Identifying Land Cover Change Using a Semantic Statistical Approach: First Results

A.J. Comber[1], P.F. Fisher[1] and R.A. Wadsworth[2]

Department of Geography, University of Leicester,
Leicester, LE1 7RH, UK.
Telephone: +44 (0) 116 252 3859
Email: ajc36@le.ac.uk

Abstract

The identification of land cover change from the 1990 LCMGB and its successor, the UK LCM2000 is problematic because of their different methodologies. A semantic statistical approach for combining different land cover ontologies was applied to the problem. Expert expressions of idealised relations between the dataset semantics were combined with information on expected spectral overlap and parcel spectral heterogeneity information to generate two characterisations for each parcel. Comparing the characterisations identified some change parcels and highlighted some systematic attribute artefacts. This approach allows a linkage ontology to be developed and provides a greater understanding of how the ontologies of LCM1990 and LCM2000 relate to each other.

1. Introduction

Repetitions of land resources inventories commonly entail re-evaluations of objectives and methodologies due to technical innovation and policy developments (Comber *et al.*, 2002). In extreme cases the re-evaluation results in a re-conceptualisation of the classification scheme for a particular phenomenon: a changed ontology. This is a generic problem, exemplified by the land cover mapping of Great Britain in 1990 (LCM1990) and in 2000 (LCM2000), descriptions of which can be found in Fuller *et al.* (1994) and Fuller *et al.* (2002) respectively. LCM1990 records 25 Target land cover classes defined on pixel reflectance values, identified using a *per-pixel* maximum likelihood classification algorithm. In contrast LCM2000 records land cover information on a *per-parcel* basis at three levels of detail, including 26 Broad Habitat classes. Parcel reflectance values were classified, assisted by ancillary data (soils) and Knowledge Based Corrections. Extensive meta-data is attached to each parcel or vector as described by Smith and Fuller (2002). The net result is a drastic change in the *meaning* behind the class labels between LCM1990 and LCM2000. Changes in ontology make it difficult to relate the information contained in the previous dataset to that of the current. To overcome this problem, Comber *et al*. (submitted) have proposed a Semantic Statistical Approach for reconciling the ontological difference between datasets.

This paper describes the results of using the Semantic Statistical Approach to identify locales of land cover change from LCM1990 and LCM2000. The approach combines expert opinion of how the semantics of the two datasets relate with meta-data descriptions of LCM2000 parcel spectral heterogeneity. This approach explores semantic and ontological links between datasets, exposes weakness in expert understanding of the data concepts and provides a methodology for integrating different data types (e.g. Raster and Vector, Vector and Vector, Intersects between vector objects).

The generic problem of spatial data integration has been addressed by various workers under a series of different headings. (Ahlqvist *et al.*, 2000; Bishr, 1998; Devogele *et al.*, 1998; Frank, 2001; Kavouras and Kokla, 2002; Visser *et al.*, 2000). A common theme to emerge from this work is the identification of semantics as the bottleneck to translation between different data ontologies. The problem of semantic translation between data concepts can be tackled using formal ontologies. An ontology in this context is an explicit specification of a conceptualisation (Guarino, 1995). The integration "problem" is that an object described in one conceptualisation may not correspond or translate to another conceptualisation. In the "ontologies approach", the capacity for data integration and sharing depends on understanding the way that the data is conceptualised. For instance Pundt and Bishr (2002) describe how members of different information communities get access to the meaning of other data if they can approach the ontologies that have been developed by those who collected the data.

In this paper we report the results of field visits to evaluate the Semantic Statistical Approach, which identified parcels suspected of having changed in the interval 1990 to 2000. In this way the utility of the LCM2000 parcel meta-data was explored and data artefacts revealed.

2. Semantic Statistical Approach for detecting change

The basic concept behind the approach is to use generate two characterisations of each LCM2000 parcel: one based on the interpreting the intersecting LCM1990 pixels through a Semantic Look Up Table (LUT); the other based on a comparison of the parcel spectral heterogeneity information with the parcel class through a Spectral LUT. It was hypothesised that parcels whose characterisations differed most between the two characterisations, were candidate locales of land cover change. Note, that due to methodological similarities the parcel spectral heterogeneity information is equivalent to LCM1990 classes.

