

Conference or Workshop Item (Refereed)

Comber, A. J.; Fisher, P. F.; Harvey, F.; Gahegan, M.; Wadsworth, R. A.. 2006 Using metadata to link uncertainty and data quality assessments. In: *Progress in spatial data handling: Proceedings of the 12th international symposium on spatial data handling, University of Vienna, Austria, 12-14th July 2006*. Berlin, Springer, 279-292.

This version available at <http://nora.nerc.ac.uk/3157>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the authors and/or other rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

Contact CEH NORA team at
nora@ceh.ac.uk

Using Metadata to Link Uncertainty and Data Quality Assessments

Lex Comber, Pete Fisher, Francis Harvey, Mark Gahegan, Richard Wadsworth

1. Introduction

In this paper we argue that the links between data quality reporting, metadata and subsequent assessments of data uncertainty need to be stronger. This builds on the ideas developed by Fisher (2003) who commented that data quality and uncertainty were like ships that pass in the night. Current data quality reporting is inadequate because it does not provide complete descriptions of data uncertainty and allow assessments of data fitness (Comber et al., 2005a, Comber et al., 2005b). As Fisher (2003) noted “data quality as it has developed in the writing of data standards, and uncertainty as it has been researched in recent years, have followed two completely different tracks”. The ideas presented in this paper are an attempt to set a research agenda to provide some glue to join together three related but distinct areas of scientific endeavour in spatial and geographical information sciences: uncertainty analysis, data quality / fitness for use assessments and metadata reporting

For a variety of reasons data purporting to record the same real world features do so in a variety of different ways. Comber et al. (2002; 2003; 2005b) describe the impacts on the final data product of changing commissioning contexts, technical developments, advances in scientific understanding and on the final data products and the specific conceptualisation embodied in the data of the objects under investigation. These aspects strongly influence the *meaning* of the data in its widest sense. At present information on meaning is not provided in metadata and is not included in metadata standards, an absence that hinders data integration efforts as much as different source data and statistical processes.

A recent workshop¹ concluded that the case for current metadata standards (ISO, OGC) is not yet proven in relation to geographical information. This is in part because of the relative and constructed nature of much geographical information (Harvey 2000; Comber et al. 2003), but also because of changing context within which such information is used: applications of geographical information are now ubiquitous; there are many more users; there are many more downloadable sources available to the user; geographical information (and systems) are being used in many areas of scientific endeavour (Comber et al. 2005b).

In this dynamic and fluid context of more users with varying degrees of experience but equal access to data, the importance of appropriate metadata reporting increases. Metadata should provide information that helps user assess the *usefulness* of a dataset relative to their problem. That is, it should facilitate assessments of the relationship

¹ http://www.niecs.ac.uk/events/activating_metadata/index.html

between measures of data quality and uncertainty (for any specific application). Current metadata categories tend towards descriptions of *usability* and do not allow users to easily identify the suitability of data for their intended application. We propose that metadata should include information describing the conceptualisations embedded in the data and information relating to user-experience with the data would allow better assessments of usefulness to be made.

2. Uncertainty, data quality and metadata

2.1 Data quality

Measures of data quality have traditionally been generated by the producers of spatial data. Data quality is described in terms of the ‘big 5’: Positional Accuracy, Attribute Accuracy, Logical Consistency, Completeness, and Lineage. Various standards for the components of metadata reporting have defined based on these measures of data quality (FGDC, 1998; ISO 2003; OGC, 2005). The interested reader is directed to Guptill and Morrison (1995) which provides an expansion and clarification of the concepts of quality used in those transfer standards. These measures originate from the historical cartographic legacy of geospatial data production (Fisher, 1998) and the need to transfer information (Fisher, 1999). In Guptill and Morrison, none of the authors discuss what the measures really mean or how they may be used in fitness-for-use assessments. Only Salgé (1995) discussed issues outside of the ‘Big 5’ and introduced the concept of semantic accuracy, noting that objects recorded in the data may not actually match the definition of those objects which the database is based. This concept of quality has had little impact upon the standards process and data quality specifications (Fisher, 2003). This is because the specification of data quality standards, previously dominated by the national mapping agencies and software companies, has now become the preserve of dedicated standards organisations such as ISO and the Open GIS Consortium.

