Exploring the use of Crowd Generated Geospatial Content in Improving the Quality of Ecological Feature Mapping

Laura Rhiannon Kinley, BSc

Thesis submitted to The University of Nottingham

for the degree of Master of Philosophy

October 2015

School of Engineering,

in collaboration with

the Horizon Centre for Doctoral Training

Abstract

Generalisation and uncertainty in ecological data are hugely problematic for environmental monitoring and modelling analyses and as such, land use change calculations, global climate change monitoring and food security analyses suffer. Low resolution, incomplete and often inconsistent land cover data sets signify that the digital representation of the Earth's biophysical surface is very poor. User generated content is gaining increasing traction within the scientific community; researchers are increasingly looking to Volunteered Geographic Information as a solution to fulfil the inadequacies of authoritative information as it can often be obtained more frequently and with less expense.

Throughout this thesis, the use of VGI for land cover classification and Phase 1 Habitat classification is trialled with the view that crowd generated content could make appropriate training data for remote sensing classifications. In the field of ecology, VGI has mostly been used in the context of validating existing data and a great deal of research is required into how ecological data can be collected most effectively and the extent to which it can be filtered to achieve usable datasets. This thesis examines the utility of both open web data (passively submitted to sources such as Geograph) and data obtained via directed crowdsourcing as ancillary data for remote sensing classification.

The open web land cover data sources analysed (OpenStreetMap and Geograph) have some very detailed and potentially minable information, though require substantial edits and filtering prior to use. Directed crowdsourcing approaches are shown to have a stronger match with their authoritative counterpart and several factors (characteristic profiling, classification time and user confidence in classification) are found to be useful ways of filtering the crowd contributions to achieve the best quality data set. The investigations show how datasets can be synthesised that, whilst not reaching the accuracies of authoritative data, could be seen as an asset to ecological monitoring and inventory.

Extending the directed crowdsourcing method to use data collected iteratively *in situ* via a mobile application forms the focus of future research recommendations. Active learning is also suggested as a complementary method which could reduce the number of training samples required by focusing sampling efforts on where existing classifications are most uncertain and where the contributor's knowledge profile is most appropriate.

Contents

Abstract 2

Author's publications and presentations 8

Online Publications 8

Presentations & Posters 8

Acknowledgments 9

List of Figures 10

List of Tables 12

Chapter 1 | Introduction 13

1.1 The Importance of Optimising the Mapping of Ecological Features 13

1.2 Uncertainty in Existing Ecological Mapping Methods 13

1.3 User Generated Geospatial Information as a Solution 14

1.4 Thesis Outline 15

1.5 Research Approach 15

1.6 Aims and Objectives 16

1.6.1 Thesis Aim 16

1.6.2 Research Objectives 16

Chapter 2 | Literature Review 17

2.1 Introduction 18

2.2 Importance of Land Cover Data 18

2.3 Traditional Land Cover Mapping Methods 19

2.3.1 Pixel Based Image Classification 20

2.3.2 Object based Image Classification 20

2.4 Assessing the Quality of Land Cover Maps 21

2.5 Limitations of Existing Land Cover Mapping Methods 22

2.5.1 Interoperability 22

2.5.2 Levels of Agreement 24

2.5.3 Issues with Existing Land Cover Data from an End User perspective 25

2.6 Crowd Generated Geospatial Content 27

2.6.1 Origins of VGI, Citizen Science and Crowdsourcing.....27
2.6.2 Defining VGI, Citizen Science and Crowdsourcing.....28
2.6.3 Directed Crowdsourcing.....29
2.6.4 Open Web Crowd Generated Content 30
2.7 Emergence of User Generated Content: Web 2.0, & Proliferation of Geospatial Technology 30
2.8 Volunteered Geographic Information as a Research Tool 31
2.9 Uncertainty and Error in VGI 32
2.9.1 Data Quality Concerns.....32
2.9.2 Technology Limitations 33
2.9.3 Sampling Bias.....33
2.9.4 Project Futurity.....34
2.10 Assessing Spatial Data Quality.....34
2.10.1 Internal Quality.....35
2.10.2 External Quality 35
2.10.3 Metrics to Measure Quality 36
2.11 Previous Research on the Quality of Ecological VGI 37
2.11.1 Citizen Science Projects Sourcing Ecological Data.....37
2.11.2 Can VGI Address the Flaws of Ecological Sampling? 39
2.12 Approaches to Address VGI Quality 40
2.12.1 Fusing VGI and Remote Sensing.....41
2.12.2 Variance in Structure & Codes of Practice 41
2.12.3 User characteristic filtering.....42
2.12.4 Ancillary information.....43
2.13 Key Research Gaps 43
Chapter 3 | Assessing the Potential for Open Web Data in Enhancing the Quality of Authoritative Land Cover Mapping.....45
3.1 Introduction 45
3.1.1 Information Mined from Existing Projects as a Data Source.....45
3.1.2 Previous Research Using Existing Open Web Data.....46

3.2 Aims & Objectives 47
3.2.1 Chapter Outline 48
3.3 Key Materials and Data Preparation 48
3.3.1 OpenStreetMap Data 48
3.3.2 Geograph Data 51
3.3.3 CEH Land Cover Map 2007 51
3.3.4 Ordnance Survey MasterMap 53
3.3.5 County boundary 53
3.4 Study Location 53
3.5 Methods 54
3.6 Results and Interpretations 55
3.6.1 OSM Polygon Analysis 55
3.6.2 Spatial Distribution of Geograph Data: 61
3.6.3 Temporal Distribution of Geograph Data 62
3.6.4 Accuracy of Attribute Data and Spatial Scale of Crowd Interpretation 64
3.6.5 Scale of Land Cover Attribution 73
3.6.6 Can Accurate Contributors be Identified? 75
3.7 Discussion 76
3.7.1 Accuracy of Open Web Data 76
3.7.2 Usability of Open Web Data 79
3.7.3 Proposed Methods for Enhancing Open Web Data Usability 80
3.8 Study Critique 81
3.8.1 Spatial scale of investigation 81
3.8.2 Choice of Data Used 81
3.8.3 Methods Used to Analyse Data 82
3.9 Conclusions 82
Chapter 4 | Refinement of Approaches to Directed Crowdsourcing 84
4.1 Introduction 85
4.2 Aims and Objectives 85

4.3 *In Situ* Land Cover Classification 85
4.3.1 Objectives 85
4.3.2 Preliminary Study Location 86
4.3.2 Methodology 86
4.3.4 Observations 87
4.4 Development of a Land Cover Crowdsourcing Web Page 91
4.4.1 Objectives 91
4.4.2 Methodology 91
4.4.3 Results of Web Classification Experiment 95
4.5 Summary and Recommendations for Next Iteration of Crowdsourcing Site 98
4.5.1 Barriers to the Successful Attribution of Land Cover 98
4.5.2 Usability of the Crowdsourcing Site & Classification Systems 99
Chapter 5 | Assessing the Utility of Directed Crowdsourcing in Habitat Classification 100
5.1 Introduction 101
5.2 Aims 102
5.3 Study Rationale 103
5.3.1 Choice of Research Approach 103
5.3.2 Selection of Classification System 103
5.3.3 Selection of Study Location 104
5.3.4 Selection of Classification Tasks 105
5.3.5 Selection of Participants 111
5.4 Methods 111
5.4.1 Participant Recruitment 111
5.4.2 Data Collection 114
5.4.3 Analysis 118
5.4.4 Adaptation of Woodcock & Gopal Method 118
5.5 Results and Analysis 119
5.5.1 Overall Accuracies 119

5.5.2 Variance between Participant Groups 120
5.5.3 Grouping by Level of Training 126
5.5.4 Variance in Inter-Class Accuracies 126
5.5.5 Correlations with Ancillary Information 130
5.5.6 Fuzzy Assessment 134
5.5.7 Uncertainty in Classification (Qualitative Analysis of Open Text Feedback) 135
5.6 Discussion of Results 137
5.6.1 Directed Crowdsourcing for Habitat Mapping 137
5.6.2 The Importance of Prioritising Data Collection 138
5.6.3 Label Weighting 139
5.6.4 Qualitative Interpretations 141
5.7 Critique of Study 141
5.7.1 Simulated Crowdsourcing Environment 141
5.7.2 Participant categorisation 142
5.7.3 Ecologist 'Ground Truth' 142
5.8 Summary 143
Chapter 6 | Synthesis of Findings and Future Research Directions 144
6.1 Introduction 145
6.2 Summary of Findings 145
6.2.1 Assessing the Potential for Open Web Data in Enhancing the Quality of Authoritative Land Cover Mapping 145
6.2.2 Assessing the Utility of Directed Crowdsourcing in Habitat Classification 146
6.3 Next Steps for Future Research 147
6.3.2 Exploring the Motivations behind Directed Crowdsourcing 148
6.3.3 In-Situ Directed Data Collection 148
6.4 Conclusions 149
References 150

Author's publications and presentations

Online Publications

Kinley, L. (2013). Towards the use of Citizen Sensor Information as an Ancillary Tool for the Thematic Classification of Ecological Phenomena. Proceedings of the 2nd AGILE (Association of Geographic Information Laboratories for Europe) PhD School 2013. *

Kinley, L., 2013. Assessing the potential for crowdsourced geospatial content to enhance the quality of authoritative land cover mapping. Nottingham, UK: AGI. **

Presentations & Posters

Ordnance Survey – Research Seminar 2012

Assessing the Potential for Crowdsourced Geospatial Content to Enhance the Quality of Authoritative Land Cover Mapping – AGI GeoCommunity 2013

East Midlands Universities Research Student Conference 2013

Assessing the Accuracy of Volunteered Geographic Information Derived Habitat Classification - International Symposia on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 2014 (Best Student Paper award)

Excerpts from the above publications are adapted and used within Chapter 2* and Chapter 3**

Acknowledgments

Laura Kinley is supported by the Horizon Doctoral Training Centre at the University of Nottingham (RCUK Grant No. EP/G037574/1) and is supervised by Professor Mike Jackson and Professor Giles Foody.

The author would like to extend many thanks to Glen Hart at the Ordnance Survey for his invaluable advice and support both during and following her internship (during which some of this work was undertaken) and her supervisors for their advice.

Thank you to Mercedes Torres and Ordnance Survey for the provision of the ground-based photography and authoritative classifications used within this thesis. Many thanks is also due to the many participants who took part in the classification experiments and those who uploaded data to the open web, in particular Jerry Clough, (OSM Nottingham meet up coordinator) who has also been a great friend.

Laura would like to extend huge gratitude to her parents, Kate and Steve and her fiancée Tom. Their endless patience, support and advice have enabled her to get through this ☺.

Data Licensing:

Ordnance Survey MasterMap data: © Crown Copyright/database right 2015.

Geograph and OpenSteetMap are licensed under the Creative Commons Attribution-ShareAlike 2.0 license (CC-BY-SA).

Land Cover Map 2007 data was licensed for use with the OS by the Centre for Ecology and Hydrology.

List of Figures

Figure 2.1 Minimum Physical Requirements of a “Forest” - Data from Lund (2004) 24
Figure 3.1: Map of OSM Contributions for the Hampshire Area with “Natural” Tag 49
Figure 3.2: Map of OSM Contributions for the Hampshire Area with “Surface” Tag 50
Figure 3.3: Map of CEH Land Cover 2007 Data for Hampshire 52
Figure 3.4 Map showing Study County in Context of Great Britain 54
Figure 3.5 Visual Comparison of OS and OSM Data 56
Figure 3.6 Visual Comparison of OS and OSM Data 56
Figure 3.7 Map Showing the Extent of OpenStreetMap’s Land Cover Data 57
Figure 3.8 Spatial Join of OSM Centroid Points with OSMM Topo Layer 58
Figure 3.9: Extent comparisons between OSM and OSMM 59
Figure 3.10 Comparison of Equivalent Polygon Extents by OSMM Featute Type 61
Figure 3.11 Geograph Coverage by Grid Square 62
Figure 3.12 Temporal Spread of Geograph Images 63
Figure 3.13: Spread of Geograph Images with Appropriate Temporal Attributes 64
Figure 3.14: The Distribution of the 129,253 Geograph Points within Hampshire 65
Figure 3.15 Geograph Points that match LCM Data 71
Figure 3.16 OSM Points that match LCM Data 71
Figure 3.17 OSM & Geograph terms that Spatially Correlate with LCM 72
Figure 3.18 Proportions of Data at the Various Hierarchical Scales 74
Figure 3.19: Top 5 Users with Matches with OSMM 75
Figure 3.20: Top 5 Users with Matches with LCM 76
Figure 3.21 Grid Square 3208 – Variance Between Photographer’s Interpretations of Land Cover 78
Figure 4.1: Fieldwork Location 86
Figure 4.2: Crowdsourcing Webpage Screen Captures 92
Figure 4.3 Average Percentage of Photos Correctly Classified by Category 96
Figure 5.1 Sites of Study 104
Figure 5.2 Example of Ecologist Field Notes 107
Figure 5.3 Classification Images Set 1 (Easy) 108
Figure 5.4 Classification Images Set 2 (Intermediate) 109
Figure 5.5 Classification Images Set 3 (Difficult) 110
Figure 5.6 Twitter Account used to Promote the Experiment 113
Figure 5.7 User Guidance Web Page 115
Figure 5.8 Example Classifications 116