An expert familiar with both datasets was asked to construct a table mapping the expected and unexpected relations between LCM1990 and LCM2000 classes. This Semantic LUT was constructed under a scenario of idealised semantic relations. From published information about the expected spectral overlap of the spectral subclass variants in the spectral heterogeneity attribute ("PerPixList") with the 26 broad habitat classes, a Spectral LUT was constructed. In both LUTs a three valued logic was used:

"-1" to indicate relationships that were unexpected;
"0" to represent relationships that were uncertain;
"+1" to indicate relationships were expected.

From the LUTS it was possible to generate a score of Unexpectedness and of Expectedness at both points in time. The 1990 score was by summing the number of pixels for each of the classes indicated as being *Unexpected* in the Semantic LUT, and those as being *Expected*. The 2000 score was derived by summing the percentages given in the PerPixList attribute that were *Unexpected* and *Expected* according to the Spectral LUT.

Comparing the two characterisations after normalising one to the other (2000 was based on percentage of the parcel area, 1990 on the number of pixels) generated a vector in a feature space of Unexpectedness (*X*) and Expectedness (*Y*). Parcels whose descriptions varied the most, i.e. had the biggest vectors, formed the set of hypothesised land cover changes. The 100 parcels with the greatest vectors were identified for three LCM2000 broad habitat classes, representing three cases of ontological change:

- Broadleaved Woodland (minor ontological change);
- Arable Cereal (moderate ontological change);
- Acid Grassland (major ontological change).

Because it was possible to calculate the *gross* vector for each parcel based on the *number of pixels* or the *percentage* vector based on the *percentage of pixels*, two sets of 100 parcels were identified for each broad habitat class. Of the 600 parcels, 100 were randomly selected and of those 93 were visited. A discussion of the issues involved selecting parcels by the size of vectors between *proportions* or by vectors based on *size* (areas) can be found in Comber *et al*. (submitted). The number of parcels visited and identified in the different directions for each LCM2000 broad habitat are shown in Table 1.

Table 1. The number of parcels of each broad habitat land cover type visited against the number identified using the semantic statistical method.

	LCM2000			
Direction	Broadleaved Woodland	Acid Grassland	Suburban and Rural Development	Total
NE		3 / 20		3 / 20
SE	2 / 7	19 / 157	11 / 56	32 / 220
SW	5 / 7	5 / 13	3 / 3	13 / 23
NW	21 / 186	2 / 10	22 / 141	45 / 337
Total	28	29	36	93 / 600

3. Results

The Semantic Statistical Method has at its core the idea that change can be identified by comparing two characterisations of the LCM2000 parcel, one from the intersecting LCM1990 pixels, and the other from the spectral subclass ("PerPixList") attribution. Characterisation was by comparing parcel attributes with expected attribution through a LUT to calculate scores of Expectedness and Unexpectedness for 1990 and for 2000. Parcels that were hypothesised to have changed were those whose characterisations were most different. Necessarily this process generated a vector between 1990 and 2000 and it was possible to analyse the results with reference to the direction and magnitude of the 1990 to 2000 vector. Placing Unexpectedness on the *x* axis and Expectedness on the *y*, movements were grouped into four basic directions corresponding to quadrants of a compass: NE, SE, SW and NW.

3.1. Analysis of Change Detection results

In this section the success of change detection is evaluated, by broad habitat, by vector direction, and by magnitude. Some examples of actual changes are also included by way of illustration.

3.1.1. Broad Habitat

A simple binary analysis of whether a field survey confirmed (or denied) that the parcels suspected of change had actually been subject to some change on the ground is shown in

Table 2.

Table 2. The number of parcels where change was found during the field visits, by broad habitat.

