Title: You know what land cover is but does anyone else?...an investigation into semantic and ontological confusion

Authors: Alexis Comber^{1*}, Peter Fisher¹, Richard Wadsworth²

Addresses:

¹ Department of Geography, University of Leicester, Leicester LE1 7RH, UK

² Centre for Ecology and Hydrology, Monks Wood, Abbots Ripton, Huntingdon, Cambridgeshire, PE28 2LS, UK

* Contact author: Department of Geography, University of Leicester, Leicester LE1 7RH, UK, Tel: +44 (0)116 252 3859, Fax: +44 (0)116 252 3854, email: ajc36@le.ac.uk

Abstract

Information derived from remotely sensed data is increasingly being used to describe land cover and landscape structure for a range of applications. Users in different disciplines may use this information, although they may have different perceptions of land cover than those who created the information. They may be unaware of the origin of the information or its meaning and they may treat it as if it were data. Current paradigms for reporting meta-data and data quality do not adequately communicate the producer's knowledge, and should be extended to describe the conceptual, semantic and ontological meanings of land cover.

1. Introduction

Products from remote sensing, particularly in the form of so called Land Use and Land Cover mappings (Land Characterisation is another term), are becoming widely used in many areas of activity which involve one form or another of landscape analysis. The remote sensing research community are becoming the providers of information to a wider society. In this letter we argue that the remote sensing community is failing to communicate all it knows about the information it is producing to those users.

We believe that the remote sensing community has concentrated too much on the technical issues (how do we produce something?) and not enough on the semantic and ontological issues (what are we trying to produce?). This represents an over-emphasis on what is uncontroversial and easily measured, and is being compounded by meta-data standards that do nothing to help the people who have to use the data; but only show that the data producer can follow a particular recipe. Every remote sensing text book describes the process of land cover mapping in a similar way; raw sensor data are collected and corrected (geometrically, radiometrically, topographically etc.), then areas with similar spectral properties are related to features on the ground (using a more or less sophisticated supervised or unsupervised algorithm). Different processing paths, algorithms, and sensor characteristics will produce different maps, and the cause and extent of these differences fills whole academic journals. But, we believe the critical issues for the wider community of users of remotely sensed products are: who decides and defines what the features of interest are; and how is information about the features

communicated to the user? We argue that there is a need for more useful meta-data descriptions due to:

- an increased number of users of land cover information (witness the number of countries with national land cover mapping programmes based on satellite imagery but delivering classified mappings);
- easier dissemination of digital information (especially on the internet and in the future over the GRID);
- widely available software for processing this information (Geographical Information Systems);
- increased influence of policy as opposed to science in defining what should be mapped (Comber et al., 2003).

2. Different representations

It is possible for different land cover maps to be derived from the same data. Variation can be accounted for under two broad headings: technical and ontological.

2.1 Technical Aspects

The stages at which technical aspects enter into the process of constructing a land cover map are well known and grounded in statistical analysis of physical measurements, although measurements recorded by the sensor are not directly or exclusively controlled by the features on the ground. Therefore, pixel values are an indirect and imperfect measure of the reflectance of those features. Some of these issues can be compensated for by geometric and radiometric corrections. The corrected data are then classified using one of a range of approaches, (hard-soft, supervised-unsupervised), but all assume that clusters of spectral homogeneity (however they are identified) relate to homogenous cover on the ground. To account for some of this variation a number of quality reporting paradigms have become established. The principal paradigms are the confusion matrix, user and producer accuracies and the kappa statistic; these describe how the map relates to an alternative, but hopefully compatible, source of information. Although these dominant paradigms are aspatial at least they show an explicit acknowledgement of the variation of land cover due to technical reasons.

2.2 Ontological aspects

An ontology specifies how an idea is conceptualised. In this case it describes what land cover actually means in a wider sense that includes the epistemology of data collection, pre-processing and processing, and the ontological aspects of determining what features are to be included in each class. Whilst some of this meaning is informed by the technical aspects (accuracy, scale, minimum mapping unit), most originates from the meaning behind the different land cover class labels, commonly described in terms of botany, biogeography and phyto-sociological association. Yet land cover classification schemes are not determined by the reflectance properties of land cover and their inferred relationship with biology alone. Rather their specification combines policy objectives at regional, national or international levels with the individual and institutional objectives of those charged with creating the derived land cover map to inform policy (Comber et al. 2003) (see figure 1). The result may be vague specifications of land cover features resulting in uncertain conceptualisations, the precise nature of which are not described in

the land cover meta-data. An example is the different grassland classes in the UK LCM2000 (Fuller et al. 2002), where meanings of the grassland classes overlap, but the overlap is only inferred by the class description.

