

Farmplan (1996), Business Planner for Agriculture, Farmplan Computer Systems, Gloucestershire.

FitzGibbon P (Ed.) (2000), Ask Giraffe. In: GEOEurope, GEOTec Media, Issue 9, p20.

Guptill, S C (1999), Metadata and data catalogues. In Geographical Information Systems, Principles and Applications. P. Longley, M. Goodchild, D. Rhind and D. Maguire (eds.), J. Wiley & Sons. Vol. **2:** 677-692.

Longley P A, Goodchild M F, Maguire D J, Rhind D W (2001), Geographic Information Systems and Science, John Wiley & Sons Ltd, Chichester, England.

MAFF (1999), Organic Conversion Information Service Pack, HMSO.

Maguire, D J, (1999), GIS customisation. In Geographical Information Systems, Principles and Applications. P. Longley, M. Goodchild, D. Rhind and D. Maguire (eds.), J. Wiley & Sons. Vol. **1:** 359-369.

MapInfo Corporation (1998), MapInfo Professional User's Guide, MapInfo Corporation, Try, New York.

NGDF (2000), NGDF Discovery Metadata Guidelines – Version 1.2, National Geospatial Data Framework (NGDF) Management Board.

Parker D, Buchanan H, Hoult C, Taylor G, Coombes M (1996), Guidelines for geographic information content and quality, Association for Geographic Information, London.

Sondheim M, Gardels K, Buehler K, (1999), GIS interoperability. In Geographical Information Systems, Principles and Applications. P. Longley, M. Goodchild, D. Rhind and D. Maguire (eds.), J. Wiley & Sons. Vol. **1:** 347-358.