LCM2000 class	**Change observed**		**Total**	**Proportion Correct**
	No	**Yes**		
Broadleaved Woodland	23	5	28	0.22
Acid Grassland	26	3	29	0.12
Suburban and Rural Development	25	11	36	0.44
Total	74	19	93	0.26

From Table 2, 26% of the hypothesised changes were found to be actual change on the ground and predictions about the "Suburban and Rural Development" were the most successful.

3.1.2. Vector Direction

Table 3 presents results by vector direction. It is apparent that most of the successfully identified changes had a vector direction NW, that is moving from high unexpectedness, low expectedness to low unexpectedness, high expectedness). These are shown in Figure 1.

Table 3. The number of parcels where change was found during the field visits, by vector direction.

Direction	**Change observed**		**Total**	**Proportion Correct**
	No	**Yes**		
NE	3	-	3	0.00
SE	28	4	32	0.14
SW	11	2	13	0.18
NW	32	13	45	0.41
Total	74	19	93	0.26

3.1.3. Comparing change and "no-change" vectors

We can compare the magnitude and direction of the vector of those parcels found to have changed with the remainder of the 100 parcels, to see whether they are systematically different. These are shown in Figure 2 for Suburban / Rural Development. This class had the most successfully identified changes and parcels identified on the *number* of pixels as opposed to the *percentage* of pixels are shown for clarity.

3.1.4. Example changes

By way of illustration Figure 3 describes six example parcels, with vectors that moved NW and that were found to have changed.

3.1.5 Summary

From the field work we can see that changes that were identified on the ground show a heterogeneity of distance, direction and starting position in the feature space of Expectedness / Unexpectedness. Generally calculating vectors based on pixel counts (area) works better than by percentage, but very small and very large segments may require opposite treatments.

Figure 1. The distance and direction of the parcel vectors for which actual change was found on the ground. Vectors for Suburban / Rural Development are in black lines, Broadleaved Woodland are dashed and Acid Grassland in bold/

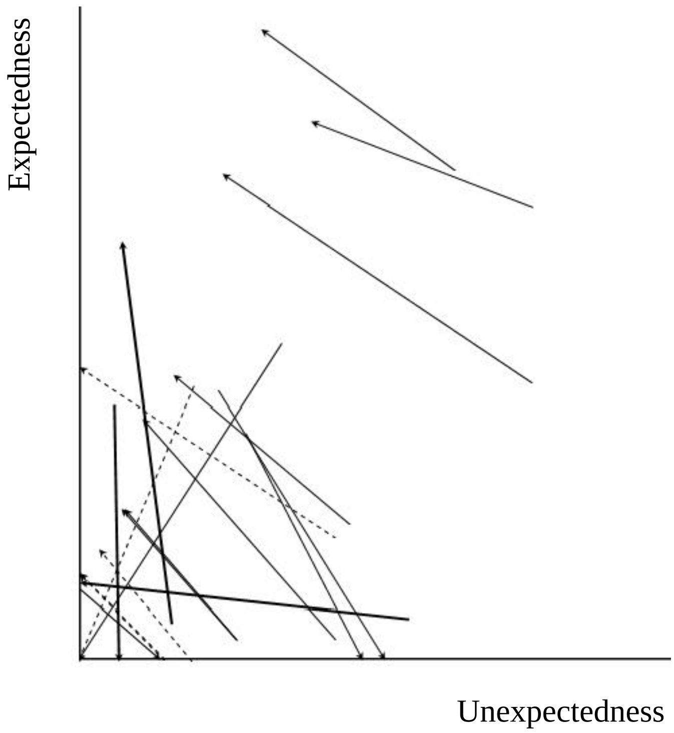


Figure 2. Vectors of 100 parcels of broad habitat class Suburban / Rural Development. Those that were identified as actual changes (1990 to 2000) are in bold.