Figure 5.9 Boxplot to Show the Variance in Classification Accuracy per Image by Classification Level 120
Figure 5.10 Group membership of Participants 121
Figure 5.11 Boxplot to Show the Variance in Classification Accuracy per Annotator by Classification Level and Participant Grouping 121
Figure 5.12: Graph to Show Broad Habitat Classification Accuracy by Participant Grouping 124
Figure 5.13: Graph to Show Sub Habitat Classification Accuracy by Participant 124
Figure 5.14: Scatter Chart to Show the Mean Count of Clicks in Relation to the Classification Accuracy per Image at both the Broad and Sub Hierarchical Classification Levels 133
Figure 5.15: Relationship between Classification Accuracy and Label Confidence 134

List of Tables

Table 2.1 INSPIRE Land Cover Data Specification Guidelines 21
Table 2.2 Land Cover Data Requirements 25
Table 2.3 Internal Quality Measures 35
Table 2.4 External Quality Measures 36
Table 3.1 Variance in polygon extents by OSMM featute type 60
Table 3.2 Unique terms extracted from Geograph 66
Table 3.3 Unique terms extracted from OpenStreetMap 69
Table 3.4 Counts of OSM & Geograph Terms that match with LCM 72
Table 3.5 Proportions of Data at the Various Hierarchical Scales 74
Table 3.6 Table to Show Top Contributors with Spatially Correlated Matches 75
Table 4.1 Issues Identifying Classification due to Spatial Scale 87
Table 4.2 Issues Identifying Classification due to Classification Inappropriateness 88
Table 4.3 Issues Identifying Classification due Insufficient Knowledge 89
Table 4.4: Breakdown of Participants 94
Table 4.5: Average and Total Scores by Classification 95
Table 4.6: Average and Total Scores by Knowledge 95
Table 4.7: Positive Comments Regarding Classification Methodology 97
Table 4.8: Negative Comments Regarding Classification Methodology 98
Table 5.1 - Combined Image Difficulty Level Conditions 107
Table 5.2 Breakdown of Participants 114
Table 5.3 Woodcock and Gopal Analysis 118
Table 5.4 Mean Accuracy per Annotator Across all Images 122
Table 5.5: Cohen's Kappa Analysis 122
Table 5.6: ANOVA – Participant Groupings 123
Table 5.7 Post Hoc Test – Tukey (Broad & Sub Levels) 123
Table 5.8: Broad Class Accuracy 126
Table 5.9: Confusion Matrix for Broad Level Classifications 127
Table 5.10: Confusion Matrix for Sub Hierarchical Level 128
Table 5.11 Analysis of Classification Guide Attentiveness 131
Table 5.12 Liked Aspects of Help Resources 132
Table 5.13 Drawbacks of Help Resources 132
Table 5.14: Pearson's Correlation Coefficient Analysis 133
Table 5.15 Pearson's Correlation Coefficient Analysis 134
Table 5.16: Percentage Accuracies and Percentage Improvements 135
Table 5.17 Comments from Participants 136

Chapter 1 | Introduction

1.1 The Importance of Optimising the Mapping of Ecological Features

Improving the accuracy of large scale ecological feaures has long been a scientific line of enquiry (Anderson, 1976, Foody, 2002, Fritz et al. 2012). The collection of high quality digital biophysical data is becoming increasingly necessary as researchers work towards solutions to abating the world's social and environmental problems such as determining the impacts of land use change, monitoring global climate change (Verburg et al. 2011) and food security analyses (Fritz et al. 2009). Such data contributes to the development of environmental health indicators and towards models to monitor and predict the condition of the environment. Jones (2008) summarises the utility of using land cover, digital biophysical data and in situ data in conjunction to enable the interpretation of environmental status, impact analysis, vulnerability and risk assessments, ecological forecasting and alternative future landscape analyses.

Land cover maps are portrayed as the most important sources of baseline terrestrial information, providing essential data for the work of many users, e.g. the Millennium Ecosystem Assessment (MEA, 2005). Despite the importance of land cover data and its subsequent applications, classification uncertainty is a prominent issue (Fritz et al. 2009); current land cover data sets are of relatively low resolution (with only a 100m resolution for the EEA's most recent CORINE dataset), incomplete in areas and contrasting in their interpretation of land cover definitions (as portrayed by a classification contradiction map developed in Fritz et al. (2011)).

1.2 Uncertainty in Existing Ecological Mapping Methods

Despite the huge importance of systematic and scientifically defensible biodiversity sampling, several factors limit the quality of ecological analyses; the accuracy of the model employed and the quality of distribution data procured signify that great inaccuracies can be present. The spatial resolution, the temporal regularity of data collection, the units of information collected and the level of detail to which data is collected varies from study to study (Hampton et al. 2013). A trade-off is presented between the need for accurate data (in terms of spatial resolution, count of detected elements and biological validity) and the cost in both time and technical skill required to gather it (Godet et al. 2009).

Inadequate sample size is a very common inductor of error in ecological data sets; changing the sample size from 50 to 300 has been shown to alter habitat conservation target analyses by up to 45% (Metcalfe et al. 2013). Several global land cover products exist including GLC-2000, GlobCover and MODIS and they have often been the data of choice for many species distribution, environmental change and food security analyses. Despite our reliance on them they have been found to have large thematic discrepancies between them (Fritz et al. 2012); existing land cover analyses often "lack crucial and accurate data" (Fritz, et al 2009). The datasets also temporally and thematically inhibit accurate modelling (Schepaschenko et al. 2012) particularly in developing countries where temporally accurate data is needed "immediately" for the monitoring of ecosystem services (Andelman, 2011).

Data recordings that are *in situ* (traditionally obtained by ecologists) are required to improve the accuracy of these products that have been derived from classifications of satellite imagery. Ground data is unfortunately logistically challenging and time consuming to acquire (Townshend et al. 2012) meaning that there is usually a reliance on the interpretation of satellite imagery to provide the training data from which classification algorithms allocate a land cover class. It is essential to develop economically viable ways of attaining more ground truth samples so that global land cover models can be developed with confidence.

1.3 User Generated Geospatial Information as a Solution

User generated geospatial content is geo enabled information submitted by volunteers allows us to "observe, analyse and visualise" (Dalby, 2012) our ecosystems and signifies that scientists now have access to a wealth of data that can be more accurate, more up-to-date and more complete than authoritative sources (Devillers et al. 2012). Despite great quality hurdles (Haklay, 2010), citizen scientists are being entrusted with increasingly complex tasks; contributors have matured from the completion of basic identification tasks such as the Christmas Bird Count to contributing to complex issues which are traditionally confined to academia. The crowd are now classifying galaxies on 'Galaxy Zoo', determining protein structures through a 'FoldIt' game, collecting hydraulic measurements (Lowry and Fienen, 2013), taking measurements of air quality and determining the structural integrity of buildings following natural disasters (Rebbapragada and Oommen, 2011). User generated geospatial content has an emerging role in research centres and is clearly providing data that is essential to scientific monitoring and discovery and humanitarian causes.

Volunteered information can play an integral role in areas where surveying would not otherwise be completed (such as the surveying of highly detailed, transient and low economic priority features), or where it can help inform and prioritise authoritative mapping efforts. User generated content could be of particular value in the context of using volunteered information as change intelligence for the mapping of the natural environment where, in the UK, a cyclic revision system denotes that rural and moorland areas are only authoritatively inspected systematically on a five-year basis (Ordnance Survey, 2015). Problematically, user generated geospatial content typically exhibits considerable quality variance ranging from imprecise digitisation to poor label accuracy. As such, off the shelf usage for planning and scientific purposes is challenging from a data quality perspective and methods of refining the data must be developed.

1.4 Thesis Outline

Chapter 2 presents a comprehensive review of the literature around ecological VGI, ascertaining the quality of crowdsourced data and looking at methods of ensuring that the quality is as high as possible. Study one (**Chapter 3**) explores a quantitative analysis of Open Web data undertaken in conjunction with Ordnance Survey. Key flaws from the use of Open Web data for land cover mapping help to suggest directed crowdsourcing as a more suitable approach through **Chapters 4 and 5**. **Chapter 4** consists of two preliminary studies which are used to aid the design of the second study, which comprises of the design, use and evaluation of a directed crowdsourcing website. **Chapter 5** details the analysis of the crowd habitat classifications collected via the website and looks at some ways of filtering the data to provide more accurate classification sets. **Chapter 6** offers a summary of key findings from the two core studies and offers some future research directions. The study methodologies are described and evaluated in the appropriate research chapter.

1.5 Research Approach

A “mixed methods” approach, as described by Creswell and Clark (2007), is used throughout this thesis. This approach combines both quantitative and qualitative techniques in order to gain a more comprehensive understanding of a research topic than if only one element was investigated (Cope and Elwood, 2009). It is an approach particularly relevant to this study due to the requirement to perform geographic

information analyses, determine accuracy and to understand the extent to which an appropriate level of engagement with the user is achieved. Open ended questionnaire answers are used in Chapters 4 and 5 where it is necessary to understand the ways in which the crowd interpret classification systems and the tasks which they are undertaking. Quantitative analyses are used throughout the thesis to gain an understanding of the accuracy of the broad approaches of open web data sourcing and directed crowdsourcing. Context derived from the qualitative data attained is used to understand any patterns in the quantitative results where appropriate.

1.6 Aims and Objectives

1.6.1 Thesis Aim

The aim of this thesis is to explore and address the gaps highlighted within this literature review (Section 2.13), specifically to determine the extent to which the addition of 'crowd generated' and open web geospatial content can be useful in improving the quality of thematic classification when mapping ecological classes and features.

The research aims to indicate what level of improvement might be expected given certain assumptions on availability of crowd-data, what form that improvement takes (i.e. whether that be improved accuracy in general terms, more up-to-date classifications, improvements for only certain types of land cover or only in certain special circumstances). The research aims to document how any such improvements may be achieved and the techniques adopted to achieve those improvements. The thesis also documents areas in which further research is recommended. The research should allow for a better understanding of the viability crowd generated content has for ecological research applications, particularly with regards to land cover and habitat data.

1.6.2 Research Objectives

The aim discussed in section 2.14.1 is addressed through a number of objectives which relate to the analytical and discussion chapters outlined in section 1.4:

1. Establish the degree of accuracy required from ecological datasets and any concerns that the potential use of crowd generated content presents to users of ecological datasets.

2. Understand how the crowd identifies ecological features field *in situ* and develop the best practices for developing a web based crowdsourcing platform.

3. Obtain ‘passively generated’ open web data and establish whether it could provide any contribution over and above that of the authoritative data collection and classification systems.

4. Develop a system for harvesting crowd contributions in a directed manner in the context of Phase 1 Habitat Mapping and quality assessment of crowd contributions.

Chapter 2 | Literature Review

2.1 Introduction

This Literature Review aims to provoke a thorough understanding of GIS, Remote Sensing and Volunteered Geographic Information, further developing research perspectives. Within this Chapter, existing research will be reviewed and synthesised with the aim of determining gaps within the current literature. The identification of these areas of required research will shape the direction of the thesis, enabling the refinement of the research question and objectives.

2.2 Importance of Land Cover Data

Land cover maps are 'fundamental' to monitoring and analysing ecosystems at a landscape level (Yu et al. 2014, Parent and Volin, 2016). The FAO defines land cover as "the observed (bio) physical cover on the Earth's surface"; which should be "geographically explicit" allowing interoperability between different organisations (FAO 2000). Maps that use land cover data and other ecological datasets are integral to a number of operations including insurance and risk modelling, monitoring climate change and in ensuring natural resources are used appropriately (Verburg et al. 2011). Land cover is very different from land use; confusion between the two can often further exacerbate interoperability issues. Land use can be defined as the "activities and inputs people undertake in a certain land cover type to produce, change or maintain it" (FAO 2000).

Land cover is continually changing owing to both natural changes and anthropogenic influence (Ban et al. 2015). Land cover change is the greatest threat to biodiversity (Vitousek, 1994) and thus must be mapped and understood to a high standard. Land Cover and the changes that occur in relation to its extent affect the planet's biophysics, biogeochemistry and biogeography (Giri et al. 2013). The multi-annual recording and monitoring of biodiversity within protected sites and biologically important landscapes is therefore critical for detecting changes which affect the status, quality and resources within ecologically significant sites (Townsend et al. 2009 and King, 2002). A particularly interesting aspect of land cover mapping is the monitoring of ecotonal areas, owing to their vulnerability and the implications that change within them has upon biological diversity. Ecotones are transitional zones between adjacent ecosystems and ecological systems. They are areas of changeable land cover, housing species that are at the limits of their physical and competitive

tolerances (Wasson et al. 2013) signifying that they are very susceptible to environmental changes and of particular importance for ecological modelling.

As well as its importance in a range of environmental research domains, land cover data informs statisticians and policy makers. From a business perspective, land cover data is relevant to a large number of applications, including urban and transport planning (Suau-Sanchez et al. 2014), soil protection and mapping and the informing of agricultural policy (Van Zanten et al. 2014). The Rural Payments Agency relies on land allocation information to make informed decisions about the allocation of more than £2 billion annually (RPA, 2012). Land allocation is also integral to flood risk analysis (Brody et al. 2014) which has implications for insurance companies and in town planning.

