Achieving and Maintaining Interoperability of Spatial Data

Carsten RÖNSDORF, United Kingdom

Key words: Associativity, positional accuracy, segment snapping, DNF

SUMMARY

In Great Britain extensive work has been done to develop a solid framework for joined-up geographic information, the Digital National Framework (DNF), which acts as an enabler for a Spatial Data Infrastructure (SDI) and, as such, may be transferable to other countries [Murray, Munday and Bush, 2005].

This paper examines the associativity of geographic data to their reference as a result of work being done to implement positional accuracy improvement throughout Great Britain, with the goal to bring traditionally surveyed topographic maps to meter-accuracy against GPS. A particular methodology to analyze and store relationships between features, the Associativity Model, is presented along with first test results.

In the context of DNF these methods can be utilized to migrate data into the data model suggested by DNF, to manage the synchronicity of datasets over time and verify the use of datasets in conjunction with each other. The latter may be used to ensure data interoperability in the light of web services and future scenarios of serving geodata from multiple servers, maintained by multiple organizations into one application.

Achieving and Maintaining Interoperability of Spatial Data

Carsten RÖNSDORF, United Kingdom

1. INTRODUCTION

Current initiatives to create SDIs throughout the world will enable digital geodata to be integrated and shared. Geographic information differs from other information types in that it never stands on its own and always refers to a geodetic framework, and in many cases a reference data such as topographic or cadastral data. This needs to be taken into account when different datasets are integrated or data is exchanged to enable more advanced applications.

The position of a feature in the real world can be described in two ways: the absolute position within an overall reference system and relative to other known local features. While the coordinate value of a street lamp, for example, enables you to find it using GPS positioning, the information that it stands outside of 55 Acacia Avenue will be much more useful for a person trying to find this address and therefore the street lamp.

1.1 Interoperability

While interoperability is usually defined on the system level as *the ability of two or more systems or components to exchange information and to use the information that has been exchanged* [IEEE, 1990], interoperability of spatial data focuses on the second part of the definition, the *use* of the data and addresses content, accuracy and quality rather than exchange formats and data models.

The cartographic generalization of a smaller-scale dataset, for example, causes that a 1:100 000 topographic map, may not be suitable to be used in conjunction with a point dataset created with 1-meter accuracy to identify buildings in the real world.

One aspect of spatial interoperability relates to the accurate geodetic framework that supports the geodata. While the current geodetic reference systems in most countries have been defined long before absolute positioning through GPS became available in the 1980s, new surveying and data capture techniques, such as differential GPS measurements or orthorectified photography, are widely used today.

Efforts to achieve geometric interoperability between traditional data and GPS-derived data are described in section 2, Positional Accuracy Improvement.

1.2 Changing references

In order to be stable, reference data should change as little as possible over time. While the most widely used geodetic reference frameworks have been stable for fifty years or more, the current move by a larger number of National Mapping Organizations to an ETRS89 (European Terrestrial Reference System 1989)-based system shows that the reference, and therefore the coordinates, of a lot of data will change. The transformation between old and new reference systems can be separated and modelled as long- and short-wave effects [Imrek, 2004]. The former can usually be modelled in a transformation such as OSTN02™ [Greaves and Cruddace, 2001] while an example for the implementation of the latter is Ordnance Survey's Positional accuracy improvement programme (see section 2).

The second element of reference data, a geographical map representation (such as topographical or cadastral vector map) that allows users to create and express relationships of data to real-world objects is subject to much more frequent changes. With the creation of new streets and buildings, the destruction of others and changes to land ownership structures, this reference map, which serves as a computerized model of the real world, should reflect these frequent changes in order to be current and accurately describe the real world.

1.3 Geographic Information Systems and data layers

Most Geographic Information Systems (GIS) deal exclusively with features represented as arrays of coordinate values. A point is stored as a set of x and y coordinates, a line as an array of x and y coordinates and a polygon as a collection of one or more (if it contains holes) polygons with the definition that the first point is identical with the last one.