The specification of quality standards continue to reflect data production interests, reporting the easily measurable and showing that the data producer can follow a recipe (Comber et al., 2005a) rather than more fully communicating the producer’s knowledge of the data. A number of quality reporting paradigms have become established, principally 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 information². The net result of this legacy is that measures of data quality are difficult for users to interpret in relation to a specific application: users do not know how to apply data quality measures in their analyses (Hunter, 2001) in order to assess the suitability of the data for their application. This assessment is done by expert analysts.

2.2 Spatial data uncertainty

Geographic objects are those structures created to impose order on the real world: objects are delineated, identified and placed into categories according to a set of criteria. Whilst many (non-geographic) objects have boundaries that correspond to physical

² The ‘Accuracy’ measures are commonly correspondences, for example when land cover from a field survey is compared to that derived from remote sensed data.

discontinuities in the world, this is not the case for many geographic objects that may be less well defined.³ Uncertainty in spatial data relates to doubt about the information that is recorded about a location (Fisher, 1999). There are different types and directions of uncertainty relating to nature of the object under consideration. Fisher et al. (2005) have proposed taxonomy after Klir and Yuan (1995) where different types of uncertainty are related to how well the geographic objects are defined (Figure 1).

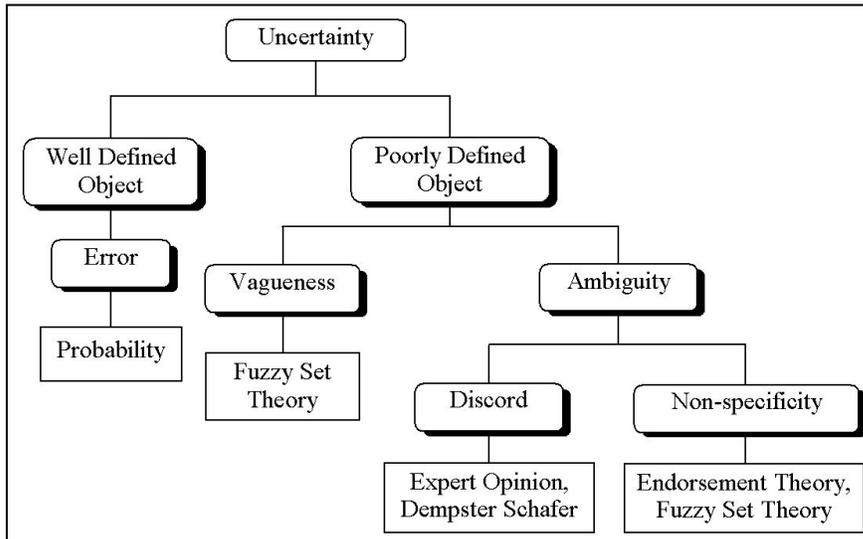


Figure 1. A conceptual model of uncertainty in spatial data (from Fisher et al., 2005)

‘Well defined objects’ are those where the object classes (e.g. “building”) are easily separable from other classes and where the individual instances are clearly distinct from other instances of the same class. For well defined objects the main uncertainties are positional and attribute errors which can be analysed using probabilities.

‘Poorly defined objects’ may be vague or ambiguous. Vagueness occurs where it is difficult to identify the spatial extent or to precisely define an object. That is, it is difficult to allocate unequivocally individual objects into a class. Ambiguity is composed of discord and non-specificity. Discord arises when one object is clearly defined but could be placed into two or more different classes under differing schemes or interpretation of the evidence. Non-specificity occurs when the assignment of an object to a class is open to interpretation.