2.3 Traditional Land Cover Mapping Methods

Remote sensing is the most widely used and most economical method of sourcing land cover data (Gong, 2012). Spectral signatures derived from the cover or phenomena on the Earth's surface are obtained via remote sensing. These measurements are usually collected via an airborne platform such as an aircraft or sensor equipped satellite. These sensors determine the level of electromagnetic radiation which is emitted by or reflected from the phenomenon being sensed (Aplin, 2004). The varying cover types on the Earth's surface and objects within the atmosphere reflect electromagnetic radiation to different extents which allows the determination of land cover type when the spectral signature of the cover is analysed. The advancement of remote sensing (RS) techniques is imperative to land cover mapping; developing more advanced sensors and better understanding how atmospheric conditions can affect spectral signatures can lead to improved knowledge of land cover extents and subsequently improve our ability to monitor and understand the environment (Aplin, 2004). Observing the earth remotely via remote sensing techniques facilitates ease in the mapping of land cover and use, resulting in reduced surveying costs and minimised *in situ* data collection (Hansen and Loveland, 2012). Validation of training sites (used to enable RS classification algorithms to learn spectral response patterns) are however still essential, maintaining a reliance on site visits (Saadat et al. 2011). Land cover characterization requires the use of effective classification algorithms which are in turn dependent on adequate training and minimum levels of human input. Various classification methods use different

algorithms to group pixels with similar characteristics. This enables the transformation of spectral signatures into a thematic land cover map. Remote sensing classification techniques can be grouped into two broad categories: object and pixel based classification.

2.3.1 Pixel Based Image Classification

Pixel based classification represents the classical method by which each pixel within an entire image is assigned a class based on the characteristics of that pixel alone. Per pixel classification can often result in a “salt and pepper effect” (Blaschke et al. 2000) whereby individual pixels are classified in an alternative manner to their neighbours despite their presence within the same physical object. The method thus shows a tendency to focus on individual spectral properties ignoring the spatial context. Pixel based methods were historically relevant whilst pixels were coarser than or similar in size to the corresponding object on the ground. As the spatial resolution of images has increased through the proliferation of advanced sensors, object based classification has become an increasingly used method for classifying imagery.

2.3.2 Object based Image Classification

In contrast, geospatial object based image analysis involves the reduction of pixels into a number of homogeneous segments via image segmentation rule sets based on one or multiple criteria. This is often helpful for very high resolution imagery. The compartmented pixels are then classified as a group based upon derived statistical information which is based on metrics such as the mean, minimum and maximum values per band, ratios and variance (Blaschke, 2010). Objects are able to carry very useful information in addition to pure spectral information; topology, form and context can all be derived from an object boundary. Change can be detected by evaluating the per object percentage, homogeneity, and form of the pixels, in comparison to the equivalent historical object stored within the database (Walter, 2004). Object based methods have been shown to provide more accurate results with regards to change detection owing to reductions in misregistration and the addition of contextual information (Johansen et al, 2010) though they are however reliant upon the accuracy of the segmentation process.

2.4 Assessing the Quality of Land Cover Maps

Land cover maps are generally assessed for accuracy in terms of the extent to which they compare to a known 'gold standard' dataset (Haklay, 2010). The data the map is compared to can be in the form of a point, pixel or block and an unbiased sampling strategy must be devised that takes into account the range of land cover classes (Stehman et al. 2012). Assessing the accuracy of a land cover map requires reference data that is often difficult to obtain owing to the remote locations of the areas being assessed. Another problem is also presented by the fact that even professional annotators can disagree with one another (Fonte et al. 2015). Recently proposed EU INSPIRE guidelines (INSPIRE, 2012) also indicate how best to measure the quality of land cover data (Table 2.1).

Table 2.1 INSPIRE Land Cover Data Specification Guidelines

Quality Measure and Sub Measure	Description
Completeness – Commission	Number of excess items in the dataset in relation to the number of items that should have been present.
Completeness – Omission	Number of missing items in the dataset in relation to the number of items that should have been present.
Conceptual consistency	Count the number of features and feature relationships which violate the conceptual schema of the dataset.
Logical Consistency – Domain consistency	Compare attribute of items within scope against acceptable attribute domain (acceptable attributes) and determine if any are outside the domain.
Logical Consistency – Format consistency	Degree to which data is stored in accordance with the physical structure of the dataset, as described by the scope.

Logical Consistency – Topological consistency	Correctness of the explicitly encoded topological characteristics of the dataset as described by the scope.

The UK's premier land cover product Land Cover Map 2007 (LCM2007) uses a generalisation process to map UK land cover into (approximately 10 million) land parcels, 99.5% of which were classified using automated methods alongside knowledge based enhancements. Validation test sets show the map to have an overall accuracy of 83% which is below the target accuracy of 85% set for land cover mapping (Thomlinson et al. 1992). The accuracy of existing Global Land Cover maps can vary between 67% and 81%, and is lower (10–50%) in some areas of the world (Tsendbazar et al. 2014).

2.5 Limitations of Existing Land Cover Mapping Methods

Unfortunately land cover is not mapped to a high standard globally (Foody, 2002), this is in part the result of a deficit of high resolution ground interpretations of land cover which are used to train remote sensing classification algorithms. The availability of land cover data is limited both temporally and spatially which is likely to reduce the prevalence of high resolution landscape analyses in the future (Tallis et al. 2012, Parent and Volin, 2016). The lack of interoperability between the land cover maps produced by varying organisations is also problematic. This deficit of reliable and high resolution land cover mapping impedes our ability to accurately monitor environmental change (Hassan et al. 2005).

2.5.1 Interoperability

Conceptualisations of geographic space vary on an individual and organisational level. Attempts to build interoperable national, continental (e.g. INSPIRE, SEIS), or even global (e.g. GMESS, GEOSS) environmental information sets and monitoring systems require ontologically harmonised spatial data with comprehensive coverage. Spatial Data Infrastructures (SDIs) rely upon the consistency of data which permits different systems to understand one another. This consistency can be described in terms of the spatial scale at which features are surveyed and represented cartographically, the format in which the data is stored, to the terminology or nomenclatures used to depict and classify geospatial features. Common methodologies for storing and representing data promote competition, simplify

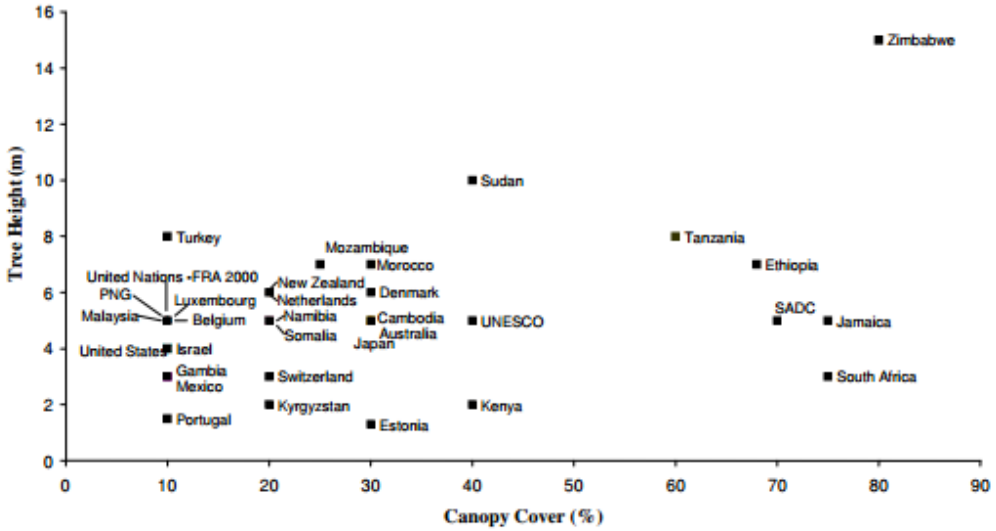
sharing and add transparency, enhancing the efficiency of data usage (as less time and effort is spent on interpreting the data). The production of data which meets these requirements is a challenge experienced across the various levels of public authority within Europe and globally.

In the UK there is no statutory requirement for the conduction of land cover surveys and therefore no single agency with the responsibility to produce and maintain it (Scottish Natural Heritage, 2012). Four national projects all address the issue of land cover: the full coverage Countryside Survey, Land Cover Map of Great Britain, the Land Cover of Scotland 1988 and the Welsh Habitat mapping project Gwylio (Lucas et al. 2011). The surveys vary in their approaches and consensuses owing to the various aims and objectives behind project initiation. The EU INSPIRE data standardisation legislation denotes that organisations must document their classification system; it does however permit data providers to distribute land cover data using their preferred classification systems and nomenclatures signifying that end users still have to translate between different systems – which may not be entirely possible and could lead to misinterpretation of data. The INSPIRE data specification depicts land cover as a “transverse theme” signifying that it should show consistency with other geographic data themes. The EEA funded EAGLE group hopes to achieve this by developing land cover products collaboratively and consistently through an open assembly of technical experts from different EU countries on land cover.

Land cover and land use mapping is particularly inconsistent; with the lack of interoperability prohibitive to efficient use. The development of the various land cover products was driven by various national and international organisations with different end goals using a plethora of techniques resulting in fundamentally differing products heavily influenced by the objectives (Pérez-Hoyos et al, 2012). Semantic transformations are often required as there is great variance in the way data are collected and the terms under which they are conceptualised (Comber, 2005). A segment of land can be classified as type A under one land cover classification system (a) and under type B in classification system b. Establishing linkages between types A and B (and vice versa) allows the two classification systems become interoperable and classifiable under both systems a and b (Baglatzi and Khun, 2012) . Problems can occur however when questions of scale are added: type A.2 and A.3 could be generalised to type A but de-generalising type A to a subset is impossible. Similarly where type A in classification system a may be a there may be no appropriate transferrable class for type A in system b. The variance in how organisations classify

one feature over another is exemplified particularly well by Lund's (2004) Graph of the minimum requirements for forest classification by country (Figure 2.1).

Figure 2.1 Minimum Physical Requirements of a "Forest" - Lund (2004)



2.5.2 Levels of Agreement

Existing comparisons between global land cover data sets show that there is a great level of discrepancy between products (Fritz et al. 2012). Pérez-Hoyos et al.'s (2012) comparison between CORINE, GLC2000, MODISLC and GlobCover using fuzzy methodology determined that there were large discrepancies between the four products with the best agreement between CORINE and GLC2000 (at 57%). Disagreement is particularly evident with heterogeneous and fragmented landscapes with a variety of cover classes and those with spectrally similar cover (e.g. grassland and eutrophied lakes). Fritz and See (2011) similarly determined large global discrepancies, finding a combined forest and cropland disagreement of 893 Mha between GlobCover and MODIS v.5 which equates to 20% of global cropland area.

Alongside the variances in sensor quality and semantic misalignment, a reason for the disparity often seen between global land cover data sets is the lack of sufficient *in situ* data, which is essential in classifier training. Ground data can be logistically difficult to access and very expensive to acquire (Foody and Boyd, 2012, Townshend et al. 2012) signifying that alternative methodologies may be necessary. Errors in ground data can have significant impact upon perceived accuracy and the outcome of change detection analyses; Foody (2009) discovered that imperfections in ground reference data produce systematic bias in the producer's estimates of change

occurrence and impact upon change detection accuracy. In a simulated experiment (with conditionally independent errors) the producer's accuracy of change detection was underestimated by 19% and the extent of change was overestimated by 6%, despite a ground reference data set with a high accuracy of 90% being used.

With regards to the UK, land cover map change detection between data sets is very complex as the classes used to within the Centre for Ecology and Hydrology's classifications have changed since 1990, the spatial structure is different between sets to permit the generalisation of data into polygons and reliable methods of segregating changes from those due to error and those due to spatial inconsistency are yet to be developed.

2.5.3 Issues with Existing Land Cover Data from an End User perspective

Increasingly there is a demand from users to have access to a resolution of land cover and land use data appropriate to their requirements. Currently the flagship Ordnance Survey MasterMap offers little in the way of land cover data as much of it derives from cartographic inheritance. A two tier classification is present in the form of nine core descriptive groups exist which denote the land cover further specified by the "descriptive term" attribute. The current provision is however, thought unsuitable for the requirements of many users as shown by the user requirements table below (Harding, 2011). Interview material from the Ordnance Survey has been grouped to understand current deficiencies in Ordnance Survey's land cover data and future requirements for land cover data (Table 2.2). The insufficient level of detail (both with regards to attribute and extent) within the data and interoperability issues are consistently raised as problems.

Table 2.2 Land Cover Data Requirements

Organisation & Use of LC data	Insufficiencies	Future Requirements
Local authority – LC data enhances flood risk assessment and selection of sites for waste disposal.	Woodland pasture information to <2ha is not available inappropriate spatial scale). Forestry commission data & OS legends don't tie up leading to confusion over data trust.	Height of hedgerows (to approximately 1 meter) to enhance knowledge of conservation sites Locations of woodland pasture Need a national standard.
Government agency – use LC information for Flood risk assessment.	"Woodland" is not meaningful… less generic land use terms are required e.g. extent and type of open areas.	Improved information regarding the extent of land cover. More detailed information e.g. Outer trunk boundaries, holes within the woodland, boundaries of conservation areas. The data must be more consistent.
Environmental consultancy - use it to indicate how vegetation slows down the movement of water and in ecological and infectious disease mapping.	Too much ambiguity in classes.	Extent and type of cover; would prefer something similar to National Land Use Database. Different models for summer and winter.
Police	OS Address layer 2 does not fulfil emergency planning requirements (it was developed to satisfy local authorities).	Detail in the height of trees, water bodies and wooded areas are particularly important.