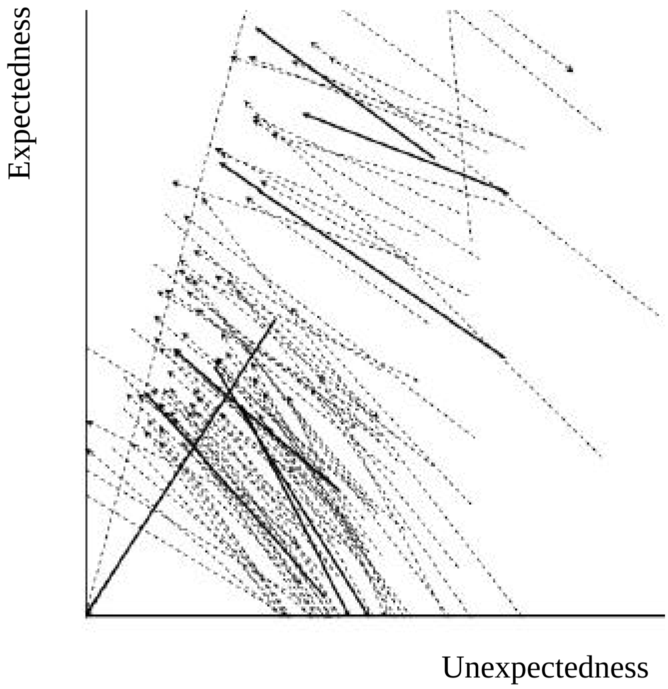
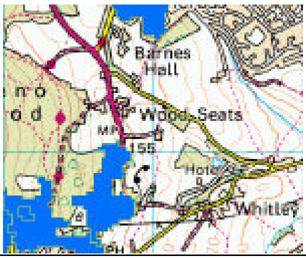
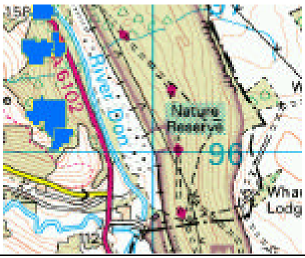
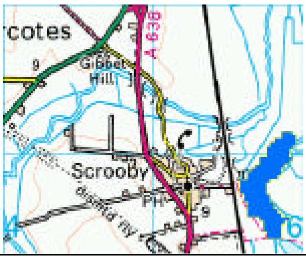
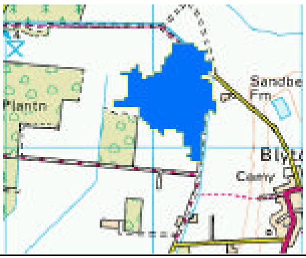
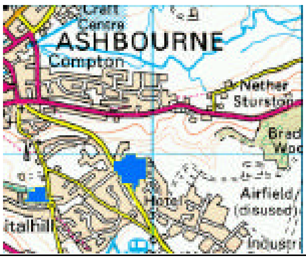
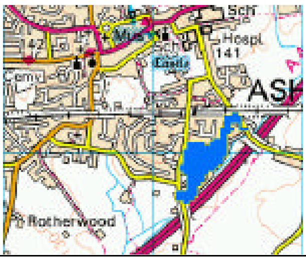


Figure 3. Six example change parcels, a brief description of the nature of the change, the original (1990) class and some context from OS 1:50000 Raster scanned maps ().

LCM2000 class, 1990 to 2000 change process, *1990 classes*	**Geo-context**	**Picture**
Broadleaved Woodland, Maturation, Open classes e.g. *grass*, juvenile trees		
Broadleaved Woodland, Maturation, Open classes e.g. *grass*, juvenile trees		
Acid Grassland, change in land quality by overstocking, degraded from *improved grassland*		
Acid Grassland, changes in land quality, let go, degraded from *Arable*		
Suburban / Rural Development, New Housing Development		
Suburban / Rural Development, New Housing Development		

3.2 Analysis of False Positive results

The second set of results is a breakdown of the reasons for the "false positive" results: a parcel being identified as a potential change area, whilst not actually having changed. It describes the origins of the large vector between the 1990 and 2000 characterisations.

Table 4. Origins of large vector size for parcels found not to have changed (false positives).

