CONVEYING OBJECT-BASED META-INFORMATION

by

Peter F. Fisher

Midlands Regional Research Laboratory
Department of Geography, University of Leicester
Leicester LE1 7RH, United Kingdom

pff1 @ leicester.ac.uk

ABSTRACT

Metadata and lineage are two related areas of research which have received considerable recent attention from the GIS community, but that attention has largely focused on meta-information at the level of the image, coverage, or data layer. Researchers working in other areas such as error and uncertainty handling, have focused on a lower level within the spatial data, but can also be considered to have been working on metadata. Users require access to the whole set of metadata from the level of the mapset to the elemental object, i.e. the point, line, polygon or pixel. This paper attempts to draw these different research strands together and suggests an umbrella framework for metadata which can be translated to a number of different flavors of metadata for which accuracy, lineage, statistics and visualization are exemplified. This leads to discussion of an interface design which enables rapid access to the metadata.

INTRODUCTION

Meta-Information or Meta-Data have been defined merely as data about data (Lillywhite, 1991). It is therefore rather surprising to find that two relatively restricted areas which fit this definition have been the focus of most research associated with metadata. One line of investigation has come from the database community and focuses on datasets as a whole; the properties and contents of each dataset being the prime information of interest (Medjyckyj-Scott et al. 1991). This research is epitomised by such systems as, for example, BIRON (Winstanley, 1991), FINDAR (Johnson et al. 1991), MARS (Robson and

Adlam, 1991) and, more recently, GENIE (Newman et al. 1992). On the other hand, lineage has been addressed by Lanter (1991) who developed the GeoLineus system to track coverage transformations in Arc/Info, an approach which is being extended to other systems. Combined with error reports it is also possible to use the lineage information to propagate overall errors (Lanter and Veregin, 1992). While these two lines of research have been identified as associated with metadata and dominate that literature, other research is being done, or the need for it is realised.

For example, under the NCGIA initiatives on the accuracy of spatial databases (I-1) and the visualization of spatial data quality (I-7) much work has been done on the error measurement, reporting, modeling, propagation, and visualization. This is all to do with metadata, since accuracy information is not usually the data, it is about the data. The most recent initiative is on the Formalization of Cartographic Knowledge (I-8), and is among many other Artificial Intelligence developments in GIS. The result of reasoning from an AI system is new data, but the process of reasoning is metadata.

The purpose of this paper is twofold. First is to suggest an ***umbrella*** framework for spatial metadata within GIS which encompasses all the strands of research outlined above as well as others which are not mentioned and which may yet emerge. The second theme to the paper, is to examine how these may be offered to a user for exploration and interrogation, both as cartographic and as textual presentations.

A HIERARCHY OF OBJECTS IN A SPATIAL DATABASE

Moving from the most general to the most specific, there is a logical hierarchy of objects in the spatial database. This hierarchy is recognised in the design of many systems, but it is necessary to recall it for the present discussion. The hierarchy is listed in Tables 1 and 2 where it is given in order from the most general to the most specific.

The Mapset is a collection of data about a particular area. Usually all members of the mapset are to a common projection (itself one of the metadata items of the mapset), and covering an identical area (although this is not essential). It is the highest level of metadata which includes spatial and aspatial information. One essential item of metadata at this level is the contents of the next level in the hierarchy, the data layers.

The Data Layers (based on a layer-based system but also transferrable to an object-oriented one) are the individual themes of information. They are variously referred to as the image, the raster, the quad, the coverage, etc. in different systems. Arguably in a true object-based system the Mapset and the Data Layer are one and the same level, but in most current systems they are distinct.

The Object Classes is the individual theme on the thematic map. So in a soil Data Layer, the Object Classes would be the particular soil map units, while in a census area data layer, the object class would be the Enumeration District, or Census Block. In a topogaphic layer it would include all the features types present.

The Spatial Objects are the elemental spatial features which GIS are designed to work with. These may be points, lines, areas or pixels. Surface data being usually coded as lines, points or pixels. Commonly one derivative data layer can be the metadata for a source layer.

TABLE 1

RAW OBJECTS

Object-Types:	**Examples and Some Key Characteristics**
The Mapset	The overall group of data - the study area
The Data Layer	A theme over the mapset area - Elevation - Soil data - Census areas
The Object Class	Individual map themes - Soil type X - Wetlands - Public Buildings - Each Block
The Spatial Object	Objects actually displayed - Point - Line - Polygon - Pixel

The four levels recognised here are identifiable in a number of different versions.