Conservation and landscape enhancement	Vegetation information is too general, inaccurate and out of date.	Extent and type of land cover. Extent height and type of vegetation.
NGO		Extent of land cover (ponds wetlands springs). Positions of large trees Extent of edge of woodland canopy – must be up to date Dynamic boundaries e.g. cliffs must be as recent as possible.

2.6 Crowd Generated Geospatial Content

2.6.1 Origins of VGI, Citizen Science and Crowdsourcing

The discovery of spatial laws and patterns has long been a human fascination throughout the scientific community and beyond. Alfred Wallace was not a traditional scientist; his socio-economic background and lack of university education denoted that the intellectual elite of the time saw his work to be of questionable value (Shermer, 2002). The work was scrutinised and monitored by the scientific community (much as crowd contributions are scrutinised today) yet his expeditions showed rigour having independently conceived the same theory of evolution as scientifically acclaimed Darwin. Citizens conducting science as a methodology of data collection spans as far back as the mid-19th Century. Dudley Stamp provided perhaps the first example of volunteer recorded land allocation with his 1930's Land Utilisation Survey of Britain. The detailed map on a scale of 6 inches to the mile, published on 1 inch to the mile, is still used as a data source for habitat connectivity change analysis (Hooftman and Bullock, 2012). Such expeditions inspired a public thirst for exploration and enquiry, setting a precedent for public interpretation and discovery that thrives today owing to the development of technology.

Public Participation GIS (PPGIS) surfaced as a movement in the 1990s and was designed to open up geospatial tools and technology to a wider range of users following a social-theory critique of existing GIS practises. It developed as a mechanism for allowing communities and users a stake in the decision making

process, democratising GIS and emphasising the role of the citizen in the creation and evaluation of data (Sui, 2008). Decisions made using the collaborative “GIS 2” (named to mark the advancement of GIS into a participatory tool that utilised the elements of the web facilitating collaboration) would represent a diverse array of views and emphasise the role of all stakeholders with applications in urban planning and beyond (Schroeder, 1996).Technology has enabled PPGIS to expand from the desktop PC and research centre to online spaces permitting open experimentation, wide ranging public participation and increasing the accessibility of decision making.

Today the crowd is a group of volunteers that utilise GPS enabled technology or Web 2.0 technology to collect data, whether that be in a directed manner (e.g. analysing images on a computer) or collecting and updating in an ad-hoc way via iSpot or OpenStreetMap.

2.6.2 Defining VGI, Citizen Science and Crowdsourcing

It is important to understand the various branches of spatial user generated content or Volunteered Geographic Information (VGI) as it is developed in a variety of ways, each giving the contributor a different degree of authority, direction and freedom (Haklay, 2013). It is generally generated through crowdsourcing techniques where participants are given a defined set of goals and objectives to contribute towards (Horita, et al. 2013, Spyratos & Lutz, 2014). VGI is defined as “geospatial data that are voluntarily created by citizens who are untrained in the disciplines of geography, cartography or related fields” (Seeger, 2008). It is a term coined by Goodchild (2007) who views the world as having six billion human sensors that are able to contribute localised spatial information to improve existing maps. VGI can be generated in an ad hoc manner involving little or no governance being stored as open web data e.g. the submission of a geolocated photograph and a descriptive blurb to a photo sharing website under a creative commons license (Elwood et al. 2012). No individual or organisation actively asked for the photograph to be taken, or directed a particular theme for what it should capture, but the public domain information and associated geolocation could be scraped from the web and used to analyse and convey patterns or processes. VGI exists in a variety of formats and ranges from textual depictions, photographs and to detailed maps (Fonte et al. 2015).

Fundamentally citizen science is the engagement of volunteers in civic themes and research by persons that do not have any specific scientific background or training (Dickinson et al, 2010). It is an umbrella concept for a varied range of activities.

Recording ranges from active interrogation in hypothesis-led science to passively collecting information via the movement of sensors (Silvertown, 2009). The questions addressed by citizens can be highly-focussed research questions or simple exercises generating data of little scientific value, but also aiming to educate. Methodologies for using citizens in data collection include leading and directing people as data collectors to incubating subsets of participants to formulate ideas for projects, assessing the data and using the information themselves (Roy et al. 2012). The addition of geolocation to these observations permits analysis of this data as Volunteered Geographic Information.

Crowdsourcing is more direct in its goals and objectives whereas citizen science offers a more exploratory approach with broad long-term goals (Wiggins and Crowston, 2011). They are similar in that they both portray projects from an extremely wide spatial scope including micro, local focussed projects and global scale research or problem solving initiatives. Goodchild (2007) introduces the concept of "Volunteered Geographic Information" marking an "explosion" in the creation and maintenance of geographic data by the general populous, offering the first throwback to the late 19th Century definitions of citizen science.

2.6.3 Directed Crowdsourcing

Nichols & Williams (2006) refer to targeted monitoring as a structured strand of citizen sensing. Targeted sensing is a form of directed crowdsourcing which is often used within citizen science projects; an organizer sets specific tasks and contributors are directed to collect or analyse information to contribute towards accepting or declining a well-defined hypotheses. Contributors often participate in directed crowdsourcing because they care about a specific cause, such as bird counting. Gamification can also be a significant draw with researchers capitalizing on the pull of online and mobile gaming to collect scientific data. Towns Conquer is a game developed specifically to source vernacular names to enhance Spain's national mapping agency's database which has in game rewards to foster motivation (Castellote et al. 2013). Gamification has also been used to determine protein structures in the hugely popular FoldIt game (Khatib et al. 2011). Both gamification and the technique of paying crowd workers via services such as Amazon's Mechanical Turk (Kittur et al. 2013) unfortunately require a great deal of cost before any quantity of meaningful data is attained. In all directed crowdsourcing projects developing and sustaining a contributor's motivation to participate is imperative (Roy e al. 2012).

2.6.4 Open Web Crowd Generated Content

A second broad type of VGI is open web VGI, which is often passively created by Web 2.0 users and made freely available under open and creative commons licensing. This data can be mined from the web and used within research. It often can be used to form a serendipitous linkage with a scientific problem (Kinley, 2013), such as Flickr photograph tags being used as an indirect source of species distribution information. Passively generated VGI often has associated ancillary information pulled from Location Based Social Networks (LBSNs) and Web 2.0 content sharing platforms (Gao et al. 2011). Gschwend and Purves (2011) correlate open web data mined from Flickr with the undulations in a Digital Terrain Model, and De Longueville et al. (2009) also show how string data from LBSNs can be used to detect forest fires from geo-located tweets. Huang and Xiao, (2015) also use Twitter data within disaster zones aiming to prioritise areas for disaster relief teams. Whilst intentionally submitted, sources such as OSM can also be used as an open web source; the intention of OSM is to create a basic map of the entire planet's features, however it can also be used for other applications such as inferring land use information and population density.

2.7 Emergence of User Generated Content: Web 2.0, & Proliferation of Geospatial Technology

O'Reilly's (2005) Web 2.0 symbolised a change in the use of the Internet rather than a physical upgrade. It involved the increased dissemination of information and the development and the uptake of tools for collaboration provoking a socially active web with an ethos of sharing and development. Since then a number of web based services have proliferated, enabling the easier and economically viable creation of geospatial content and mapping. Cloud software and data distribution has permitted the support of spatiotemporal intensive applications (Yang et al. 2011) and the scalable use of complex tools and data on web based platforms. This has enabled a huge rise in the number of collaborative online geospatial tools.

Open Source has played a large role in the proliferation of home grown mapping and spatial analysis techniques with initiatives such as PostGIS, MapServer and OpenLayers providing free infrastructure for the storage, interactive publishing and rendering of spatial data respectively. The Open Source Geospatial Foundation's OSGeo-Live stack provides a self-contained installation free option making spatial

tools, publishing and sharing options accessible to a greater number of users (Leidig and Teeuw, 2015). The software is easy to modify signifying communities have developed and shared solutions within an expanding and tight knit community (Lakhani and Von Hippel, 2003). Open Source developments have and will continue to facilitate the increased development and use of user-edited maps (Silvertown, 2009).

By 2017 it is expected that there will be more mobile devices in use than people on the planet with 2.7 billion of these being smartphones capable of sensing and interrogating the environment around us (CCS Insight, 2013). Consumer devices incorporate data collecting technologies as standard signifying that all smartphone users become 'citizen sensors' whether actively or passively. The popularity of low-cost sensing technologies signifies that data can be collected and analysed in real time (Meek et al. 2014) which opens up great possibilities for directed crowdsourcing in time critical investigations or where data is required iteratively. The proliferation of citizens as sensors and the quality measures that are continuously being developed signifies that VGI can be more accurate, more up-to-date and more complete than professional sources (Devillers et al. 2012, Wiggins et al. 2011, Sullivan et al. 2013).

2.8 Volunteered Geographic Information as a Research Tool

Using crowd data in place of information traditionally collected by professional teams may seem alarming to data quality professionals, however, VGI is gaining popularity in situations where the knowledge of local people can be of great use to the application in question and also where authoritative geographic data fails in terms of update frequency and detail.

There are numerous examples of how VGI has benefited society from improving accessibility to GI, improving our phenological monitoring to facilitating improved responses to natural disasters (Roche et al. 2013). Many National Mapping Agencies (NMAs) are comparing their datasets to VGI to enable the detection of potential errors and recent changes (Devillers et al. 2012), showing how VGI has evolved to become a respected and valued source of information. Where timely data is required VGI is increasingly becoming a solution in the world of academia. The origins of citizen science saw volunteers undertaking simple identification tasks e.g. the 1930's Land Utilisation Survey of Britain (Jiang et al. 2013) and the RSPB Bird Count. Through technological developments, citizen science has now evolved to the extent that contributors are now able to solve complex scientific problems which previously

required a great deal of scientific training to undertake. Shamir et al. (2014) describe how crowd classifications have been combined with machine learning techniques to identify whale calls. Citizens have also been challenged to undertake complex hydrology measurements such as stream stage via the CrowdHydrology project (Lowry and Fienen, 2013). Crowds have also been directed to create 3D building models using architectural software (Uden and Zipf, 2013).

VGI presents a “powerful opportunity” (Hampton et al. 2009) to understand current and future environmental changes. If ways are found of mitigating quality issues, VGI enables us to access data that was previously unattainable owing to its prohibitive expense. This is an exciting and increasingly accepted prospect in many domains of academic research. There is however, a wide range of concerns associated with the use of VGI (Budhathoki et al. 2008); the quality control mechanisms employed by professional mapping organisations is seen by some as irreplaceable and thus Goodchild’s (2007) view that VGI will eventually replace authoritative mapping is far from a reality. Using VGI and citizen science data to complement authoritative data and in areas where professional geographic information is unavailable is a more realistic view.

2.9 Uncertainty and Error in VGI

2.9.1 Data Quality Concerns

Geospatial data quality is incredibly important; for inferences made from GI to be trusted and relied upon it is crucial that the GIS community uphold standards of data production depicted in Table 2.1. People have the potential to radically alter representations of place and space by creating geospatial information (GI), enhancing it through annotation and uploading it to the World Wide Web. VGI is very difficult from conventional professionally produced geographic data (Coote and Rackham, 2008); whilst there are many dedicated volunteers producing high quality content, not every ‘produser’ of information has sufficient knowledge and skill to create reliable and usable content (Bruns, 2006 and 2008). There is no control over who records what or how they record it (Comber et al. 2013) which is of particular concern if the data is used for the purposes of scientific analyses without taking this into consideration. The lack of uniformity between humans and how they interpret place and space signifies that crowd submitted content is extremely variable; that there is a lot of semantic heterogeneity in how features and space are described. Academic literature expresses concerns over the validity of “bottom up” data generation with Flanagin and

Metzger (2008) suggesting that it threatens “the veracity of geodata” by providing “poorly organised, out-of-date, incomplete or inaccurate” data.

2.9.2 Technology Limitations

VGI contributors will traditionally collect data (e.g. for OpenStreetMap and geotagged photographs) with a GPS unit, either integrated within a smartphone or using a GPS device. The accuracy of a modern GPS unit with WAAS/EGNOS capabilities varies from 2-3 meters (ESA, 2015). Accuracy can however be reduced to 20 meters when tall buildings or trees obstruct the pathway to the EGNOS satellite or by the occurrence of multi path errors (Blum et al. 2013). The cost of the GPS components will vary between smartphone devices which may lead to variability in the accuracy of crowd submissions.