	LCM1990	LCM2000	Empty PPL	Incorrect PPL	Expert & PPL	Total
Broadleaved Woodland	5	12	4	2	-	23
Acid Grassland	-	2	-	15	9	26
Suburban / Rural Development	6	9	1	9	-	35
Total	11	23	5	26	9	74

LCM1990: the LCM1990 pixels were not consistent with the actual land cover
LCM2000: the LCM2000 parcel classification was not consistent with the actual land cover
Empty PPL: empty spectral heterogeneity attribute field
Incorrect PPL: spectral heterogeneity attributes not related to parcel class
Expert & PPL: expert error where the expert failed to fully understand the change ontology *and* either an empty or incorrect spectral heterogeneity attribute

From Table 4 it is possible to comment on the reasons why the false popstives (i.e. the parcels with large vectors were considered not to have changed) were identified as change parcels using this methodology:

- 46% of the parcels that were found not to have changed either had a large number of LCM1990 pixels that were inconsistent with “no change” or the LCM2000 parcels were incorrectly classified;
- 5 (7%) parcels had empty PerPixList fields;
- 35% of the parcels had a PerPixList attribution that was inconsistent with the parcel classification;
- 12% of the parcels were erroneously identified because of expert misunderstanding of the changed ontologies between 1990 and 2000 combined with inconsistent or empty PerPixList fields.

The impact of all of these artefacts was to create a large vector between 1990 and 2000 which resulted in the polygon being identified as a potential change area. From Table 4 it is evident that false positives for Acid Grassland, the class with the greatest change in ontology between 1990 and 2000, were all related to PerPixList (spectral attribution) anomalies. Artefacts for the other two classes are evenly spread.

3.3. Explanation of Negative Results Using Vector Directions

The false positive results were then analysed by vector direction as shown in Table 5, to identify any further rules or filters.

Analysis of the false positive results by direction presents a much clearer picture than analysis by land cover class (Table 3). In the sections below, the anticipated explanations for movement in each direction between 1990 and 2000 are described in the subsections below, followed by analysis of the field validation.

Table 5. The vector direction origins of the false positive results with true positive results in bold

Direction	LCM1990	LCM2000	Empty PPL	Incorrect PPL	Expert & PPL	Total
NE	-	-	-	-	3	3 (**0**)
SE	-	2	-	23	3	28 (**4**)
SW	-	-	5	3	3	11 (**2**)
NW	11	21	-	-	-	32 (**13**)
Total	11	23	5	26	9	74

3.3.1 Movement NE: increased Expectedness and increased Unexpectedness

Parcels with a NE vector were anticipated to contain conflicting spectral attribute information. This may have been due to a changed ontologies or parcels with high spectral heterogeneity. All of the parcels identified in this category were Acid Grassland. Movement was due to a failure by the Expert to understand the links with the LCM1990 class of Moorland Grass and due to major ontological change. The effect of these was a low starting point (in 1990) and conflicting attribute information in 2000.

3.3.2 Movement SE: decreased Expectedness and increased Unexpectedness

Parcels with a SE vector were expected to contain spectral attribution inconsistent to the parcel class. Despite some change being found (4/32 parcels) we believe this was by chance rather than design. All of the parcels with vectors moving in this direction had anomalous spectral attribution. Most of this was due to the use of single date imagery rather than composite two date imagery, resulting in an attribute of “100% Water, variant c” for 20 out of the 32 parcels. Of the other 12:

7 were indications of genuine heterogeneity (1 Acid Grassland parcel and 6 Suburban with “Urban” variants classes with no *believed* relation to Suburban);

3 were parcels classified incorrectly and therefore their attribution did not match the class (2 Acid Grassland, 1 Suburban);

2 were due to the changed ontology, defining Acid Grassland in terms soil acidity.

Movement in this direction therefore indicates spectral attribute anomaly.

3.3.3 Movement SW: decreased Expectedness and decreased Unexpectedness

Movement in this direction was expected to indicate increased uncertainty about the parcel because the LCM2000 attributes would be contributing less Expectedness and Unexpectedness than the LCM1990 pixels. The parcels that were identified in this category can be placed into two groups: those with empty spectral attribution fields and those containing Spectral subclass not described in the information that accompanied the release of the LCM2000 data. Vectors moving in this direction indicated systematic artefacts in the parcel attributes.