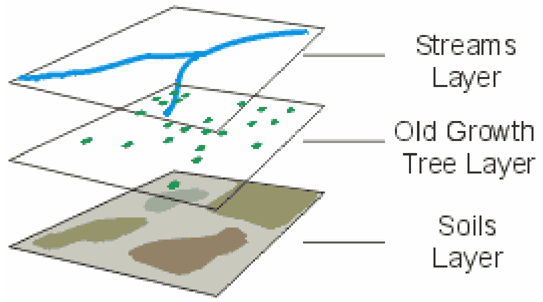

Figure 1: Layer concept
[City of Birmingham, 2004]

Separate datasets are stored in separate, independent data layers, as illustrated in figure 1, but no relationships are stored. Examples for a layer-based storage are the ArcView® Shape and MapInfo® TAB formats. If reference data is available to a suitable accuracy, it is a common process to digitize user data layers against the reference data. At the point of creation this data has a very distinctive relationship to the reference data, particularly if the new geometries are snapped to reference data. In this case existing points in the reference data and their coordinate values are copied and used in the new, digitized user dataset.

Over time, the reference dataset may change due to changes in the landscape such as new buildings, positional accuracy improvement or the introduction of a new coordinate reference framework. As a result, synchronicity between user and reference data needs to be maintained over time to guarantee data interoperability. If the two datasets are not synchronized, the data may be misinterpreted, which could lead to wrong decisions. The only way to do this in the layer concept is to flag up all reference data changes and check if the relationship to the user data has changed. This can partly be automated, but is still labour-intensive. Since no relationships are stored, the data model itself does not contain an integrated mechanism to guarantee the integrity between reference and user data.

1.4 Digital National Framework (DNF)

The DNF is a model for the integration of geographic information of all kinds – from national reference datasets to application information at the local level. Its implementation provides a permanent, maintained and definitive geographic base to which information with a geospatial content can be referenced [Ordnance Survey, 2004]. It is concerned with the relationship of geographic information and principles such as reusing existing data to create and maintain new datasets are fundamental to the DNF. In Great Britain DNF principles were successfully implemented as an index for land ownership parcels and a new database to of land now open to public access [Murray, Munday and Bush, 2005].

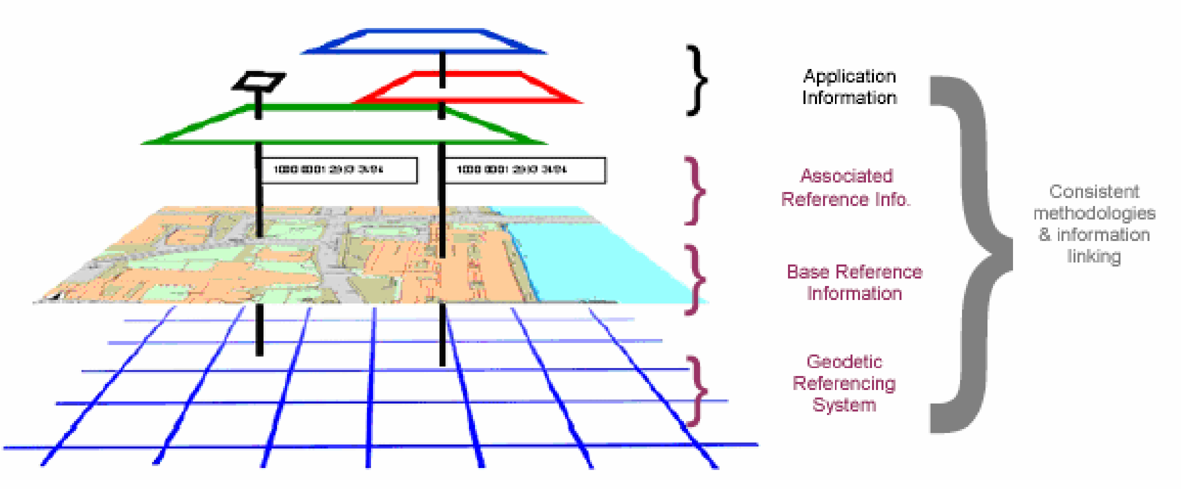

Figure 2: Digital National Framework [Ordnance Survey, 2004]

The reference information is provided in the form of a reference map or base map, a topographical dataset, for example, while application information, or user data, is linked into the reference information. Both are based on a common geodetic reference (figure 2).