2.9.3 Sampling Bias

VGI contributors do not necessarily have any regard for how they sample geographic features unless explicit instructions are given on how exactly to do so. VGI sampling is not systematic and the spatial scale over which people contribute is affected by where a contributor lives, what they are interested in and a range of socio-economic factors. Haklay (2010) describes how the coverage of OSM data is variable because of the digital divide; people simply do not want to map council estates. Inaccessibility may also be an issue with certain mappers reliant on public transport and the accessibility of footpaths and certain contributors are only interested in mapping post-boxes (OSM Wiki, 2015). In the case of mobile citizen science applications such as ‘iSpot’, which are used on ‘Bioblitz’ days, it is highly likely that a species will be sampled more than once rendering the species count information unusable in a research context. Appropriate sampling is incredibly important in ecology and in obtaining remote sensing training samples (Rocchini et al. 2011); decisions regarding the focus of conservation projects could be misinformed and any modelling using crowd generated content will have to have a degree of assumed inaccuracy.

2.9.4 Project Futurity

The maintenance and incubation of such projects is an arduous and delicate task to undertake. The success of a volunteered geographic information project depends on the maintainability of quality content over time. VGI is reliant on a small proportion of the populous contributing significant quantities of time and effort in collecting and managing information. As such, it is incredibly important to keep the crowd motivated to keep submitting data (Budhathoki et al. 2008; Coleman and Georgiadou 2009). Crowdsourcing initiatives have recently been popularised within the media and within the domain of GIS; the direct value of humanitarian mapping is very clear from projects like OpenStreetMap which has no doubt inspired many to try their hand at submitting information to such projects. Whether this is seen as a novelty or the beginning of a sustained VGI movement is questionable. VGI is vulnerable from a sustainability perspective and as the UN's Future Trends in Geospatial Management (2013) crucially depicts, whether or not enthusiasm, intrigue and dedication to such projects will linger over extended periods of time is yet to be seen. Regardless, the short term potential for volunteered data is huge.

2.10 Assessing Spatial Data Quality

Much focus has been placed on finding methods to evaluate the quality of VGI (Goodchild, 2007; Coote and Rackham, 2008; Mooney et al. 2010). Spatial data is traditionally assessed in terms of both its internal data quality (taking into account logical consistency, completeness, precision and attribute accuracy) and external quality (accessibility, representational data quality and contextual quality) which if assessed in combination, can give the assessor a strong indication of a dataset's fitness for purpose (Wang and Strong, 1996, Devillers and Jeansoulin 2006).

Assessing both internal and external accuracy for a given dataset is difficult (Brando, 2010) and as such many approaches to VGI accuracy assessment have been used. Many researchers have tested volunteered content against the specifications of the ISO 19157 Geographic Data Quality measures to quantify the inaccuracies of the information (Haklay, 2010). The ISO measures attempt to standardise data quality assessments and provide a framework from which quality can be assessed.

2.10.1 Internal Quality

Internal quality represents the similarity between the produced data and a "nominal ground" or ideal data set. An ideal data set may be argued to be non-existent (Devillers and Jeansoulin, 2006); in which case a data set thought to be of higher accuracy should be used to test the data set for internal quality. There are several criteria to depict the internal quality of spatial data (Table 2.3) as described by the ISO 19113 standards (ISO, 2002):

Table 2.3 Internal Quality Measures

Quality measure	Description
Completeness:	Presence and absence of features, their attributes and relationships
Logical Consistency**:**	Degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical).
Positional Accuracy:	Accuracy of the position of features
Temporal accuracy:	Accuracy of the temporal attributes and temporal relationships of features
Thematic accuracy:	Accuracy of quantitative attributes and the correctness of their non-quantitative attributes and of the classifications of features and their relationships.

2.10.2 External Quality

External quality refers to the correlation of data with that of the user requirement for the data – thus implying that quality is never absolute (Devillers and Jeansoulin, 2006). The requirements of spatial data are likely to vary between various people and organisations. Environmental planners for example are likely to require high precision data whereas attribute detail and accuracy may be a higher priority for a perspective tourist finding out about a particular location. Harding (2011) depicts how from a provider's perspective there is a strong emphasis on "fitness for purpose". It is therefore essential that an analysis of external quality be undertaken with regards to this project. Indeed given that VGI provides free and often high quality data (Parker, 2011), certain users may be willing to relinquish a degree of accuracy and quality control (Coleman, 2009) if data liability is not an issue. Data quality is at its optimum

when the difference between the quality of the data and that of the user's needs is minimal (Devillers, 2006). Wang and Strong (1996) have depicted four aspects of external data quality which can be tested (Table 2.4).

Table 2.4 External Quality Measures

Quality measure:	**Description**
Intrinsic data quality**:**	Believability, accuracy, objectivity, reputation.
Contextual data quality**:**	Value added, relevancy, timelines, completeness, appropriate amount of data.
Representational Data Quality**:**	Interpretability, ease of understanding, representational consistency, concise representation.
Accessibility Data Quality**:**	Accessibility, access, security.

As quality can be both External (in terms of trust) and Internal (e.g. positional accuracy) the development of a methodology for assessment is very difficult (Brando, 2010). Furthermore, as VGI exists in many forms e.g. geotagging, digitisation and GPS upload – multiple approaches to validation are needed. VGI users are often encouraged to "map what is important" to themselves (OSM Wiki, 2011); it is thus unreasonable to view VGI purely in terms of intrinsic data quality as subjectivity is important with respect to quality depending on what the data is to be used for (Flanagin and Metzger, 2008).

2.10.3 Metrics to Measure Quality

2.10.3.1 Assessing Completeness

Assessing data completeness usually involves looking at the presence or absence of one dataset in comparison with another to determine rates of commission (items that are not necessarily meant to be in the dataset) and omission (items missing that should be in the dataset) (Coote and Rackham, 2008). This is particularly difficult with VGI as the data collected (particularly if open web data) was not necessarily created with the same end use in mind, different data may have been collected intentionally. Data completeness is easy to calculate within a GIS; common datasets can be measured and compared in terms of area covered, length of features, proportions of

the same attribute type present and point density (Koukoletsos et al. 2012). VGI datasets are traditionally compared to a dataset of higher quality (Goodchild and Hunter, 1997) such as that of a national mapping agency (Haklay, 2010) however this can be problematic when the area has not been previously mapped (Maué and Schade, 2008).

2.10.3.2 Assessing the Accuracy of Attributes

Attributes of VGI data are usually in nominal format meaning that for a given attribute there is the potential to be a range of possible values. Assessing the accuracy of the attribute is simply done by comparing the given attribute to the actual attribute (on an authoritative dataset) and calculating the resultant level of agreement. Differences between strings often arise because of spelling mistakes, abbreviations, capitalization and use of grammar such as apostrophes. Work can also be done to ascertain the degree of error; the calculation of a Levenstein distance (Girres and Touya, 2010) can be used to ascertain the distance between two attributes and whether or not the difference is more than likely the result of a spelling or syntax error. The distance is calculated by ascertaining the least number of edits required to modify one string to match another. Fuzzy accuracy assessment techniques can also be used where an attribute may not be wholly correct but may not be entirely incorrect e.g. where hierarchical relationships are involved in land cover classification (Woodcock and Gopal, 2003).

2.11 Previous Research on the Quality of Ecological VGI

2.11.1 Citizen Science Projects Sourcing Ecological Data

There are several crowdsourcing projects that involve the creation of geospatial data; the UNEP protected areas project aims to crowdsource information regarding protected areas on the Earth's planet and display it alongside authoritative information from national agencies and international organisations in order to develop Millennium Development Goal indicators of protected area coverage. It thus aims to engage and share information whilst offering tools to citizen scientists to feed their knowledge into real world goals. Another example is the Rapid Assessment of Land Use Change In and Around Protected Areas (RALUCIAPA) site, which allows contributors to view Vegetation Continuous Field (VCF) percent forest cover data with overlain layers of protected areas and base satellite images from Google (Mulligan, 2008). Similarly the

Earth Watchers project (Deforest Action, 2012) allows volunteers to monitor an allocated patch of forest as new satellite imagery is released, developing temporally accurate cover estimation. Crowd contributions in the domain of land cover can generally be grouped into two categories (Fonte et al. 2015): crowd validation and crowd classification.

2.11.1.1 Crowd Validation

The Degrees of Confluence Project (DCP) aims for a photograph to be uploaded of each latitude and longitude intersection on the planet from each cardinal direction. Like Geograph multiple photographs have been uploaded for each point which signifies that the 111631 project images can be used as a reference data bank for multiple time periods. Foody and Boyd (2013) tested a sample of images from 99 DCP points as a validation source for the Globcover land cover map. The results showed a great level of inter-annotator disagreement and a low agreement with the land cover map though this could be due to inaccuracies present within the Globcover data. Iwao et al. (2006) also used DCP data as a validation tool for various land cover maps: GLC2000, MOD12, UMD and GLCC. Interpretations from the DCP images were shown to have the same or greater agreement with the land cover maps than interpretations from satellite imagery showing their use in validation. Estima et al. (2014) and Estima and Painho (2013) have also tested VGI as a validation source; a study using Flickr images to validate the land cover map for a Portuguese municipality found a poor distribution of photographs and discovered that more than 50% of the images did not depict any discernible land cover class suggesting that the source could only be of use if combined with other banks of validation information. This shows that it is very important to select a source of VGI or open web data that is appropriate for the purpose of use.

2.11.1.2 Crowd Classification

A web-based platform “Geo-Wiki” is the only crowdsourcing platform specifically for global land cover in existence which asks volunteers to choose a land cover type for randomised Google Earth imagery tiles. Volunteers are instructed to view areas of disagreement between global land cover classifications with Google Earth’s high resolution imagery (sub-meter to 4-m resolution) underlain. Georeferenced photographs also aid contributors in their task of assessing the accuracy of land cover within a given range. The results are stored and are collated to form a hybrid land cover data set to be utilised in future analyses. Many studies have been conducted

using data collected from GeoWiki: (Perger et al. 2012, Comber et al. 2013, Fritz et al. 2013, See et al. 2013). Comber et al. (2013) compare 7,657 VGI land cover classifications attained through Geo-Wiki to control data and found the accuracy to be 62%. Perger et al. (2012) also determined crowd accuracy to be between 67% and 77%. Despite aiming to provide a high resolution dataset, the randomised nature of the locations presented for classifications denotes that users cannot use their localised knowledge to contribute *in situ* classifications; instead non-localised, old and often poor resolution imagery is used to derive the land cover.

2.11.2 Can VGI Address the Flaws of Ecological Sampling?

Talks with Britain's National Mapping Agency, Ordnance Survey, (G Hart, 5/5/13) have identified ways in which the following types of land cover were identified as particularly problematic features where VGI could assist in enhancing the classification of features owing to their similarities in spectral responses:

1. Turbid and weed-filled water bodies misclassifying as bare ground
2. Confusion between bare ground and manmade surfaces – the diversity of manmade surface types make this really tricky to achieve automatically
3. Dead trees misclassifying as buildings (usually quite rare, becomes more of a problem at the extreme ends of the year for leaves on the trees i.e. late April and early-Oct imagery
4. Possibly splitting grass/crops class into agricultural/recreational land – the Ordnance Survey do not currently attempt this but their capture specification now includes it.

VGI is increasingly being used and seen as an appropriate source of data within ecological and biogeographical research communities as many professional organisations (such as the Ordnance Survey, Land Care Research and the Woodland Trust) are developing their own crowdsourcing projects to help meet their own research and data collection requirements. Peters (2010) depicts the merits of using non-traditional knowledge in ecological sampling and Teacher et al. (2014) also champions the use of new technological developments and provides a positive review of the use of smartphone applications in ecology, particularly with regards to in field image processing capabilities. If orchestrated effectively, citizen sensing projects have the potential to provide samples that meet the requirements set by the principal research investigator. If the observations are of sound quality, replicable results,

useful in the modelling and prediction of ecological spatial-temporal distributions, can be generated (Getzin et al. 2012). The use of crowdsourced data makes sense in areas where there is a great need for the data; the risks of using it (Haklay, 2010) are surely outweighed by the potential to source data that would otherwise not be collected.

2.12 Approaches to Address VGI Quality

The research community has ascertained the extent to which VGI provides an accurate and reputable data source (Haklay, 2010, Girres & Touya, 2010 and Goodchild & Li, 2012) but the best practices for assuring quality with respect to the different types of VGI still need to be defined. Though VGI does evidently have irremovable quality issues, they are not insurmountable. Currently, there is a great emphasis within the academic community in trying to tease out the evident benefits of a willing volunteer cartographic workforce and ensure that submissions can be used in conjunction with authoritative information. Some approaches include geometric corrections to improve the positional accuracy of the data (Fairbairn and Al-Bakri 2013) whereas others concentrate on aligning the thematic nature of the various datasets. Semantic integration has been trialled as a method of conflation addressing the lack of consistency between how the untrained volunteer interprets something and the official terminology. Finding the linkages between these data permits a reduction in heterogeneity and the establishment of common ground between otherwise identical and real phenomena. Ramos et al. (2013) use domain ontologies to the map classes and their properties from different sources onto one another and Du et al. (2013) utilise semi-automatic methods to align the geospatial vocabulary with that of the crowd. Dynamic surveying approaches have also been trailed; iteratively querying the quality of data as information is submitted allowing for the real time validation of information (Pawlowicz, 2011).