The analysis of uncertainty for well defined geographic objects is advanced. There are many examples in the literature where these aspects of data quality have been modelled using descriptions of error and accuracy, held by the metadata and combined with probabilistic approaches (e.g. Monte Carlo simulations, Bootstrapping). Most discussions of uncertainty in geographical information relate to these types of approaches to

³ Barry Smith and David Mark have written extensively on this subject exploring the concepts of *fiat* and *bone fide* boundaries, corresponding to *fiat* and *bone fide* geographic objects (Smith, 1995; 2001; Smith and Mark, 2001).

modelling error. This is not the case for poorly defined geographic objects, characterised by vagueness, ambiguity, discord and non-specificity, and yet we would argue that these are the most important and most frequent cases when working with geographical information.

2.3 Metadata

We have defined metadata as ‘information that helps the user assess the *usefulness* of a dataset relative to their problem’. That is, in the context of uncertainty assessments, it should allow the user to determine data quality and fitness for their analysis or data integration activity. Necessarily this involves relating one view of the world, as encapsulated by a particular dataset, to another or to the objective of the analysis (Ahlqvist, 2004; Wilson, 2004). Assessing data fitness or suitability involves understanding data limitations (such as mismatches between user and data concepts) and quantifying the direction and magnitude of the uncertainties associated with integrating activities. In many cases environmental information supply is either a monopoly or an oligopoly and users do not have a choice about which dataset to use. Current metadata reporting does not give any information to the user about how to best exploit the data.

Whilst current metadata standards (e.g. ISO, 2003) are adequate to guide assessment of technical constraints on data integration caused by Structure (raster to vector) or Scale (generalizations to lower level classes), they do not describe the concepts (ontologies) embedded in the data. Using the above definition they are inadequate: they do not facilitate user assessment of the applicability of the data to their problem; they convey nothing about the meaning of the data. Such information could include:

- the organizational (cultural) or epistemological context which gave rise to the data in the first place;
- the commissioning (often policy-related) context of the data.

Thus tools for user assessments that have been developed hitherto (e.g. Devillers et al., 2005) have to assume a perfect match between the concepts encapsulated in the data with those of the user application. They do not address one of the fundamental aspects of working with geographic information – it is relative, subjective and constructed.

3. Geographic Information – origins and use

The nature of geographic information and its increased use provide the context metadata to be expanded to include data concepts in order to make the link between uncertainty and data quality assessments.

The creation of geographic information is inherently relative. It involves the abstraction of the real world into some kind of data object and therefore results in information loss. There many choices to be made when constructing a geographic database– what to include, what not to include – for a number of data characteristics: granularity, scale, class numbers, class definitions etc. Each one of these contributes to the overall conceptualisation of the real world that gets encoded into the database and each choice implicitly specifies some information loss. How important that information loss is depends on the application in hand.

The end result is considerable variation in geographic information databases which requires far more than the simple class descriptions in order to understand. By way of example, it is instructive to review how the concept of “Bog” changed between 1990 and 2000 land cover mappings of the UK. The 1990 Land Cover Map of Great Britain (Fuller et al., 1994) defined Bog in terms of standing water, permanent waterlogging, waterlogging, surface water and the presence of characteristic plant species (*Myrica gale* and *Eriophorum spp.*)⁴. In the Land Cover Map 2000 (Fuller et al., 2002) Bog was defined on peat depth greater than 0.5m⁵. The consequences of this change in conceptualisation are significant: for a 100km x 100km area (Ordnance Survey tile SK) in the English midlands there were

- 12 pixels of ‘bog’ (<1 ha) in 1990
- 120728 pixels of ‘bog’ (~75 km²) in 2000.

It is worth noting that for both surveys the same class label was employed, the datasets were developed and constructed using similar remotely sensed data, by the same team from the same research institution.