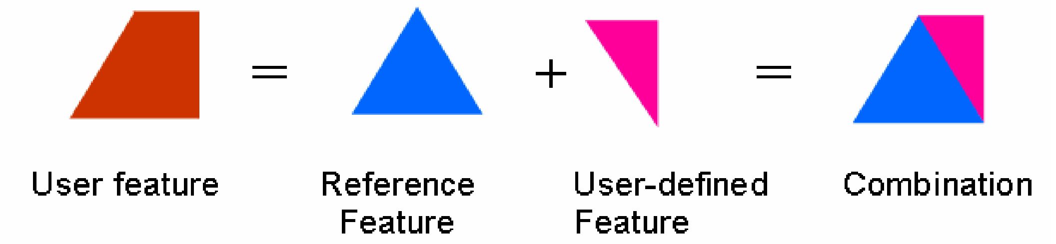

Figure 3: Feature collection using reference and user-defined features

Within the DNF, relationships between features are allowed to be built on three different levels. While level one does not involve geometry, level 2 (integrated geometry) establishes references between user and reference data in the way that the user data is assembled by building blocks from the reference data (reference features) as a feature collection.

In the case that the reference data does not offer the required geometries, the model can be extended by introducing additional user-defined features as shown in figure 3. This implies that the synchronicity between reference features and user-defined features needs to be managed.

In contrast to the layer model, relationships between reference and reference data are explicitly stored within the DNF data model.

2. POSITIONAL ACCURACY IMPROVEMENT

The improvement of reference data initially generated before GPS became available is currently being discussed by a number of organizations, such as the U.S. Bureau of Census [US census, 2005] and Ordnance Survey [Ordnance Survey, 2005]. Principally, this could be achieved either by a complete resurvey or by improving the existing data. The differences between a traditional and the improved reference dataset in a built-up area near Stratford-upon-Avon in Great Britain are illustrated in figure 4. It was found that the improvements and therefore shifts in identical points were random and could not be mathematically modelled.