Quality can also be assured via community monitoring. Bug checking software (such as Open Street Bugs and Map Dust) can highlight areas where errors are prevalent so experienced users can edit them to an appropriate standard. In this way the community is able to monitor new users and unusual submissions to ensure that the data is of high quality. Community checking is however more relevant to larger, more established projects such as OSM where there is a close community of users that take responsibility and pride in the condition of the map. Bug and user monitoring can also be used in small scale projects; Comber et al (2013) used Geowiki data to calculate the degree of accuracy users had in identifying certain classes at 299 control

points. The results of the user control point analysis were then utilised to weight the reliability of pixels associated with volunteer generated submissions according to four measures, providing an overview of data reliability. Whilst this does not enhance the quality of the data itself, it does permit users of the data to make informed choices about whether and how it should be used.

A framework focusing on 'seven pillars' of data quality validation and assessment is being developed (Meek et al, 2014) investigating the various elements of crowdsourced information that are of cause for concern. The nature of the workflow, which requires iterative checks, exemplifies the complex nature of crowdsourced data and in verifying its quality.

2.12.1 Fusing VGI and Remote Sensing

VGI has been used frequently for the validation of thematic map outputs (such as Geo-Wiki and the Degrees of Confluence Project (Foody and Boyd, 2013) but there is a need to focus on how VGI can be best combined with machine learning and remote sensing techniques to provide training data for the generation of thematic maps. Schnebele and Cervone (2013) use machine learning techniques to generate an updated flood hazard map, but VGI has not been utilised as training data for land cover classification. VGI could certainly alleviate the shortfalls of existing ground truth data by providing up to date samples which could be useful in detecting environmental change in areas where there is no research funding to provide an official survey. Developing methods of obtaining the best quality training data possible (perhaps through weighting the data), in the desired areas is required to ensure that VGI can be used to train remotely sensed imagery effectively.

2.12.2 Variance in Structure & Codes of Practice

Directed crowdsourcing, whereby users are grouped and have specific goals and objectives, could prove useful in enhancing the overall quality of volunteer created information. If a system receiving crowdsourced input can recognise and interpret the confidence of the user in labelling geographic content and their degree of success in doing so, it can then direct users to tasks where they are likely to perform better, improving the overall accuracy of submissions. Control data such as that used by Comber et al. (2013) would be subject to a number of assessments including portmanteau accuracy (Comber et al. 2012), user and producer's accuracy alongside

measures of user confidence (Tuia and Munoz-Marí, 2012). The outputs of this could then be used to direct users to areas that would be more appropriate for their classification skills.

The design of a task is very important to obtaining a high quality data set (Nichols & Williams, 2006). Certain recordings (such as contributor locations) can be automatically sourced from the devices gyroscope and GPS unit (Meek et al. 2012) so as to reduce the error through human input. The Crowdcrafting.org platform allows researchers to dictate the number of tasks available to a contributor and the type of task they are permitted to perform. This enables certain tasks to be delivered to certain users if they are deemed to be good at a particular task. It is important to evaluate the design of a crowdsourcing project to best understand how certain elements of the project design and structure could be influencing the quality obtained from contributors.

2.12.3 User characteristic filtering

Contributions by 'spammer' users (Foody, 2013) can often do more damage than good with regards to the overall accuracy of crowd classifications. Establishing the inter-annotator accuracy and filtering the results accordingly could improve the classification by eliminating the classifiers that do not agree with the majority of the crowd (Rodrigues et al. 2013). Dickinson et al. (2010) even suggest the exclusion of new user contributions and from those who continually submit erroneous data. An analysis of intra-annotator accuracy over time (Foody, 2013) shows that high frequency contributors are not necessarily likely to provide submissions of greater accuracy. It is therefore sensible to encourage new users, which ensure the sustainability of the project and could offer new insight on classification tasks. Knowing the likelihood that a sample is accurate is very useful; using only the annotations of the most skilled volunteers has been shown to improve consensus labelling accuracy by Kumar & Lease's (2011) probabilistic multi labelling approaches. The suggested exclusion of new participants should be seen as a last resort; it is important to build a sense of community in crowdsourcing projects and a disregard for new contributors can deter participation.

Annotator metadata is often a required component for filtering submissions; information about a contributor's motivation, skill level and interests can help determine whether or not a contributor is likely to be good at a classification task. Kazai et al. (2011) have shown how certain personality traits can show correlations

with the level of accuracy obtained from crowd labels. Personality and demographic traits (Li et al. 2014, Ipeirotis, 2010 & Ross et al. 2010) are also shown to be effective methods of predicting annotator accuracy. Anonymysed submissions procured from open web present a problem in this regard; the data are hard to filter as little is known about the contributors; label accuracy must be determined in other ways.

2.12.4 Ancillary information

Goodchild and Li (2012) suggest looking at geographical rules as a form of quality assurance. Logic can be used to perform quality checks which can indicate the accuracy of a dataset. For example, in the case of vegetation mapping it would be prudent to compare the location of the recorded instance of vegetation to known species distribution models. If uncertain, the species' physical tolerances (such as preferred climate) can also be used to indicate the likelihood of an instance of the species occurring at a given location. The buffer zone of the species distribution model to which the geo-tagged contribution falls in can be used to calculate the statistical likelihood that the submission is accurate. The lack of open, interoperable and complete databases of such species distribution models could however inhibit this method with the crowd data simply indicating the presence of landscape change.

2.13 Key Research Gaps

This review of literature pertinent to the theme of ecological VGI data quality has presented several interesting research gaps that merit investigation.

- Despite evidence from the literature that crowd workers can be used for complex classification tasks (Shamir et al. 2014), the crowd are rarely used to classify ecological data. Most of the current research focuses on using the crowd to verify existing data rather than address the key need of collecting fresh biodiversity data. This is addressed throughout the thesis.

- Directed crowdsourcing is recognised as a technique but few research initiatives demonstrate its potential for creating temporally accurate land cover maps. There is a need to assess the fitness for use of both open web data and directed crowdsourcing data (Fonte et al. 2015). This is addressed throughout the thesis.

- There is also a requirement to rapidly assess the quality of VGI so that it can be used with minimal alteration. Determining ways of quickly filtering the data so that it can be used with confidence is a research priority. User characteristic grouping and confidence rating appear to be fruitful options investigated in machine learning studies that are yet to be explored in the domain of VGI. These will be addressed in Chapter 5 of the thesis.

Chapter 3 | Assessing the Potential for Open Web Data in Enhancing the Quality of Authoritative Land Cover Mapping

	Research Objectives Addressed In This Chapter:
1	Establish the degree of accuracy required from ecological datasets and any concerns that the potential use of crowd generated content presents to users of ecological datasets.
2	Obtain open web data and establish whether it could provide any contribution over and above that of the authoritative data collection and classification systems.
3	Understand how the crowd identifies ecological features field in situ and develop the best practises for developing a web based crowdsourcing platform.
4	Develop a system for harvesting crowd contributions in a directed manner in the context of Phase 1 Habitat Mapping and quality assessment of crowd contributions.

3.1 Introduction

3.1.1 Information Mined from Existing Projects as a Data Source

The EU's most frequent assessment of land cover is the LUCAS survey (an annual area frame sampling survey whereby photographs are taken along transects) which is very sparse in coverage. Whilst consistent with the boundaries of OS MasterMap, the UK's flagship land cover product, the Centre for Ecology and Hydrology's Land Cover Map (LCM2007) was last produced in 2007 making it inappropriate for studies dependent upon up to date and thematically accurate data. VGI could certainly be considered as a tool in alleviating these deficiencies.

Application Programming Interfaces (APIs) signify that geographical objects can be mapped using web mapping services (such as Google Earth & ESRI's Basemaps) with high resolution satellite imagery underlain. These mapping services are readily distributable and can easily be made available to the public over the web for collaborative mapping projects. The VGI submitted to these sources can be exploited for alternative purposes because of their open attribution licenses which permit the inputs to be used as 'open web data'.

An example of such a project is “Geo-Wiki” which is specifically designed to collate global land cover by asking volunteers to choose a land cover type for randomised Google Earth imagery tiles (Fritz et al. 2012, Comber et al. 2013). Open web data can also be sourced from projects that are not necessarily designed to collect geospatial information but do extract coordinates from the user’s device. Image files often contain Exchangeable Image File Format (EXIF) information which is used to map photographs on sites such as Flickr, Instagram and Geograph. Tweets from social networking site Twitter contain associated geo-location data which is publicly available and can be used to look at what users are saying in association with the location from which a tweet was sent. There are however drawbacks in mining this data such as the fact that tweets may not have geo enabled and that those that do may often contain nil in the way of useful information. The success of mining (the process of extracting) geoweb data for research use is reliant on effective filtering methods.

3.1.2 Previous Research Using Existing Open Web Data

A study comparing open web data to existing Global Monitoring for Security and Environment Urban Atlas (GEMUSA) land use data determined it is possible to extract incomplete Urban Land Use information from OpenStreetMap at acceptable levels of accuracy (Arsanjani et al. 2013). Photos and their associated descriptions have also been extracted from the Degrees of Confluence Project for use to validate existing Land Cover Maps (Foody & Boyd, 2013) with subsequent non-expert interpretation of forest cover within 1.38–9.08% of that of the authoritative assessment. Estima et al. (2014) have also concluded that the use of geotagged Flickr photographs in developing land cover maps may be useful when used as an ancillary source of information in conjunction with other data sources.

Established in 2004, OpenStreetMap is now a mature VGI project and highly regarded to the extent that it is considered as a potential alternative authoritative data. Over 2.2 million users (as of August 2015) have generated a map that in parts is more complete, spatially accurate and semantically harmonised than comparable proprietary datasets (Arsanjani et al. 2015). Having determined an indication of the data quality, the research focus is increasingly shifting towards the extraction of value-added information from the OSM database and understanding how it can be used to fill any shortfalls in professional datasets e.g. deriving population estimates (Bakillah et al. 2014) and the estimation of building types (Fan, 2014).

Restricted to the United Kingdom, Geograph is a comparatively small scale project but does contain a wealth of photographic geolocated data source with associated descriptions and tags. Geograph has been analysed in terms of participation patterns (Antoniou and Schlieder, 2014) and the photographic database has been utilised as a means for testing machine vision classification techniques (Torres and Qiu, 2014) however, it is thought that this study presents the first use of Geograph's crowd interpretations as an indirect source of land use/cover data. Geograph also offers an interesting test case owing to the spatially explicit nature of data collection and the element of gamification that being the first to associate data to a previously unvisited grid square provides.

Prior to the completion of this work, no studies on deriving land cover patterns from open web data had been reported. Iwao et al. (2006) and Foody and Boyd (2013) look at the Degrees of Confluence Project as a source of land cover data but these sources are not labelled by users and require volunteers or professionals to classify them prior to comparison with authoritative sets. Subsequent research (Estima & Painho 2013, Estima et al. 2014) has investigated the viability of using similar ancillary sources, however, the techniques used within this chapter aim to present a unique analysis of two sources at multiple hierarchical levels. String comparison is used to determine off the shelf viability as well as a look at the producers of the data to provide further insight into the possibility of quality improvement if filtering methods were to be used.

3.2 Aims & Objectives

The research presented aims to explore the adequacy of using passively generated geospatial content in conjunction with authoritative data to produce a fit for purpose and deployable land cover product. To achieve this, various indirect sources of land cover data are trialled and evaluated to establish the emerging terminology used to describe land cover data and its utility in the context of authoritative mapping. Conflation with existing Ordnance Survey Master Map data is trialled in order to discover any virtues and weaknesses of crowd generated content. The research questions addressed within this study are presented below:

1. Can indirect sources of land cover information be used to provide an accurate land cover class?

2. Do open web data producers describe features at an appropriate scale?

3. Can core producers of accurate open web data be identified?

4. Is it viable to use indirectly sourced crowd generated content to enhance existing authoritative data sets and what limitations must we be aware of when working with such data?

3.2.1 Chapter Outline

In this chapter indirect land cover data sources are evaluated to determine whether crowd contributions can be used to provide a land cover class and to determine whether the class identified by the crowd is correct. This evaluation was made using Geograph and OpenStreetMap tag information which were extracted from the whole of Hampshire and then assigned to points and acted as the sample data for comparison with Ordnance Survey MasterMap and LCM2007. An overview of the utilised data sets is presented in 3.3. The applied methodology is presented in section 3.4, followed by the presentation and discussion of the obtained results (3.6). Conclusions are drawn which are used to inform and refine the methodology for Chapters 4 and 5.

3.3 Key Materials and Data Preparation

3.3.1 OpenStreetMap Data

OSM data is either collected via GPS recording or via satellite imagery digitisation and is freely available under the Creative Commons 2.0 License. The OSM project organizes contributions into either points, lines or polygons each of which is given a tag consisting of a key and a value. The key and value are free text fields, however there are complex tagging conventions which users are encouraged to use for common features. The ability for users to generate new tags denotes that the contributed features often lack logical consistency and thematic accuracy (Hagenauer and Helbich, 2012). OSM contributions also frequently overlap each other signifying that one given area may be given multiple tags (Arsanjani et al. 2013). This provides uncertainty in the tag making off the shelf usage of this data very difficult.

Land cover information is readily available in OSM but not necessarily under the appropriate “land cover” tag. The data is spread across many tags including natural and surface. Some tags under the “landuse” group are in reality land cover data, for example land use = grass. Land cover can often be inferred on OSM by looking at the

other tags present e.g. a park tag is represented under land use may but it is possible to infer that it is covered by grass.