The reasons for the change in construction of the class of ‘Bog’ are to be found in the different commissioning contexts of the 2 surveys (Comber et al., 2003). The 2000 dataset specifications were much influenced by a number of behind the scenes factors that were not present in 1990:

- as a response to national legislation as a result of the Rio Earth summit;
- the shift of responsibility for the environmental away from the government to environmental agencies;
- the dynamic interaction of those agencies, the government and interpretations of policy with the processes on the ground that could be discerned using remote sensed data.

This is all information that would help the user understand more fully the data they incorporate into their analyses, but which currently is not included in metadata reporting paradigms.

None of the issues described above that are involved in geographic data construction, its subjective nature and changes in ontology between surveys are recent phenomena. They have always existed: different information collected by different agents for different purposes record the same real world features in different ways related to different conceptualisations and ontologies (although not under those labels). What has changed is the context within which this variation in data exists.

First, obtaining spatial data traditionally was a lengthy process. It meant entering into a dialogue with the data producer who would be concerned about whether their data would be used inappropriately, about their reputation, that of the data and the results of the original studies for which the data was commissioned. It was then an iterative exchange of information between user and producer. Potential users of the data would themselves be experienced and aware of the issues highlighted above: they would be spatial data-literate.

⁴ <http://science.ceh.ac.uk/data/lcm/classM.htm>

⁵ <http://science.ceh.ac.uk/data/lcm/lcmleaflet2000/leaflet3.pdf>

Second, the situation today is different. GI and GIS applications are ubiquitous in the public realm (e.g. GPS, in-car Sat-Nav, etc). The number of explicit users of GI and GIS has increased as reflected in the presence of GIS departments in local, regional and national government, health authorities, GIS is starting to be taught in high school and not just as a specialist post-graduate activity. Similarly, many new areas (such as insurance assessments) and academic disciplines (from archaeology to microbiology), now routinely use GIS where previously they did not. Such users may not be aware of variation in geographic information and may not understand specific uncertainties of the data they incorporate into their analyses. Users may not fully understand what the data represents – its meaning or semantics- and they will assume that it fits their conceptualisations because of familiar class names and labels that apparently match their prototypical categories with those names (Comber et al., 2005b).

Third, the number of applications and users are set to rise further with recent cyber infrastructure initiatives. The EU INSPIRE project seeks to make available “relevant, harmonised and quality geographic information to support formulation monitoring and evaluation of Community Policies”⁶. Similarly the development of the computing GRID is providing “pervasive, dependable, consistent and inexpensive access to advanced computational capabilities, databases, sensors and people”⁷. Broadly these cyber infrastructures and other eScience activities seek to connect users to spatial data without them having to go through the broker, the door keeper or the intermediary to the data and the process of dialogue that ensues.

Fourth, there has been a decline in the survey memoir as metadata - the book *about* the data describing the concepts and mapped features was always more interesting to researchers than the map it described – as noted by Fisher (2003) and the increased ease of access to digital data (ftp, web-portals, etc), these developments bring with them a risk of opening a Pandora’s box of issues that have previously never had to be explicitly addressed because they were known.

However, users have to be pragmatic and use the data that is available, despite the fact that the existing data was (usually) collected for a different purpose (and those purposes change over time). Metadata does not communicate the data producer’s model of the world embedded in data and, consequently, users are invited to treat information (an interpretation) as if it were an objective measurement. The consequences and uncertainties of unknown mismatches between the (remotely held) data objects and those of the user will be far more profound than those due to positional or attribute accuracy. In this context, assumptions may not generally be reported as a caveat to the “results” of a report or research project. This keeps the customer happy and allows the user to be seen as a “good” researcher.

⁶ <http://inspire.jrc.it/> INfrastructure for SPatial InfoRmation in Europe

⁷ <http://www.escience-grid.org.uk>

4. Linking uncertainty, data quality and GI

4.1 Preamble

The preceding sections have commented on the implicit relationship between Uncertainty and assessments of Data Quality, and have indicated that current metadata reporting are insufficient to explicitly link them. This is in part due to the constructed and subjective nature of much geographic information and in part due to the increased analytical dangers of data mis-use due to increased numbers of users and the ongoing development of spatial data or cyber infrastructures. The questions that are thrown up by this relationship are:

- How to provide appropriate information about the data such that users can either make informed decisions about which data is most suited to their analysis?
- How do enable users to understand the limitations of the results of any analysis using that data?
- How do we link uncertainty and data quality assessments to do this?