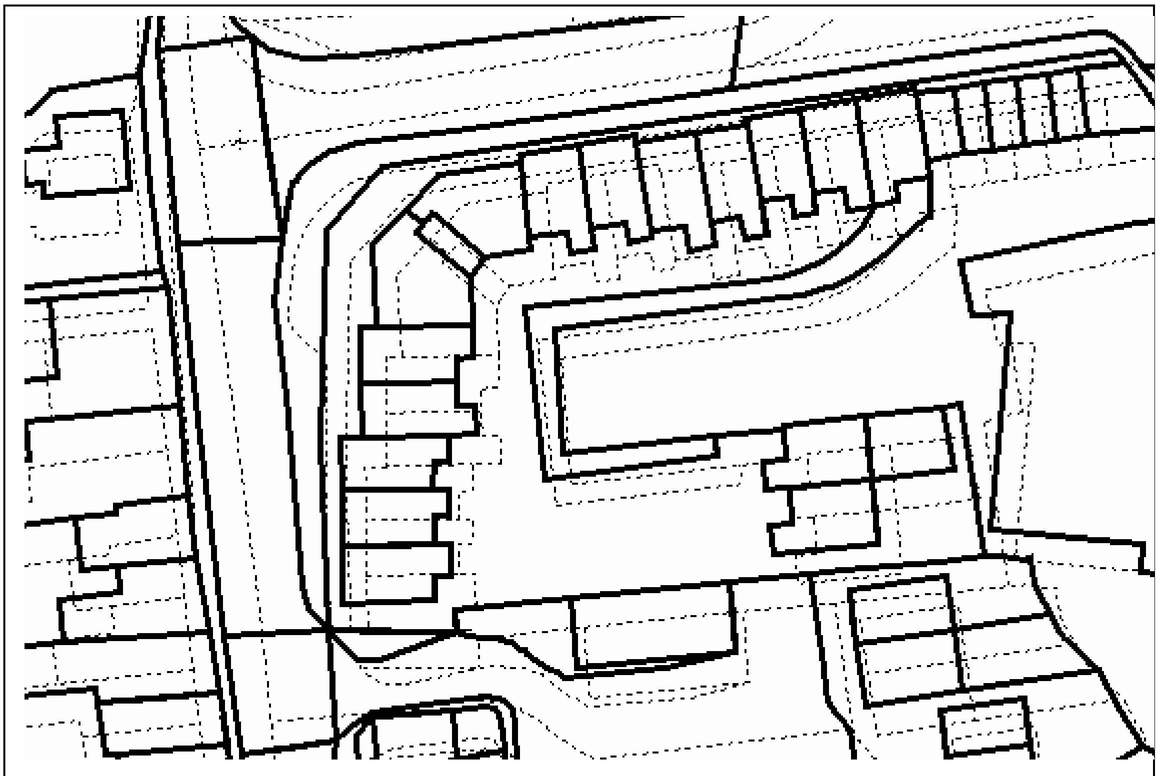

Figure 4: Traditionally-derived base data (bold line), new, improved base data (thin dashed line) Crown © Copyright

If new or updated reference data is provided by a National Mapping Organization or any other data provider, the new mapping may have a knock-on effect on user data that was captured against the previous reference data. Data users may need to correct this data to bring it back in sympathy with the improved reference map to make the two datasets interoperable. This scenario highlights the importance of the relationship between reference and user data, and has triggered the analysis and developments described in the following sections.

3. RELATIONSHIPS BETWEEN DATASETS

In many cases the relative relations between geographical features can be considered very important or even more important than the absolute coordinate values (position) of a feature in the reference system. The position of a street lamp, for example, could be determined as the distance and direction from the nearest house. In this example the easiest way for most people wanting to locate this feature in the real word would be to use the address of the house and distance to the street lamp. Within a GIS, however, the position of the street lamp is usually stored as a coordinate – the house is independently stored as a sequence of coordinates and no relationship is recorded. In this case the relative position of the street lamp against the house can be calculated from the known coordinates. This means that in today's GIS the relative relationship between two features (or two points) can be calculated but is not explicitly stored.

3.1 Polygon data

In conjunction with user data that had to be managed for positional accuracy improvement, the relationship between user data features and the reference data was analyzed for polygon data and three dominant cases of relationships were found. The base data was originally captured at a scale of 1:2500 with an absolute positional accuracy against GPS of 2.8 m RMSE, which will be improved to 0.4 m–1.1 m RMSE.

Figure 5 on the following page illustrates three important relationships between reference data (thick grey lines) and user data (thin dashed lines) that were found. In figure 5.1 the user data completely follows the geometry of the user data – the vertices of the user data relate to the vertices of the reference data. Within a GIS a lot of this data is created by digitizing using a snapping algorithm. Hence this data can be described as snapped user data. In figure 5.2 the reference data does not contain all the geometry to all the user data to reference to. Part of the bounding linework follows the reference data, another part doesn't. This case, characterized as partly snapped user data, requires additional vertices in the user data geometry. Figure 5.3 depicts a common scenario where the user data is digitized against the reference data without using a snapping algorithm. In this case the user data vertices are not identical with the reference data vertices, but close to them. Since the quality of digitization could be better (as in figure 5.1), this relationship can be described as roughly digitized user data.