Inconsistencies in the uptake and deployment of the OpenStreetMap ontology denote that the data is not incorporated under one land cover tag directory. The “Natural” tag is predominately used to depict areas of land cover (Figure 3.1) however, as can be seen below (Figure 3.2) other tag directories such as “Surface” also contain a great deal of information that could be overlooked if the user assumes all the land cover information to be in the appropriate place. Furthermore, there are also many terms that one would assume to find within the land cover category within the landuse category. “Grass”, “pond” and “forest” are just a few of the tags that have significant quantities of land area within the land use class. Whilst if merged the tags found within OSM would probably amount to a fairly comprehensive ontology of land cover they are not currently in such a format which promotes the appropriate classification of land cover.

Figure 3.1: Map of OSM Contributions for the Hampshire Area with “Natural” Tag

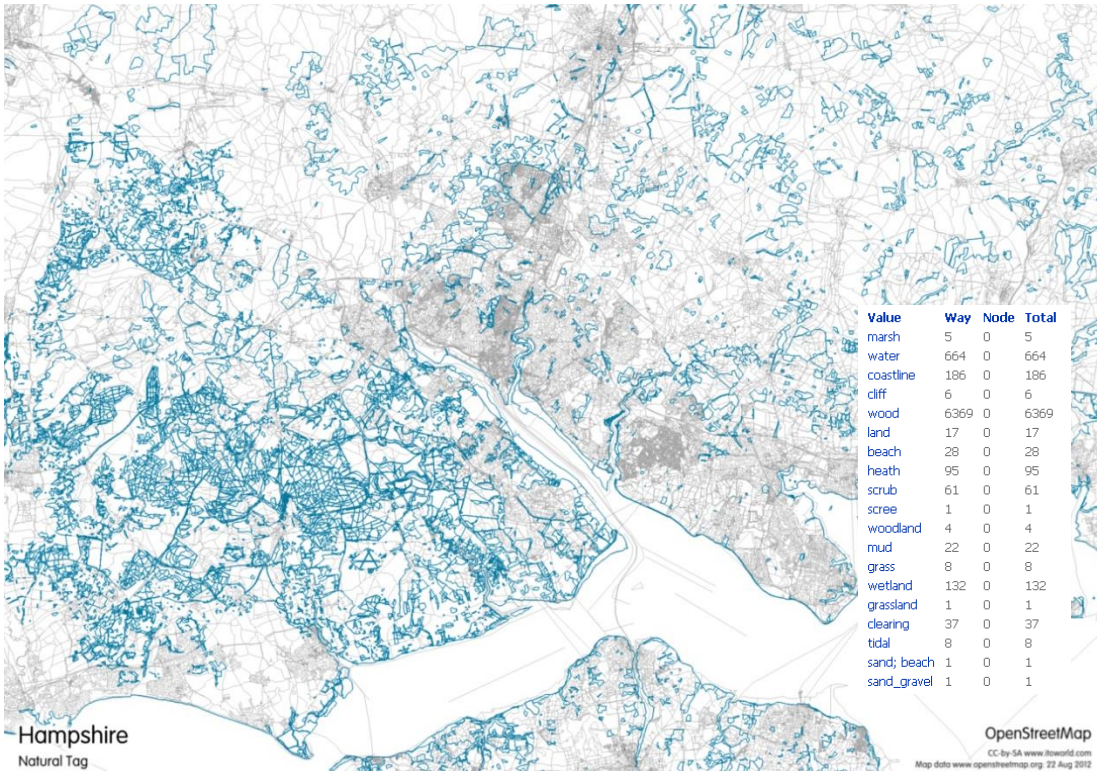
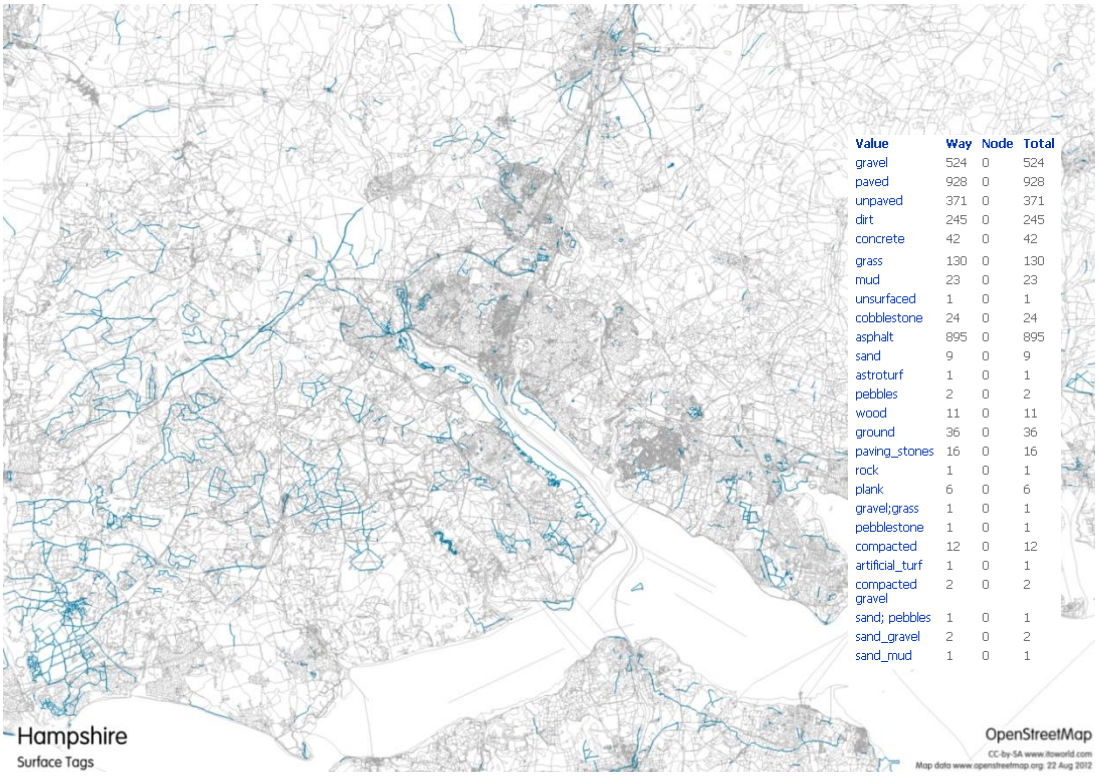


Figure 3.2: Map of OSM Contributions for the Hampshire Area with "Surface" Tag



Value	Way	Node	Total
gravel	524	0	524
paved	928	0	928
unpaved	371	0	371
dirt	245	0	245
concrete	42	0	42
grass	130	0	130
mud	23	0	23
unsurfaced	1	0	1
cobblestone	24	0	24
asphalt	895	0	895
sand	9	0	9
astroturf	1	0	1
pebbles	2	0	2
wood	11	0	11
ground	36	0	36
paving_stones	16	0	16
rock	1	0	1
plank	6	0	6
gravel;grass	1	0	1
pebblestone	1	0	1
compacted	12	0	12
artificial_turf	1	0	1
compacted gravel	2	0	2
sand; pebbles	1	0	1
sand_gravel	2	0	2
sand_mud	1	0	1

For the "landuse" attribute there are 29 core values (forest, farmyard, industrial, grass, etc.). OSM editor software encourages the use of these through the provision of drop-down-list selection functionality. However, contributors are able to type different values in as free-text, supplying their own value for an attribute. This flexibility in the tagging structure and the lack of senior mapper checking the ontological alignment with existing tags, leads to many OSM tags being inappropriate.

OSM data can be downloaded from Geofabrik.de in SHP format; however, these files do not contain "multipolygon" features owing to the limited processing capabilities of Geofabrik. To attain the full set of features the OSM data was downloaded from OSM planet in XML format after clipping it to the extent of Hampshire. After inspecting the tags present within OSM it was deemed that tags that fall under the categories of "landuse", "natural", "surface", "waterway" and "landcover" should be extracted from

the XML data to avoid many redundant data columns which would significantly slow the subsequent XML to SHP conversion. The XML was clipped to the study area using Osmosis (a command line Java application for processing OSM data) before being extracted into a GML document containing the specified tags using a Python conversion script. The data (which represents a snapshot of the state of OSM's land cover data in July 2012) was then converted into SHP format using the FME gml2shape script prior to further analysis.

3.3.2 Geograph Data

Geograph is a crowdsourcing project aiming to create a resource storing "geographically representative" photographs taken at each grid square in the UK. It is not directly associated with land cover but the user descriptions and tags associated with the uploaded photographs do sometimes contain land cover terminology or comments that can infer a particular type of land cover. Each photograph is stored within a database alongside the associated metadata such as the date on which the image was taken, subject location and photographer location.

An email was sent to the administrators of the Geograph web site in July 2012 requesting the attributes required for the project of each photograph within a bounding box of coordinates. The data was received in tsv format and converted to dbf format so that it could be inputted and exported as an ESRI shapefile. Over 129,000 Geograph points are included within the file which provides a strong coverage of the study area.

3.3.3 CEH Land Cover Map 2007

Land Cover Map 2007 (LCM2007), released in July 2011, is the latest comprehensive land cover map of the UK. It is a parcelled, vector dataset derived from satellite images collected between 2005 and 2008 and Ordnance Survey MasterMap parcels. Ancillary information such as the soil type and terrain data have also been used to further refine the classifications e.g. grass land at a low altitude in a tidal flooding plain is likely to be saltmarsh (Morton et al. 2011). The land cover attributed to each parcel is based on 23 UK Biodiversity Action Plan Broad Habitats which are widely used to report and monitor UK countryside. It has a minimum mapping unit of 0.5ha and a minimum feature width (MFW) of 20m. The data has been spatially generalised using splitting and merging techniques as the detail of OSMM is too great to use with the 20x20m pixel data used for production. The overall accuracy of the dataset is 83%

though some classes are detailed as having greater spectral variability than others and as such are less certain.

The Centre for Ecology and Hydrology's Land Cover Map 2007 (LCM) data was obtained from the Ordnance Survey Research department and used under their license which covered the internship work. The LCM data was then clipped to the extent of Hampshire (Figure 3.3).

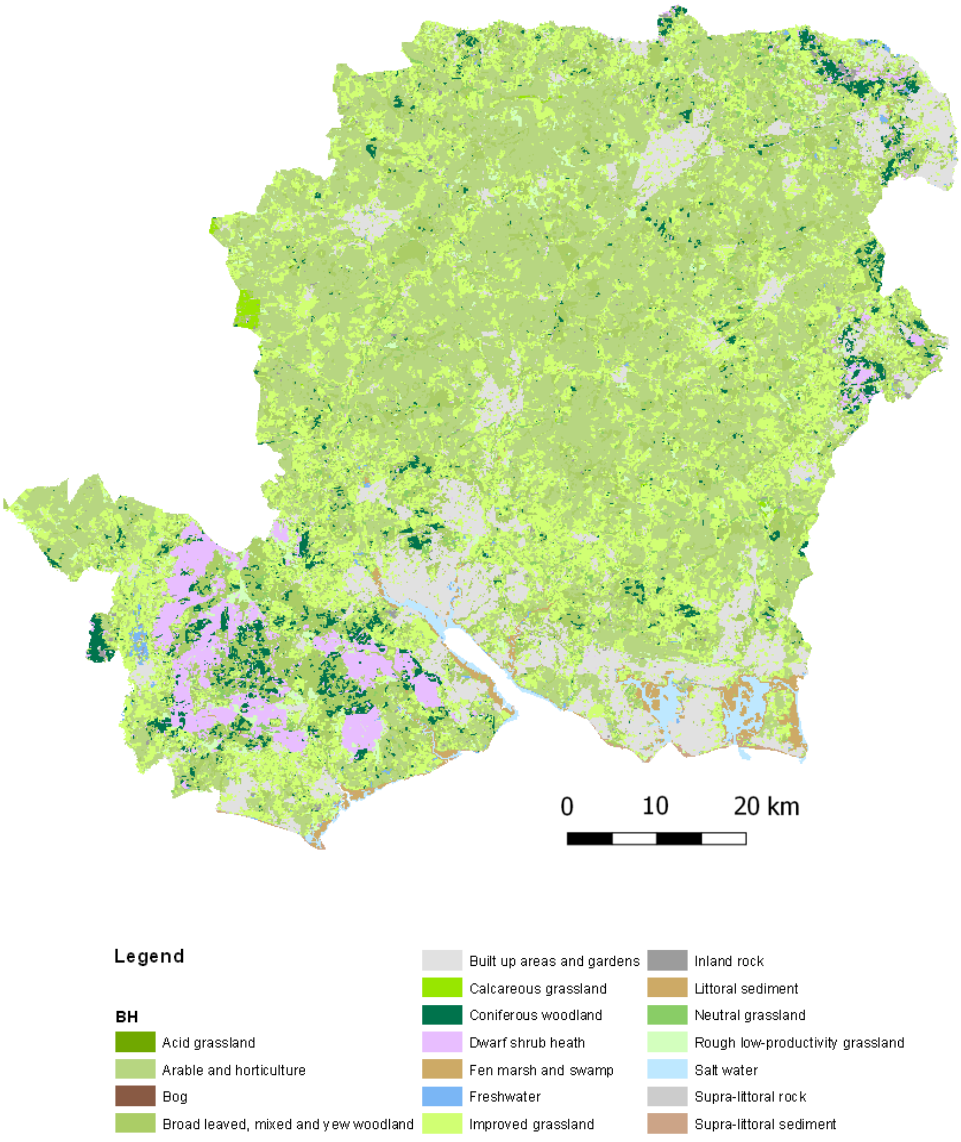


Figure 3.3: Map of CEH Land Cover 2007 Data for Hampshire

3.3.4 Ordnance Survey MasterMap

OS MasterMap is British Mapping Agency, Ordnance Survey's flagship vector mapping product with 450 million geographic features found in the real world. The map is updated every 6 weeks and contains a topography layer which maps key features (buildings, heritage and antiquities, land, rail, structures, terrain, and water etc.). The data has an absolute accuracy (how closely point coordinates with the coordinates of the same point on the ground) of 0.9m (Ordnance Survey, 2015).