3.3.4 Movement NW: increased Expectedness and decreased Unexpectedness

Parcels with vectors moving in this direction were thought have increased certainty about parcel classification, relative to the way the way the parcel area was classified in 1990 and therefore indicative of land cover change. Parcels with vectors moving in this direction indicated either land cover change or misclassification in 1990 or 2000.

3.3.5 Summary

In summary the four directions of movement are attributable to the following:

NE – (spectral) attribution was inconsistent with the parcel land cover class and there was a failure by the Expert to understand the relationships between the 1990 and 2000 ontologies;

SE – attribute inconsistency (heterogeneity);

SW –spectral heterogeneity attribute is empty **or** there are subclasses not described in the Spectral LUT;

NW – either there has been actual change on the ground **or** an error in one dataset.

4. Discussion

Some inconsistencies were observed between the parcel land cover target class and the broad habitat class in 1990 and 2000 respectively. In 1990 Broadleaved woodland is confused with Grass and Suburban with Arable. In 2000, Broadleaved woodland is confused with Arable and Suburban with Arable. These findings are consistent with other descriptions of common spectral confusions (Geoff Smith, *pers. comm.*, 2003).

The major finding of this work is the extent to which systematic analytical and data artefacts can be identified and used to generate filters for further analyses. In this work we have not filtered the analysis in any way – all parcels were included if the vector was in the subset of the 100 largest vectors for their class, regardless of the origins of that vector. The analysis of results show that vector direction is a useful indicator of different artefactual trends in the data: specific anomalies or artefacts are indicated by specific vector directions. From these we can construct a set of filters or rules to eliminate parcels with specific characteristics from analysis. The following artefacts with associated actions have been identified for omission from future analyses:

- Parcels with empty PPL fields;
- Parcels with PPL attributes not described in the meta-data accompanying the release of the data;
- Parcels identified from Single Scene Imagery;
- Parcels where the "knowledge based" correction is inconsistent with the PPL attributes.

A further important finding was the extent to which the field visits revealed inconsistencies in Expert understanding of how the LCM2000 class of Acid Grassland related to various 1990 classes. The results show that their LUT was not consistent with observed phenomena in the field for a specific class. The Expert used in this study did not fully the impact of the changed ontology of Acid Grassland. In such circumstances it may be preferable to use another Expert. We took the results back to the Expert and revisited their understanding of the data concepts and they confirmed that they were more familiar with lowland vegetation and ecosystems.

Future Work will be directed in a number of areas. Firstly, the use of different expressions of expert opinion. Here we have used a single expert under the scenario of their *idealised mapping of how the semantics relate*. Other experts and mapping of concepts are available. For instance a Change Mapping LUT of the expected transitions between land cover classes and a Technical Mapping LUT of how different land cover class concepts relate based on heuristic knowledge of where spectral confusion may occur. Evidence from these might support or be contrary to the existing statements of expectedness and unexpectedness, and in turn may allow stronger or weaker inferences to be made about change for specific parcels. Secondly, these multiple statements of Expectedness and Unexpectedness from different experts, under different scenarios would be suitable for combination using uncertainty formalisms such as Dempster-Shafer or Rough Sets. Thirdly we hope to develop a “Cook Book” for users of LCM2000.

From these results and further analyses of other field data we shall develop rules to filter data and attribute artefacts and to identify some of the analytical limits of the data and meta-data. There is an issue about the size of the parcel and the use of parcel vectors generated using percentage or area. It may be that parcels bigger than a certain area have to be eliminated from analyses, and that above a particular threshold (to be determined) vectors based on percentage change need to be applied to assess large, noisy parcels. Similarly for small parcels, there may be a minimum size below which we can say nothing about the probability of change

5. Conclusions

This work has shown that change analysis is possible using the Semantic Statistical Approach developed by Comber *et al.* (submitted). Change areas are identified by characterising the LCM2000 parcel in terms of its spectral attribution and the intersection of LCM1990 data, and then comparing the two characterisation. Typically change was indicated by a vector of direction NW in a feature space of Unexpectedness (X) and Expectedness (Y) between 1990 and 2000.