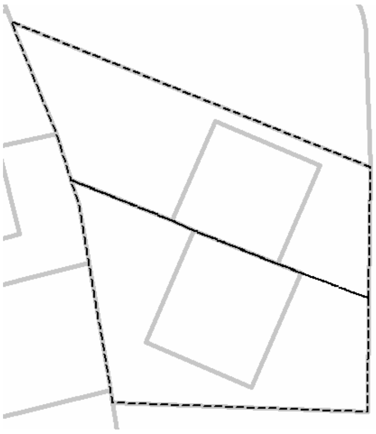

Figure 5.1:
Snapped user data

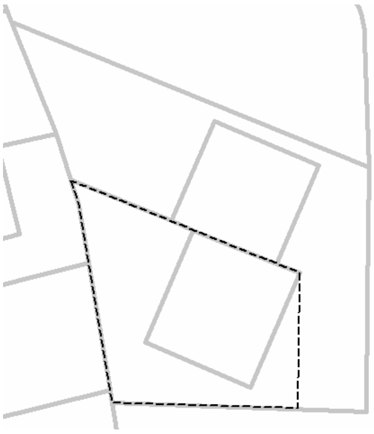

Figure 5.2:
Partly snapped user data

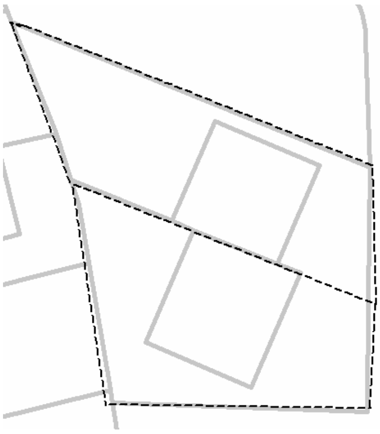

Figure 5.3:
Roughly digitized user data

4. THE ASSOCIATIVITY MODEL

The associativity model has initially been designed to analyse the relationship between user and reference data and bring user data back in sympathy with reference data after positional accuracy improvement of the latter. Since this is just one special case for the maintenance of spatial data integrity, it has been extended to cover the general aspects of maintaining spatial integrity and delivering interoperable data.

4.1 Distance function

A spatial relationship between two polygons can be described by a distance function [Straub and Wiedemann, 2000]. The distance function reports the minimum distance separating two polygons as function f(n) of the perimeter n of polygon A. Where two polygons are coincident, the function will report a zero minimum distance. $f(x) > 0$ indicates that the perimeter of polygon A falls within polygon B and therefore overlaps while $f(x) < 0$ shows that the perimeter of polygon A falls outside polygon B.

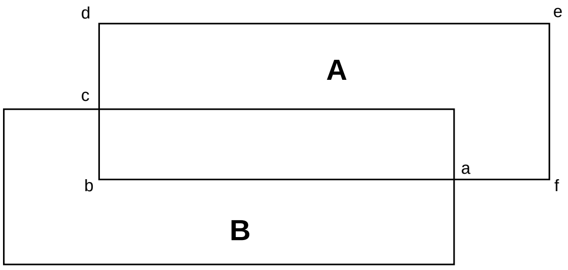

Figure 6.1: Simple polygons

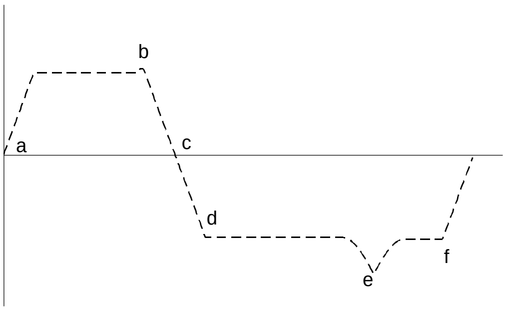

Figure 6.2: Distance function expressing the relationship between the polygons in 6.1

For the two polygons A and B, as shown in figure 6.1, the distance function between the boundary of polygon A and polygon B is displayed in figure 6.2. It is created by following the perimeter of polygon A from intersection point a over b, c, d, e and f back to point a. For all points on the perimeter (or a number of points that are placed in small discrete intervals on the perimeter) the shortest distance to polygon B is calculated. The distances can be expressed as a function of the perimeter as shown below.
Applying these principles to geographic information, distance functions can be calculated and stored between a polygon in a user dataset and the union of all polygons in a chosen reference dataset it overlaps. An example for a more complex user polygon, indicated by the solid line, and a topographic reference map (dashed line) can be found in figure 7.1, along with the resulting distance function in figure 7.2 [Stephenson and Rönsdorf, 2004].