METADATA EXAMPLES

Some types of metadata can be tracked throughout the hierarchy identified above, and the examples of *accuracy*, *lineage*, *statistics* and *visualization* are discussed.

Accuracy

Accuracy can be tracked from the highest level to the lowest. At the *mapset* level it refers to the accuracy of the global coordinates to which the mapset is registered, while at the *data layer* level, it refers to the accuracy of any spatial transformations the layers may have undergone to become registered to the mapset area and projection; both are measures of positional accuracy. Occasionally all data layers will have the same error reports and so it may be recorded at the mapset level. At the data layer level attribute accuracy may also be recorded as the Percentage Correctly Classified, the Overall Accuracy, or the Root Mean Squared Error. These are all different broad measures of accuracy used for different data types including soils, remote sensing/land cover and elevation.

TABLE 2

ACCURACY METADATA

Object-Types:	**Examples and Some Key Characteristics**
The Mapset	- Coordinates of control points on some wider-area reference system (long/lat)
The Data Layer	- Percent Correctly Classified (overall accuracy) - RMS Error
The Object Class	- Producer and User accuracy - Map unit inclusions
The Spatial Object	- accuracy of classification - positional accuracy - accuracy of area estimates

The *object class* may be associated with a number of different types of accuracy measures. In soil data this may include the types and frequency of map-unit inclusions (Fisher, 1991), and in land cover data it may include the producer and user accuracy (Campbell, 1989). Among the *spatial objects* in any polygon coverage there will be a unique value for the accuracy of the area of the polygon, while lines may have a unique value for the accuracy of the length of the line. In a classified remotely sensed image the fuzzy memberships of each and every pixel belonging to all land cover types may be recorded, although these have to be considered as a data layer in their own rights as well as metadata to the usual hard classification. Accuracy in the *viewed objects* is very poorly researched, but may include a degree to which human observers may confuse the symbolism used for one object class with that used for another object class in the display space. Metadata on the viewed spatial objects may report the degree to which the object has been deformed in the process of cartographic representation (the number of points removed, the ratio of the length shown and the actual length, etc.)

Lineage

Spatial datasets are derived by some process, either by analysis of other data layers or by digitizing. The history of these transactions and events is the lineage of the data. At the level of the *data layer*, Lanter (1991) has given a thorough treatment of lineae issues, particularly developing Geolineus, a system which uses history files to produce the events in the lineage of any *data layer*, and incorporating some further measures such as error propagation (Lanter and Veregin, 1992). Various widely available GIS also give some measure of lineage of *spatial objects*, or can be forced to. EPPL7 and GRASS, for instance, enable point-and-click presentation of spatial objects (pixels) in multiple data layers, although the default method of presentation in both is merely to display one layer and list the values in other, named layers, which may be related by lineage or just by being of the same area. The user needs to know the lineage of any data layer for this method to work. Lanter and Essinger (1991) illustrate a more automated version of this, but again no impression of lineage is given.

As in accuracy, lineage is a multilevel concept within the hierarchy. The *mapset's* lineage may include the subsetting of the basic layers from some other dataset at some other projection (in a very diverse *mapset*, this information may be specific to the *data layer*). Data layers need to record how they are individually derived, and so for data layers at the root of a lineage tree, this may include the name of the digitizer, and the scale of the paper map. In derived data layers they it records progressively the parent data layers.

By cross referencing to the parent data layers and the conditions used to execute the transformations, it is possible to present the user with lineage information on the *object classes*, and *spatial objects*. With object classes the lineage data should report the threshold values which transformed other classes in the derivation of the current class, while for spatial objects the user should be presented with parent values at that location.

The lineage of the *viewed object* identifies any transformations made to either object classes or spatial objects which have been executed for the purpose of display only; the change from a numeric value to a color on a screen (in some systems this is a hard numerical transformation, in others it is programmed, but in still others it may be specified).