We believe that expanded metadata should provide information about the wider context of the data. It is instructive to review recent developments in thinking in the areas of uncertainty and data quality in spatial data.

4.2 Uncertainty

Early approaches to uncertainty in spatial data were concerned with ‘error’. The focus was on the measures of data quality derived from the correspondence table (also referred to as the confusion matrix, data validation, correspondence table (...CAN ANYONE THINK OF ANY MORE?)) and established a number of quality reporting paradigms: user and producer accuracies and the kappa statistic (e.g. Congalton; 1991). These describe how the data relates to an alternative source of information allowing *predicted* to be compared with *observed*. There are a number of issues with these that approaches: first, they are aspatial assuming error to be evenly distributed across the data; second, very different measurement processes are used to generate *predicted* and *observed*. An example of the impact of using different measurement approaches is provided by the Countryside Surveys (CSs) in the UK of 1990 (Barr et al., 1993) and 2000 (Haines-Young et al., 2000), where field survey data was used to validate satellite derived land cover. Unsurprisingly, the stock of any given land cover derived from these 2 surveys was found to vary (Fuller et al., 1998). Yet, this is the major paradigm for reporting on ‘accuracy’ in remote sensing analyses, providing terms for ‘error’ – called variously user & producer accuracy, Group 1 and 2 uncertainties, Type I and II errors – which are then used in probabilistic approaches for modelling uncertainty: Monte Carlo simulations, epsilon error bands, bootstrapping.

More recent work has seen a movement away from probabilistic assessments of error toward uncertainty approaches that place greater emphasis on assessments of conceptual and semantic data aspects. Approaches based on formal ontologies (Frank, 2001; Pundt and Bishr, 2002; Visser *et al.*, 2002) and concepts of interoperability (Bishr, 1998; Harvey *et al.*, 1999) have identified differences in semantic concepts were identified as *the* major barrier to data integration. This has had as its focus a concern with the “what it is we are measuring” and necessarily draws on the work describing the indeterministic nature of many geographic features or objects (Burrough and Frank, 1996; Smith, 1995;

2001; Smith and Mark, 2001; Gahegan and Brodaric, 2002). The concept of ‘fuzzy objects’ has been explored by different authors for example by Fisher et al. (2004) for mountains and by Chen (2002) using the concept of a ‘dune’. Dunes are well defined (that is, they have a strong “what”) but are difficult to precisely locate (that is they have a weak “where”). Other geographic objects are more relative. For example defining and locating land cover classes (e.g. forest) is more difficult as they have a weak “what” and are difficult to locate because they have a weak “where”.

Issues of indeterminacy and relativity, relating to different data conceptualisations, object definitions and semantics have been further discussed in relation to ontologies (Argawal, 2005) and the uncertainties associated with specific integration problems (Gahegan and Ehlers 2000; Kavouras and Kokla, 2002; Comber et al., 2004; Feng and Flewelling 2004; Comber et al., 2005c), where an understanding of the definitional and conceptual aspects of the data has allowed more representative assessments uncertainty to be made.

4.2 Data quality and uncertainty

The objective in both reporting data quality (through metadata) and in much uncertainty research is to enable users to assess the limitations of using a particular dataset. As it is difficult to anticipate every use case, any producer oriented data quality description will at some be found to be inadequate. As a result of this paradox Chrisman recommended that the need for user experience of the data be included in the specification of data quality in the original report from the standards committee (DCDSTF, 1988). This was not included in the final specifications for the Spatial Data Transfer Standards that were included in the metadata content standards (FGDC, 1998). This omission has been propagated through most standards specifications (Buehler and McKee, 1998), despite the fact the need for user assessments were first identified in the draft standards in 1988. FGDC was an operationalised version of what came out of this committee (Chrisman, pers comm.)