In the process of determining the extent to which change is reliably identified, certain data artefacts have also been revealed:

- We identify parcels that were anomalous in terms of their attribution and developed rules or filters with which to eliminate data with specific characteristics from analyses according to the direction of their vectors.
- Failings in the Expert understanding of data concepts were revealed, which may lead to the use of different experts or revision of Expert understandings.

At the time of writing there is little guidance about how the extensive LCM2000 meta-data may be used and analysed. Examining the results of a field validation confirms that the Semantic Statistical Approach provides a greater understanding of how the ontologies of LCM1990 and LCM2000 relate to each other than is provided by data documentation, specification and class descriptions alone. This gives greater insight into the *meaning* of the LCM2000 data structures with reference to their attribution. It is hoped that the work reported here contributes to a developing area involving the analysis and use of metadata to evaluate and revise the base information.

6. Acknowledgements

This paper would not have been possible without the assistance those involved in creating LCM2000, especially Geoff Smith. This paper describes work done within the REVIGIS project funded by the European Commission, Project Number IST-1999-14189. We wish to thank our partners in the project, especially Andrew Frank, Robert Jeansoulin, Alfred Stein, Nic Wilson, Mike Worboys and Barry Wyatt.

7. References

AHLQVIST, O., KEUKELAAR, J. AND OUKBIR, K., 2000. Rough classification and accuracy assessment. *International Journal of Geographical Information Science*, **14**, 475-496.

BISHR, Y., 1998. Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science*, **12**, 299-314.

COMBER, A.J., FISHER, P.F. AND WADSWORTH, R.A., 2002. Creating Spatial Information: Commissioning the UK Land Cover Map 2000. In *Advances in Spatial Data*, edited by Dianne Richardson and Peter van Oosterom (Berlin: Springer-Verlag), pp. 351-362.

COMBER, A.J., FISHER, P.F. AND WADSWORTH, R.A., 2003. Actor Network Theory: a suitable framework to understand how land cover mapping projects develop? *Land Use Policy*, in press.

COMBER, AJ, FISHER, PF, and WADSWORTH, RA, submitted. A semantic statistical approach to negotiating heterogeneous ontologies: how to relate similar features mapped in different ways. *Submitted to IJGIS*.

DEVOGELE, T., PARENT, C. AND SPACCAPIETRA, S., 1998. On spatial database integration *International Journal of Geographical Information Science*, **12**, 335-352.

FRANK, A.U., 2001. Tiers of ontology and consistency constraints in geographical information systems, *International Journal of Geographical Information Science*, **15**, 667-678

FULLER, R.M., GROOM, G.B. AND JONES, A.R., 1994. The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, **60**, 553-562.

FULLER, R.M., SMITH, G.M., SANDERSON, J.M., HILL, R.A. AND THOMSON, A.G., 2002. Land Cover Map 2000: construction of a parcel-based vector map from satellite images. *Cartographic Journal*, **39**, 15-25.

GUARINO, N., 1995. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, **43**, 625-640.

KAVOURAS, M. AND KOKLA, M., 2002. A method for the formalization and integration of geographical categorizations. *International Journal of Geographical Information Science*, **16**, 439-453.

PUNDT, H. AND BISHR, Y., 2002. Domain ontologies for data sharing-an example from environmental monitoring using field GIS. *Computers and Geosciences*, **28**, 95-102.

SMITH, G.M. AND FULLER, R.M, 2002. Land Cover Map 2000 and meta-data at the land parcel level. In *Uncertainty in Remote Sensing and GIS*, edited by G.M. Foody and P.M. Atkinson (London: John Wiley and Sons), pp 143-153.

VISSER, U., STUCKENSCHMIDT, H., SCHUSTER, G. AND VOGELE, T., (2002). Ontologies for geographic information processing. *Computers and Geosciences*, **28**, 103-117.