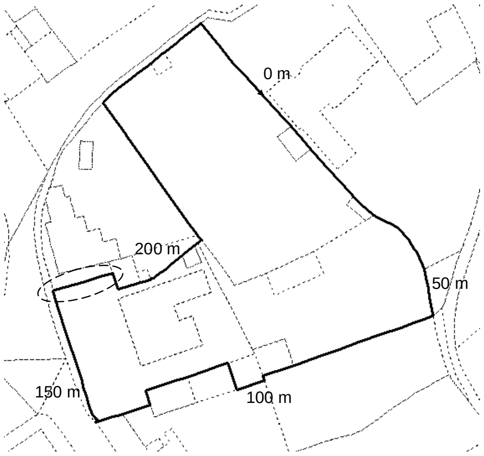

Figure 7.1: Complex polygon

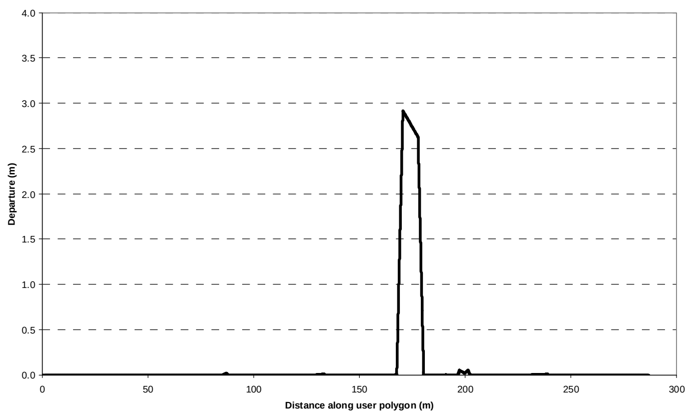

Figure 7.2: Related distance function

The distance function describes the relationship between polygons A and B, but in this form is insufficient to recreate polygon B from polygon A and the distance function. If $\vec{x}(n)$ is the vector describing the shortest distance between A and B and $\rho(\vec{x}(n))$ the orientation of this vector, a unique representation of the distance can be expressed by the 2-dimensional function storing the shortest distance and the orientation along the perimeter of polygon A, the oriented distance function. In this case polygon B can be recreated by plotting the distance oriented vectors along the perimeter of polygon A, which can be interpreted as "adding" the distance function to polygon A.

The distance function for snapped user data (see figure 6.1) constantly equals zero, while the distance function for partly snapped polygons (see figure 6.2) will be zero for the snapped part of the polygon and have a distinctive peak for the perimeter that is not coincident with the user data. For the scenario shown in figure 6.3, the roughly digitized data, the distance function will be close to zero, with some noise indicating the difference between the user data and underlying reference data.

The distance function in figure 7.2 shows a small peak at around 200 m of the perimeter. This can be interpreted as a digitizing error. The distance function can be utilized in a segment snapping algorithm in which distance function values below a low threshold are brought to zero. For the polygon geometry this means that between two points on the perimeter the fitting piece of line geometry of the related polygon is inserted.

While the model will work with line and polygon user data, it requires polygonized reference data. Further research is needed to determine the best way to store distance functions in files and databases.

4.2 Results

The test data shown in figure 8.1 shows a property polygon that was roughly digitized against the reference data. The distance function between the polygon and the underlying base data shows only noise close to zero. Using the segment-snapping method outlined in section 4.1, the geometry of the polygon can be altered to snap to the underlying reference polygons. A reference can be established to the reference polygons indicated as A and B in figure 8.2. The process ensures that the node structure of the improved user data coincides with the node structure of the reference data. In the case that the distance function rises beyond a threshold that relates to the accuracy of the data, the segment can be flagged to an operator to decide if the user polygon should coincide with the reference data. This decision may be supported by an automated rule set.

Figure 8.1
Roughly-digitized polygon (dark, thin line) node structure

Figure 8.2
Improved user data geometry (dark, thin line) with with improved node structure

(Reference data are displayed in the grey, thick line in both figures)

The same process can be used for partially snapped data (as illustrated in figure 5.2). In this case the distance function will report values significantly greater or less than zero. These can either be used to create a user-defined reference polygon or the distance function can be stored to define the relationship in conjunction with a reference to the overlapping reference polygons.