An example is provided by Robbins (2001) who demonstrated the different conceptualisations of forested landscapes held by local farmers and state foresters in Rajasthan. Robbins showed land cover class definition to be an inherently political exercise: it determines what is recorded as existing and thereby influences future management decisions (Robbins, 2001).

[Insert figure 1 about here]

In this context the semantics provide the link between the statistical delineation of land cover and ecology-based descriptors of class definitions. Variation in land cover maps caused by the political processes that determine classification schemes may have just as profound an influence on the land cover map as do the technical aspects. But these aspects of land cover conceptualisation are neither reported nor described by current remote sensing paradigms in meta-data reporting.

3. Why land cover variation is an issue now

The remote sensing community has long been aware of technical and ontological variation (although perhaps not under those labels); compare the variation in the European CORINE classification (EEA, 2001), the USGS Land Use and Land Cover Classification (Anderson et al., 1976) and Land Cover classifications used in the different land cover maps of Great Britain or anywhere else (Fuller et al., 2002; Comber et al. 2003). Variation due to technical issues is well discussed within the remote sensing literature and hitherto there has been little need to describe explicitly the sources of variation due to ontological issues. The result is that estimates of land cover stocks and spatial configuration are often commented upon without reference to ecological process, the ontological meaning of the land cover features or the epistemology of data processing. Griffiths et al. (2000) note that ‘none of these data limitations or methodological problems should overshadow the more profound difficulty of providing a valid ecological interpretation for the results’ (p2702). However, land cover is being used as a surrogate to describe the landscape structure and character by an increasing number of users who may be unaware or ignorant of land cover information origins and its meaning.

We are concerned about the impact of user ignorance. Land cover is perceived differently by different disciplines and their perceptions inform their assessment of the data and their analyses. Ecologists usually define land cover in terms of the presence and abundance of specific plant species (e.g. Barr et al. 1993). Soil surveys may interpret land cover as an indication of the underlying soil type (e.g. Macaulay Institute for Soil Research 1984). In these cases the meaning of the land cover information is related to ecological and pedological processes. Landscape Ecology is concerned with patterns. Analyses and interpretation are therefore partly dependent on scale, spatial resolution, interpretation

and classification (e.g. Wickham and Riitters 1995). However, landscape ecologists may not be concerned with the way that land cover is conceptualised and what its meaning is. Gulinck et al. (2000) note the divergence between the vocabularies and methods in landscape studies and remote sensing: 'For landscape researchers and planners, the remote sensing origins of information may be irrelevant' (p. 2553), but we would contend that it is never irrelevant because it controls what can be discerned.

All of the above disciplines produce information but most rely on external sources for land cover and may not understand the precise meaning of the information in the same way as the producers. Whilst users are commonly concerned (like the producers) with the technical artefacts (scale and granularity, spectral ambiguity, boundary issues, etc), the veracity of the knowledge embedded in the land cover map is taken on trust. Users from different scientific disciplines are not provoked into considering what the land cover information they are presented with means or the assumptions underpinning it. The implications of not understanding the land cover ontology are two-fold. Firstly, users impose their own specific disciplinary or institutional views of the meaning of land cover. Secondly, because users are encouraged to ignore the different ontological aspects that contribute towards the meaning of the land cover information, they are implicitly encouraged to treat it as a neutral, objective fact – as data to be used in their analysis.

4. Conclusions

Researchers in remote sensing have always been aware of land cover variation due to the technical issues and ontological aspects in creating information from raw remotely sensed data. The resulting maps contain many features as a result of the methods of its construction embedded within it. This is not described by current meta-data which is frequently concerned with the easily measurable aspects of the scale, spatial resolution, format, and (aspatial) accuracy. The remote sensing community accepts Technical variation but ignores Ontological or conceptual variation.