The ‘Big 5’ have dominated data quality reporting, standards (e.g. STDS, 1994; FGDC 1998) and reference works (e.g. Guptill and Morrison, 1995). Work describing the use of data quality parameters in actual applications is rare. DeBruin and Hunter (2003) describe an approach for assessing the value of different decisions about agricultural payment to farmers. A financial value was derived from the time stamp of remote sensed used to determine if a field was ploughed before a certain date, and the relative value of inspecting it. DeBruin *et al.* (2001) describe an application that assesses the value of two DEMs with respect to the extent they control an error process for determining the volume of sand required to build a new port area. There is an “expected value of control” from being able to control how an uncertainty such as data positional accuracy / error resolves. In both of these examples, decision analytical techniques (a cost benefit analysis in DeBruin and Hunter (2003) and an expected value of control in DeBruin *et al.* (2001)) are used to make an informed trade-off between the improved decision quality and increased cost. That is a real value is placed on the somewhat uncertain decision to use the data or not. In neither case is the quality measure of the information expressed in terms of any of the STDS / FGDC concepts. Rather an explicit value is placed on the assessment of uncertainty in using spatial data in the context of decision making.

5. Conclusions

In this paper we have identified a number of hitherto separate developments: increased use of spatial data, the ubiquity of geographic information applications, increased access to spatial data through cyber infrastructures and a decline in the tradition of metadata reporting (the survey memoir). In parallel we note a number of spatial data characteristics:

- Data are collected for all sorts of reasons, but we can't predict future use or value;
- The "real" metadata often resides with the individual scientist (and often only in their memory);
- There is increasing pressure on commercialisation, IPR, "spin-offs" etc. reduces desire to collaborate / cooperate;
- There is a naive belief that technology (e-science, grid) leads to a view that integration is trivial.

We have argued that in order to facilitate more robust data usage metadata needs to be expanded to link data quality and uncertainty assessments. A recent workshop on activating metadata for geographic information proposed that metadata include:

- Information describing the data conceptualisations;
- User generated metadata.

This is important for a new tranche of spatial data users to ensure that they avoid conceptual mismatches between the user and data ontologies (Ahlqvist 2004; Wilson 2004; Vasseur et al., 2004) they are able to assess the suitability of data for their use and that they are able to understand and assess uncertainties associated with any integrating activity.

If we are going to benefit from the new cyber infrastructures for spatial data then we need to provide metadata describing the data concepts and tools to allow users to assess data fitness for their application. Work to develop tools to populate and mine expanded metadata (user experience, deeper user understanding) is ongoing.

Acknowledgements

The ideas in this paper were developed through a workshop sponsored by the National Institute for Environmental eScience (www.niees.ac.uk).

References

- Agarwal, P. (2005) 'Ontological considerations in GIScience', International Journal of Geographical Information Science (IJGIS), Vol. 19, No. 5, pp. 501 – 535
- Ahlqvist O. 2004. A parameterized representation of uncertain conceptual spaces. *Transactions in GIS* 8(4) 493-514
- 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., Ness, M.J., 1993. *Countryside Survey 1990: main report. Countryside 1990 Series: Volume 2*. Department of the Environment, London.