Lineage as the term is used here refers to logical or mathematical transformations of data. It is therefore not dissimilar to logical arguments, and the logical inference of an expert system may also be viewed in the same way. Indeed Lanter's (1991) flow diagram presentation of lineage (see also Berry, 1987), is very similar to the presentation of goals and rules in a graphic representation of an expert system's logic, and is one of the best ways to show a user the reasons for an expert system's judgement.

Statistics

Statistical reports should perhaps be regarded in the same general framework as the examples of metadata discussed above; going back to the definition of metadata statistical summaries are information about the data. Indeed, many of the accuracy measures widely recognised as metadata are statistical summaries of the hierarchical level. A consideration of the hierarchies presented should lead to an appreciation that there are many other summary statistics which can be presented at the different levels identified. For example, the correlations and regressions between different data layers is metadata at the level of the *mapset*, the spatial autocorrelation is at the level of the *data layer*, the *object class* or even the *spatial object* (depending on the measure used).

TABLE 3

STATISTICAL METADATA

Object-Types:	**Example Visualization Metadata Types**
The Mapset	- Inter-layer correlation matrix
The Data Layer	- Spatial autocorrelation
The Object Class	- Join counts statistic
The Spatial Object	- local correlation

Furthermore, and less commonly recognised, such summary statistics all have equivalents in the *viewed objects*. The measures may be similar but the values will frequently be very different. Arguably one specific objective in cartographic reproduction is to match such summary statistics between the viewed objects and the raw data (Dykes, in press). Thus a very important item of metadata can be the degree of match between the values in the raw data and in the viewed objects.

Visualization

The visualization of the raw data has a hierarchical set of unique metadata, whether on a screen or on paper. These are exemplified in Table 4. At the level of the Mapset there are properties of the projection used for display, and the characteristic pattern of distortion. It may need to be recalled that the display projection is not necessarily the same as the digitizing projection or the projection for storage, and that final projection necessarily introduces its own artifacts into the data.

The name of the palette used to color a data layer, and its properties are crucial to how the user views the data; how they see it. In a situation where many windows may be in use at once, many different palettes may be in use. This is independent of the object class (legend) since that can be populated with any palette. The reasons for the use of the palette may be no more than it being the default, but they may be deeper.

The reasons for each object class being painted the color or using the symbol shown may be metadata at the next level. Some measure of the degree to which classes may be confused would also be useful information assisting design.

The mapped data groups items by class, so all rivers are shown as blue lines. The actual names of those rivers are then metadata, and a point and click system can give the user access to the source information. Where many different groups are classed as one for mapping, that too can be presented to the user as metadata. In another sense the algorithm used for generalizing some feature for display may be shown, and indeed literally the original detail of the feature is metadata to the visualization.

TABLE 4

VISUALIZATION METADATA

Object-Types:	**Example Visualization Metadata Types**
The Mapset	- The map projection used for display - Distortions caused by the projection
The Data Layer	- colour palette in use - reasons for using this palette - objectives of the design
The Object Class	- reasons for the color/symbol selection - confusion between classes
The Spatial Object	- names and actual classes of the features shown - the number of points omitted in generalizing a line

A USER INTERFACE

A graphic user interface is proposed to enable access to the different levels and types of metadata. Figure 1 shows the broad structure of such an interface in the display space there are a number of different live areas, each of which will produce a response from the system giving metadata on a particular object within any hierarchical class.

A menu of metadata types of which the system is aware needs to be shown. System awareness could be enabled by either documentation files, history files or history fields within documentation files, and the like. How this is enabled is not the subject of this paper. Selecting from this menu would cause changes in the actual metadata displayed. Within a full GIS, this could be a pull-down metadata menu or sub-menu from the main system menu, it would therefore be another aspect of the GIS functionality, presumably in the area of system control.

We then have a set of 4 active areas (possibly distinct windows) each showing an active area with or without subareas, and clicking within these areas will cause different levels of metadata to be displayed. Area 1 is the active area for the mapset, and Area 2 is for the data layer. The object classes corresponding to the data layer presented in Area 2 are then shown in Area 3, and the corresponding spatial objects are shown in Area 4. In these last two areas the individual themes and objects are themselves active and metadata on those can be retrieved by pointing and clicking. The metadata-type selected from the metadata menu is then displayed. A further menu may offer alternative methods for displaying the metadata. Thus error may be displayed in textual or many cartographic forms.