After a while the reference data may change, as shown by the circles in figure 9.1. In this case a new distance function can be calculated between the user polygon and the changed reference data and compared with the original one, and the resulting associativity changes can be flagged up for individual segments as shown in figure 9.2 (thick dashed line) or automatically dealt with by using a suitable rule set.

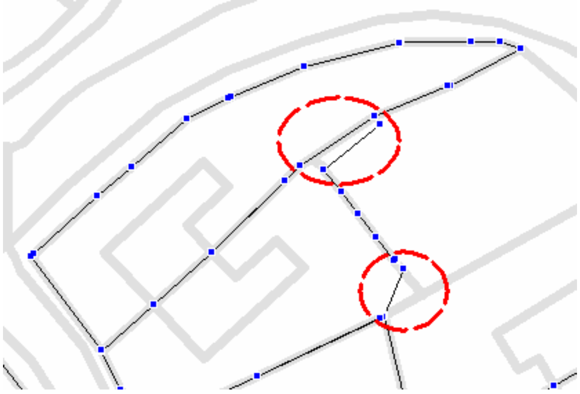

Figure 9.1: Changed reference data

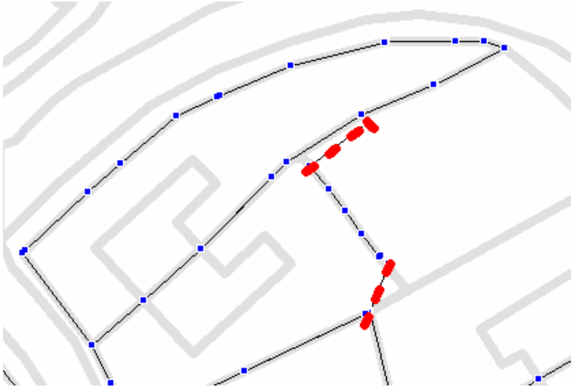

Figure 9.2: Associativity changes

5. USE OF ASSOCIATIVITY WITHIN THE DNF

5.1 Integration

The Associativity Model can be used to establish the relationship between reference and user data. In the case of roughly digitized data, it is capable of improving the user data by segment snapping as shown in section 4.2 to allow referencing to reference feature. The method caters for fully and partially snapped data to create user-defined reference features according to figure 3.

The methodology can be used to migrate existing user data into a DNF-compliant data model.

5.2 Managing change

Presumably, a more powerful application lies in maintaining the integrity between reference and user data. If the relationships are stored, they can be verified after reference data (or any other dataset a particular user dataset relates to) are updated within the system. In case of a conflict these can be flagged (as described in section 4.2). This method will be particularly efficient if combined with an incremental change-only update mechanism of the reference data as it is implemented for topographic data in Great Britain.

5.3 Validation of datasets

The more advanced level 3 of the DNF assumes that reference data will be supplied on demand, for example, using an Open Geospatial Consortium (OGC) Web Feature Service (WFS), while the users are only holding their specific user-references. User features will be built on the fly by combining the WFS-served base reference with locally held user-references.

Today most organizations maintain reference data on their own network and exercise full control over it. If the reference data is maintained by a different organization and just streamed into the application, the end-user does not control the relationship between user and reference data anymore. The user either trusts the reference data provider or puts a process into place that allows him to verify the use of the user data against the particular vintage of reference data that is accessed remotely. In this case the Associativity Model will allow the user to store the associativity to the externally maintained reference data and automatically validate it once the data is used within an application.

6. CONCLUSION

In the GIS world, which is characterized by employing predominantly layer-based data models, relationships between datasets, particularly between user and reference data, are often overlooked. The DNF as a framework concept aims to make interoperable spatial dataset by relating user to reference datasets a reality. Within this framework the Associativity Model may be a suitable method to manage relationship information, migrate datasets into a DNF-compliant data model and validate datasets within an open WFS-environment.