- Bishr, Y., 1998. Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science*, 12(4):299-314.
- Buehler, K., and L.McKee (eds). 1998. *The Open GIS Guide: Introduction to Interoperable Geoprocessing and the OpenGIS Specification*. 3rd Edition, Draft Specification, <http://www.opengis.org/techno/guide/guide980615/Guide980601.rtf>
- Burrough, P.A. and Frank, A.U., (eds.) 1996. *Geographic Objects with Indeterminate Boundaries*. (Taylor & Francis, London).
- Cheng T., 2002. Fuzzy objects: Their changes and uncertainties *Photogrammetric Engineering and Remote Sensing* 68 (1): 41-49
- Comber, A., Fisher, P., Wadsworth, R., (2003) Actor Network Theory: a suitable framework to understand how land cover mapping projects develop? *Land Use Policy*, 20: 299–309.
- Comber, A., Fisher, P., Wadsworth, R., (2004). Integrating land cover data with different ontologies: identifying change from inconsistency. *International Journal of Geographical Information Science*, 18(7): 691-708.
- Comber, A., Wadsworth, R. and Fisher, P., (2005c). Reasoning methods for handling uncertain information. Pp. 153-168 in *Qualité de l'information géographique*, (eds. Rodolphe Devillers and Robert Jeansoulin), IGAT, Hermes, France
- Comber, A.J., Fisher, P.F. and Wadsworth, R.A., (2002). Creating Spatial Information: Commissioning the UK Land Cover Map 2000. pp. 351-362 in *Advances in Spatial Data*, (eds. Dianne Richardson and Peter van Oosterom), Springer-Verlag, Berlin.
- Comber, A.J., Fisher, P.F., Wadsworth, R.A., (2005a). You know what land cover is but does anyone else?...an investigation into semantic and ontological confusion. *International Journal of Remote Sensing*, 26 (1): 223-228
- Comber, A.J., Fisher, P.F., Wadsworth, R.A., (2005b). What is land cover? *Environment and Planning B: Planning and Design*, 32:199-209.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37(1): 35-46.
- DCDSTF (Digital Cartographic Data Standards Task Force). 1988. The proposed standard for digital cartographic data. *American Cartographer* vol 15 (1), pp. 9-140.
- de Bruin, S, Hunter, G.J., (2003). Making the trade-off between decision quality and information cost. *PHOTOGRAMM ENG REM S* 69 (1): 91-98
- de Bruin, S., A.Bregt, and M. van de Ven. 2001. Assessing fitness for use: the expected value of spatial data sets. *International Journal of Geographical Information Science* vol 15 (5), pp. 457-471.
- Devillers, R, Bédard, Y and Jeansoulin. R, 2005. Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use Within GIS, *Photogrammetric Engineering & Remote Sensing*, Vol. 71(2): 205–215.

- Devilleers, R., Bédard, Y., Jeansoulin, R. and Moulin. B., (Submitted). Quality information system to support experts in the assessment the fitness for use of geospatial data. Submitted to *International Journal of Geographical Information Science*
- Feng C-C and Flewelling D M 2004 Assessment of semantic similarity between land use and cover classification systems. *Computers, Environment and Urban Systems* 28: 229–46
- FGDC (Federal Geographic Data Committee). 1998. *Content Standard for Digital Geospatial Metadata*, FGDC-STD-001-1998, National Technical Information Service, Computer Products Office, Springfield, Virginia, USA.
- Fisher P, Wood J, Cheng T. 2004 Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute Of British Geographers* 29 (1): 106-128
- Fisher, P., Comber, A., and Wadsworth, R., (2005). Approaches to Uncertainty in Spatial Data. Pp. 49-64 in *Qualité de l'information géographique*, (eds. Rodolphe Devillers and Robert Jeansoulin), IGAT, Hermes, France
- Fisher, P.F. 1998. Improved Modelling of Elevation Error with Geostatistics. *GeoInformatica* vol 2 (3), pp 215-233
- Fisher, P.F. 1999. Models of Uncertainty in Spatial Data. In *Geographical Information Systems: Principles, Techniques, Management and Applications*, edited by P.Longley, M.Goodchild, D.Maguire, and D.Rhind (Wiley and Sons, New York) vol 1, pp 191-205.
- Fisher, P.F. 2003. Multimedia Reporting of the Results of Natural Resource Surveys, *Transactions in GIS*, 7 309-324.
- Frank, A.U. 2001. Tiers of ontology and consistency constraints in geographical information systems, *International Journal of Geographical Information Science*, 15 (7): 667-678.
- Fuller, R.M., G.B., Groom, A.R. Jones, 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., G.M. Smith, J.M. Sanderson, R.A. Hill, and A.G. Thomson, 2002. Land Cover Map 2000: construction of a parcel-based vector map from satellite images. *Cartographic Journal*, 39: 15-25.
- Fuller, R.M., Wyatt, B.K. and Barr, C.J., 1998. Countryside Survey from ground and space: different perspectives, complementary results. *Journal of Environmental Management*, 54: 101-126.
- Gahegan M and Brodaric B 2002 Examining Uncertainty in the Definition and Meaning of Geographical Categories. In *Proceedings of the Fifth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Melbourne