It is increasingly easy for the non-expert to obtain digital Land Cover information (over the World Wide Web or by direct purchase from a producer organisation) and to incorporate it into a GIS. Unfortunately they often assume that land cover information corresponds to their disciplinary conceptualisations of land cover and treat it as data, just as the producer of the land cover data may have treated the original satellite imagery made up as it is of Digital Numbers as data.

Meta-data is the subject of extensive standards (e.g. FGDC 1998) which do not specify or intend that the conceptualisation or meaning of classes be reported, only definitions or descriptions of classes. Whilst users may not understand the meaning of the land cover they incorporate into their analyses, producers do. Historically producers of most natural resource maps embedded all this meta-data in paper reports which accompanied the map information, and Fisher (2003) has argued that multimedia presentation methods are an appropriate digital method to link the spatial information with the report and to describe the full ontology of the information. We would like to recommend that meta-data standards are extended to include a full description of the ontology, at the least to indicate

who decided what the features of interest were and why, and the producers view of the integrity and discernability of those units (possibly a pairwise comparison of all mapping classes). As part of this process it might be instructive for producers to get out into the field more (to understand the ontological) and for users stay in the lab more (to understand the technical).

We would like to pose two questions to the remote sensing community about land cover meta-data and quality reporting:

- Do they adequately describe the conceptual and semantic meaning of the data ontology?
- Do they address fundamental problems of remote sensing as articulated by Verstraete et al. (1996, pXX) which is ‘to establish to what extent radiation measurements made in space can in fact provide useful information for ... applications’?

Acknowledgements

Ideas behind this letter were presented at RSPSoc 2003, Nottingham and describe work done within the REVIGIS project funded by the European Commission, Project Number IST-1999-14189.

References

- ANDERSON, J.R., HARDY, E.E., ROACH, J.T. and WITMER, R.E., 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. U.S. Geological Survey, Professional Paper 964, Reston, VA.
- BARR, C.J., BUNCE, R.G.H., CLARKE, R.T., FULLER, R.M., FURZE, M.T., GILLESPIE, M.K., GROOM, G.B., HALLAM, C.J., HORNUNG, M., HOWARD, D.C., and NESS, M.J., 1993. Countryside Survey 1990: main report. Countryside 1990 Series: Volume 2. Department of the Environment, London.
- COMBER, A. FISHER, P. AND WADSWORTH, R., 2003. Actor Network Theory: a suitable framework to understand how land cover mapping projects develop? *Land Use Policy*, 20: 299–309.
- FISHER, P.F. 2003. Multimedia Reporting of the Results of Natural Resource Surveys. *Transactions in GIS*, 7: 309-324.
- EEA, 2001, CORINE Land Cover - Technical guide, European Environment Agency http://etc.satellus.se/the_data/Technical_Guide/ [available 27/08/01].
- FGDC, 1998. Content Standard for Digital Geospatial Meta-data. Meta-data Ad Hoc Working Group, Federal Geographic Data Committee, FGDC-STD-001-1998, www.fgdc.gov/standards/documents/standards/meta-data/v2_0698.pdf [available 24/10/0]
- 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.
- GRIFFITHS, G.H., LEE, J. and EVERS HAM, B.C., 2000. Landscape pattern and species richness; regional scale analysis from remote sensing, *International Journal of Remote Sensing*, 21(13-14): 2685–2704.
- GULINCK, H., DUFOURMONT, H., COPPIN, P., and HERMY, M., 2000. Landscape research, landscape policy and Earth observation. *International Journal of Remote Sensing*, 21(13-14): 2541–2554.
- MACAULAY INSTITUTE FOR SOIL RESEARCH, 1984. *Organization and methods of the 1:250 000 soil survey of Scotland*. The Macaulay Institute for Soil Research, Aberdeen.
- WICKHAM, J.D. and RIITTERS, K.H., 1995. Sensitivity of landscape metrics to pixel size. *International Journal of Remote Sensing*, 16 (18): 3585-3594.
- VERSTRAETE, M.M., PINTY, B. and MYNENI, R.B., 1996. Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing of Environment*, 58(2): 201-214.

Figure 1. The contributing influences to land cover classification development