REFERENCES

City of Birmingham, 2004. GIS website, Birmingham, UK [http://gisweb.informationbirmingham.com/imf/sites/birmingham_publicdev/custom/Compass/default.html]

Commission of the European Communities, 2004. Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE), Brussels {SEC(2004) 980}

GOS, 2005. Geospatial One Stop service [http://www.geodata.gov/gos]

Greaves, M. and Cruddace, P., 2001. The OS's new Coordinate Transformation for Great Britain, Geomatics World, Volume 10 Issue 1 November/December 2001, Stevenage [http://www.gps.gov.uk/additionalInfo/Geomatics%20world.pdf]

IEEE, 1990. Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, 1990

Imrek, E., 2004. UTM in the Austrian Cadastre - Present steps to a higher absolute accuracy, EuroSDR workshop proceedings, Improving the Positional Accuracy of Spatial Data, Dublin [http://www.dit.ie/DIT/news/events/conferences/gi-databases/eurosdr_pai_imrek.pdf]

Murray, K., Munday, B. and Bush, I., 2005. Enabling information integrity within SDI's – the Digital National Framework concept, FIG working week 2005 and GSDI-8 conference proceedings, Cairo

Ordnance Survey, 2004. The Digital National Framework – evolving a framework for interoperability across all kinds of information, white paper, Southampton [http://www.ordnancesurvey.co.uk/oswebsite/aboutus/reports/dnf_white_paper.pdf]

Ordnance Survey, 2005. Positional accuracy improvement website [www.ordnancesurvey.co.uk/positional]
US census, 2005. MAF/Tiger Accuracy improvement project [http://www.census.gov/geo/mod/maftiger.html]
Stephenson, K. and Rönsdorf, C., 2004. PAI Associativity Model, Ordnance Survey, Southampton
Sutcliffe, M., 2003. Making the move, creating pos-PAI candidate geometries. Aylesbury Vale District Council
Straub, B. M. and Wiedemann, C., 2000. Towards the update of geodata by automatic object extraction, Image and Signal Processing for Remote Sensing VI, SPIE Proceedings Volume 4170. Serpico, S.B, ed, Bellingham, WA: The International Society for Optical Engineering, 2000, p. 304 [http://ipi216.ipi.uni-hannover.de/html/publikationen/2000/straub/straub_gerke_wiedemann.pdf]

BIOGRAPHICAL NOTES

Involvement in GIS and spatial data since 1995 in various different market sectors and countries. Lead Consultant for Positional Accuracy Improvement of large-scale geodata for Ordnance Survey of Great Britain since 2002. Leading role in organizing the FIG commission 3-sponsored EuroSDR workshop *Impacts of Improving the Positional Accuracy of GI databases* in Dublin, 2004 and the upcoming workshop *Achieving Geometric Interoperability of Spatial Data*, in Munich, 2005. Also engaged in EuroSDR distance learning courses teaching *positional accuracy improvement*.

Previous work included planning of worldwide mobile telephone networks with T-Mobile® International, developing a Land Information System for the Indonesian National Land Agency in West Sumatra, Indonesia, as part of a development project managed by the German Technical Cooperation Agency GTZ and project managing a corporate GIS implemented in an electricity, gas and water multi-utility company in Germany.

Academic qualification: Dipl.-Ing. in Geodesy, University of Bonn, Germany, 1998

Member of the Geospatial Information and Technology Association of the United States

Publications:

- 2004, The pole position – high accuracy geodata and geometric interoperability in Great Britain, Europe and the US, Geomatics World, Volume 12 Issue 6, September-October 2004, Stevenage [http://www.pvpubs.com/read_article.asp?ID=32&article_id=192]
- 2004, Positional Integration of Geodata, Intergeo East conference proceedings, Belgrade
- 2004, Improve your business position, GITA conference 27 proceedings, Seattle
- 2005, Make the case to deliver real value, GITA conference 28 proceedings, Denver

CONTACTS

Carsten Rönsdorf
Ordnance Survey
Romsey Road
Southampton, SO16 4GU
UNITED KINGDOM
Tel: +44 (23) 8030-5733
Fax: +44 (23) 8079-2039
Email: Carsten.Roensdorf@ordnancesurvey.co.uk
Website: www.ordnancesurvey.co.uk