- Gahegan, M. and Ehlers, M. 2000: A framework for the modelling of uncertainty between remote sensing and geographic information systems. *ISPRS Journal of Photogrammetry & Remote Sensing*, 55: 176-188
- Haines-Young, R.H., Barr, C.J., Black, H.I.J., Briggs, D.J., Clarke, R.T., Cooper, A., Dawson, F.H., Firbank, L.G., Fuller, R.M., Furse, M.T., Gillespie, M.K., Hill, R., Hornung, M., Howard, D.C., McCann, T., Morecroft, M.D., Petit, S., Sier, A.R.J., Smart, S.M., Smith, G.M., Stott, A.P., Stuart, R.C., & Watkins, J.W., 2000. *Accounting for nature: assessing habitats in the UK Countryside*. (DETR, London).
- Harvey F 2000 The social construction of geographical information systems. *International Journal of Geographic Information Science* 14: 711–23
- Harvey, F., W. Kuhn, H. Pundt, Y. Bishr, and C. Riedemann, 1999. Semantic interoperability: A central issue for sharing geographic information. *Annals of Regional Science* 33(2): 213-232.
- Hunter, G.J., 2001. Spatial Data Quality Revisited, *Proceedings of GeoInfo 2001*, 04–05 October, Rio de Janeiro, Brazil, pp. 1–7.
- ISO, 2003. *19115:2003 Geographic Information – Metadata*. (International Organization for Standardization, Geneva).
- Kavouras, M, and M. Kokla, 2002. A method for the formalization and integration of geographical categorizations. *International Journal of Geographical Information Science*, 16 (5): 439-453.
- Klir, G.J., and B.Yuan. 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Englewood Cliff
- OGC, 2005. OpenGIS Consortium. URL: <http://www.opengis.org/> (last date accessed: 10 June 2005).
- Pundt, H. and Y. Bishr, 2002. Domain ontologies for data sharing-an example from environmental monitoring using field GIS. *Computers and Geosciences*, 28 (1): 95-102.
- Salgé, F. 1995. Semantic Accuracy. In *Elements of Spatial Data Quality* edited by S.C.Guptill and J.L.Morrison. (Elsevier, Oxford), pp. 139-151
- Smith, B. 1995. On Drawing Lines on a map. *Spatial Information theory: Lecture Notes in Computer Science*, 988 475-484.
- Smith, B. and Mark, D., 2001. Geographical categories: an ontological investigation. *International Journal of Geographical Information Science*, 15 591- 612.
- Smith, B., 2001. Fiat Objects, *Topoi*, 20 131-148.
- Spatial data Transfer Standard (SDTS) (1998). New York: American National Standards Institute
- Vasseur B., van de Vlag D., Stein A., Dilo A., 2004. Quality-aware ontology for dynamic spatial objects with indeterminate boundaries, in: 3rd International Symposium On Spatial Data Quality, Frank, A.U., (ISSDQ conference) Bruck Am Der Lehta, Austria, April 15-17 2004.

- Visser, U., H. Stuckenschmidt, G. Schuster, and T. Vogele, 2002. Ontologies for geographic information processing. *Computers and Geosciences*, 28, 103-117.
- Wilson, N. (2004). The beginnings of a logical semantics framework for the integration of thematic map data. *International Journal of Geographical Information Systems*, 18(4) : 391-417.