Multiple triplets of data-layer/object-class/spatial-object windows could exist within a mapset, and be toggled in a seventh window. As currently envisaged the display would only show one data layer at any time, but this is not a necessary restriction in a multiwindowing operating system.

CONCLUSION

In this brief paper it has been possible to present an argument which suggests that it is necessary to develop a complete view of metadata. Users wishing to find all the information about their data (at whatever level) can only be satisfied by such a complete view being taken. The nuts and bolts of this are essential, are being actively researched by many groups, and implemented in commercial systems.

A number of important issues do emerge from the discussion:

1. the umbrella structure for metadata is suggested as a necessary recognition;
2. the large number of possible different types of metadata;
3. the extent to which data, with its own metadata, are themselves metadata for derivative data layers must also be recognised;
4. the need for a consistent interface giving access to all this metadata should become an important part of any system.

Knowledgable users are aware of the number of different pieces of information which are available for data sets they are manipulating. They may often be frustrated by failure to access that information. Indeed, arguably all users deserve access to this information, since it is exactly this information which informs the user as to the suitability of the data and the analysis they have performed. But that access needs to be through a consistent and complete interface design itself based on a complete recognition of the metadata.

REFERENCES

Berry, J., 1981. Fundamental operations in computer-assisted map analysis. *International Journal of Geographical Information Systems*, Vol. 1, pp. 119-136.

Dykes, J., in press.

Johnson, D., Shelly, P., Taylor, M., and Callahan, S., 1991. The FINDAR Directory System: a meta-model for metadata. In Medjyckyj-Scott, D.J., Newman, I.A., Ruggles, C.L.N., and Walker, D.R.F. (editors), *Metadata in the Geosciences*, Group D Publications, Loughborough, pp.123-138.

Lanter, D.P., 1991, Design of a lineage-based meta-data base for GIS: *Cartography and Geographic Information Systems*, Vol. 18, pp. 255-261.

Lanter, D.P., and Essinger, R., 1991, *User-Centered Graphical User Interface Design for GIS*. Technical Paper 91-6, NCGIA, Santa Barbara.

Lanter, D.P., and Veregin, H., 1992, A research paradigm for propagating error in layer-based GIS: *Photogrammetric Engineering and Remote Sensing*, Vol. 58, pp. 825-833

Lanter, D.P., 1993, A lineage meta-database approach toward spatial analytic database optimization: *Cartography and Geographic Information Systems*, Vol. 20, pp. 112-121

Medjyckyj-Scott, D.J., Newman, I.A., Ruggles, C.L.N., and Walker, D.R.F. (editors), 1991, *Metadata in the Geosciences*. Group D Publications, Loughborough.

Newman, I.A., Walker, D.R.F., Mather, P., Ruggles, C.L.N., and

Medjyckyj-Scott, D.J., 1992. GENIE - The UK Global Environmental Network for Information Exchange, GENIE Research Report 1, Department of Computer Studies, Loughborough University.

Robson, P., and Adlam, K., 1991, A menu-aided retrieval system (MARS) for use with a relational database management system. In Medjyckyj-Scott, D.J., Newman, I.A., Ruggles, C.L.N., and Walker, D.R.F. (editors), *Metadata in the Geosciences*, Group D Publications, Loughborough, pp. 171-186.

Walker, D.R.F., Newman, I.A., Medjyckyj-Scott, D.J., and Ruggles, C.L.N., 1992, A system for identifying datasets for GIS users: *International Journal of Geographical Information Systems*, Vol. 6, pp. 511-527.

Winstanley, B., 1991, The approach of the ESRC Data Archive's BIRON system to the handling of spatial metadata. In Medjyckyj-Scott, D.J., Newman, I.A., Ruggles, C.L.N., and Walker, D.R.F. (editors), *Metadata in the Geosciences*, Group D Publications, Loughborough, pp. 115-122.

BIOGRAPHICAL SKETCH

Peter Fisher has degrees from Lancaster and Reading Universities and Kingston Polytechnic. He has taught at Kingston, at Kent State University, OH, and is now lecturer in GIS at Leicester University, where he directs the MSc in GIS and is an associated researcher with the Midlands Regional Research Laboratory. His recent research has been in the reporting, modelling, propagating and visualizing error and accuracy in spatial data. He is soon to be editor of the International Journal of Geographical Information Systems.