Currently OS MasterMap offers little in the way of land cover data as much of it derives from cartographic inheritance. A two tier classification is present in the form of nine core descriptive groups exist which denote the land cover further specified by the "descriptive term" attribute. This attribute, if present, gives further classification information about the feature such as a descriptive term specifying the natural land cover type present in the area. A feature may have multiple descriptive term attributes, but most features have zero or one descriptive term attributes.

The OS MasterMap data for Hampshire was obtained from Ordnance Survey during an internship in approximately 250 GML chunks. The GML was transformed into ESRI shapefile format over a number of days using the conversion facilities within FME (a piece of commercial data integration software). All data was then transformed to the appropriate projection enabling the comparison of attributes.

3.3.5 County boundary

The Hampshire, Southampton Unitary Authority and Portsmouth Unitary Authority boundaries were obtained through the Open Data portal (http://data.gov.uk/) as ESRI shapefiles.

3.4 Study Location

Hampshire was chosen as the area of study owing to the availability of authoritative land cover data (supplied by Ordnance Survey) and the accessibility for ground truthing sites. The cover of the study area is a mix of rural and urban, representing a plethora of land cover types for testing such as water bodies, agricultural areas, heathland and artificial surfaces.

It would have made for a more comprehensive study if the whole of the United Kingdom was analysed. This would have permitted an investigation into the spatial

heterogeneity of the coverage and terminology of crowd generated land cover content however was unfeasible given the time constraints and processing power required to extract the data and perform spatial analysis operations on the shapefiles. Given that a subsequent investigation used an area of 32 km² areas for a similar study (Arsanjani, 2013) 4000 Km² county scale investigation is certainly significantly large enough from which to draw sound conclusions. A map of the study area is provided in Figure 3.4.

Figure 3.4 Map showing Study County in Context of Great Britain



3.5 Methods

Two open web data sources (OSM and Geograph) were chosen for analysis owing to the significant amount of natural feature data they contained and the sheer number of geolocated data points (Geograph: 129,253, OpenStreetMap: 11,783). Other open web sources such as Flickr (Estima, 2013) and Twitter may contain more geolocated data points (Estima, 2013) however, these are primarily used for the purpose of sharing experiences rather than a methodological representation of the Earth's surface and would require a great deal of sifting prior to analysis (Frias-Martinez and Frias-Martinez, 2014). Other projects such as the Degrees of Confluence Project (Foody and Boyd, 2013) and others listed within a review of VGI for the validation of

land cover maps (Fonte et al. 2015) could have been used but lack the appropriate coverage for a study at this scale.

The attributes of the crowd generated content were assigned to the relevant OSMM & Land Cover Map parcel via a spatial join orchestrated in FME owing to the size of the data. The following analyses were then performed:

- Point density calculations to assess the extent of heterogeneity in coverage and the density of points within an appropriate time frame for use in image classification. See 3.6.1-3.6.2.

- Extraction of temporally appropriate data points and calculation of grid squares with sufficient temporal coverage using a 1km² fishnet grid. See 3.6.3.

- Python word matching scripts and spatial joining to filter the data for land cover information and to determine the extent of spatial and attribute agreement between the attributes (e.g. title, description, tags, class and context) of the crowd generated content and existing data. See 3.6.4.

- Scale of attribution analysis by attributing each term with a level of detail code. This enabled an analysis of the scale at which contributors describe the land's surface. See section 3.6.5.

- The users with the most descriptive matches with LCM2007 data were extracted and analysed – see 3.6.6.

3.6 Results and Interpretations

3.6.1 OSM Polygon Analysis

As described previously, the OpenStreetMap XML data has been downloaded, land cover data extracted and converted to SHP format. The data has been clipped to the study county area (Hampshire). The Ordnance Survey MasterMap data for the same area has been downloaded for coverage comparison given that OSMM represents a

gold standard dataset (Haklay, 2010). Figures 3.5 and 3.6 show two sites within Hampshire used for an initial visual comparison.

3.6.1.1 Spatial Distribution of OpenStreetMap Data

Figure 3.5 Visual Comparison of OS and OSM Data

OSM Data OS Data

Figure 3.6 Visual Comparison of OS and OSM Data

OSM Data OS Data

An initial look at OpenStreetMap data shows the coverage of land cover data to be very heterogeneous. Some areas such as the one shown in Figure 3.5 are relatively complete in terms of extent with there being few major differences with the Topography layer. Other areas however can be very sparse in terms of coverage; Figure 3.6 shows an area with hardly any OSM land cover data in an area where there should be many agricultural fields according to OSMM Topography layer. The visual analysis of OSM indicates that the benefits of OSM towards land cover extent could be very minimal particularly as even if the attribute data is indeed useful, only the comparatively few parcels of land contained within OSM would benefit from enhancement.

There is also much spatial heterogeneity within the OpenStreetMap land cover data coverage across Hampshire; as can be seen from Figure 3.7 the New Forest is particularly well mapped along with the North East tip of Hampshire. The remainder

of the OSM data is particularly poor in terms of spatial coverage with only 9.83% of the extent of Hampshire covered by land cover information (Figure 3.7). This heterogeneity could be due to certain users being more prolific with regards to mapping the areas immediately around them or could be the result of the type of land cover that people like to map or find easier to determine the boundary of.

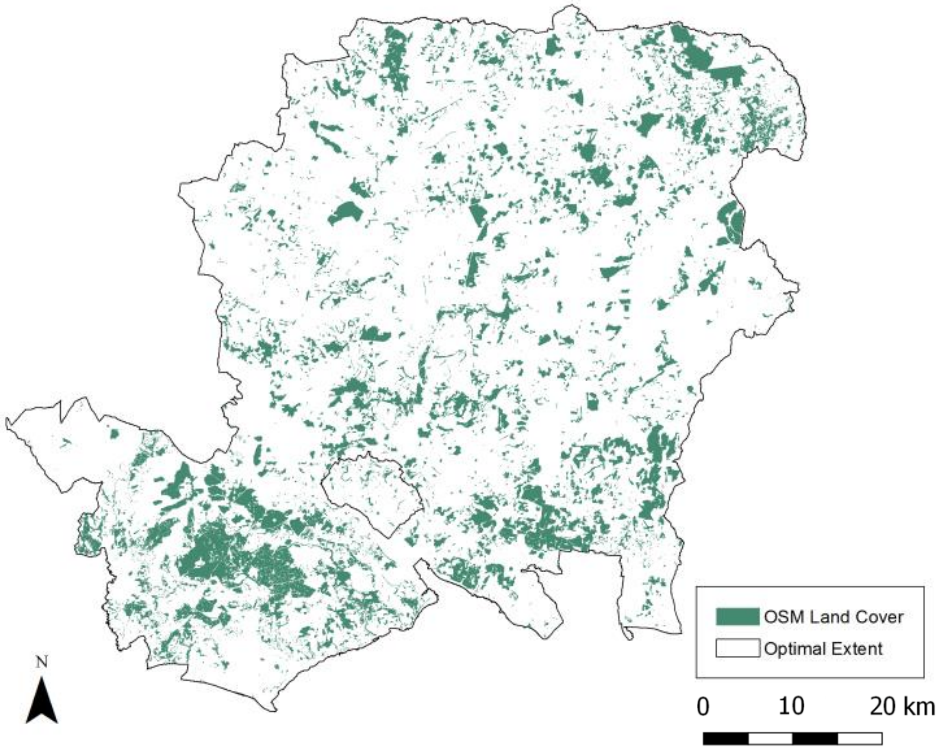


Figure 3.7 Map Showing the Extent of OpenStreetMap's Land Cover Data

3.6.1.2 Equivalent Parcel Extent Analysis

In order to determine the similarity of OSM to OSMM in terms of extent it is necessary to spatially join the data to compare equivalent land parcels. As the target features (OSM polygons) were not completely within the source feature (OSMM), it was necessary to calculate the centroid of each OSM polygon using the field calculator. These points were then plotted and exported whilst retaining the original polygon's attributes. Whilst the centre points of OSM polygons may not necessarily coincide with the same feature within OSMM owing to inaccuracies within the OSM data (thus making attribute data incomparable), it is interesting to see whether the type of land cover attributed to the polygons at the same location is appropriate. A section of the resultant join is shown within Figure 3.8.



Figure 3.8 Spatial Join of OSM Centroid Points with OSMM Topo Layer The data associated with each centroid point is joined to the OSMM Topo attribute data at the relevant location.

The OSMM and OSM extents were compared (Figure 3.9) and the OSMM data was found to have larger polygons with OSM's land cover data spanning 59.54% of the OSMM reference data. This shows that OSM's land cover data is evidently far from complete and that users are interpreting the extent of land parcels very differently from how a professional cartographer would. OSM represents a bottom-up data model (Haklay and Ellul, 2010), the lack of systematic data collection and the freedom of the crowd to map "what is important" to them (OSM wiki, 2011) denotes that naturally, some areas will be far from completion. However whilst in most cases the area of land cover within OSM is lower than that of the OS Topo counterpart at a given location, for the 11783 polygons present within OSM containing land cover information 48.88% of the OSM polyons have a greater area than that of their OSMM counterparts. This could signify the following:

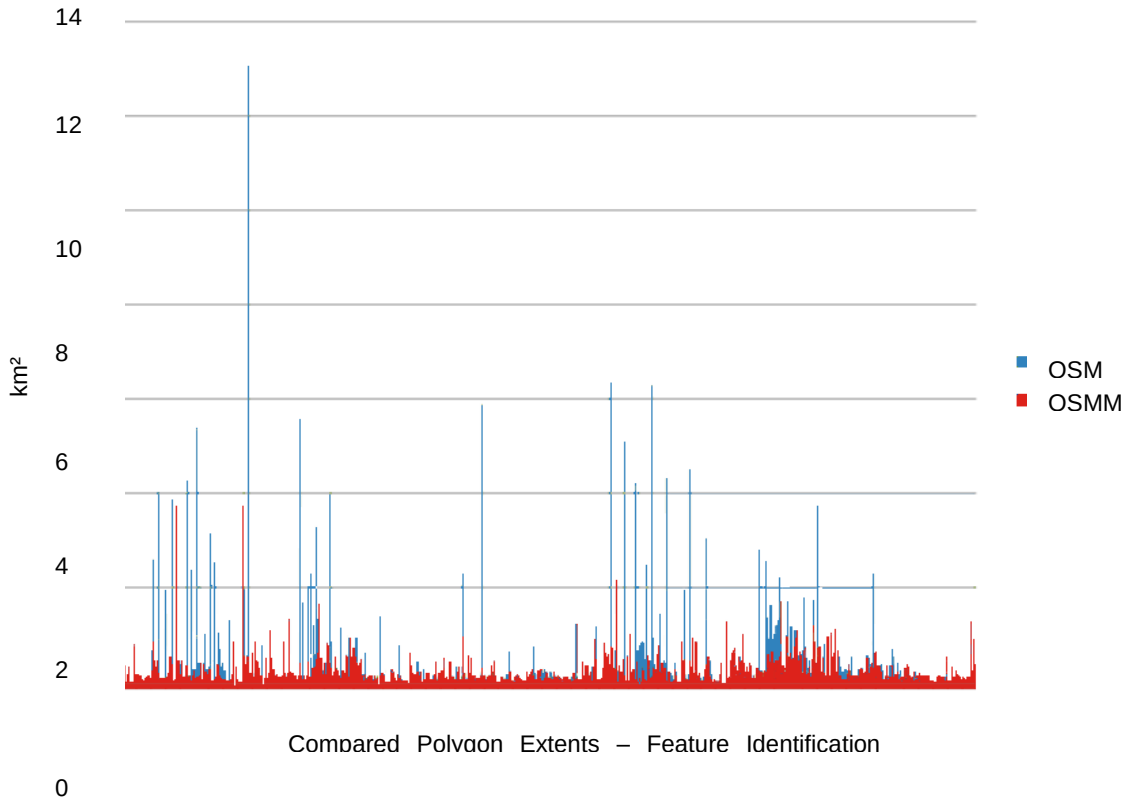
> The centroids of the OSM polygons are not joined to the appropriate feature and thus OSM has poor positional accuracy.

> OSMM has more segmentation within areas of the same land cover signifying the OSM features are joined to the correct feature spatially but exist in a more generalised format.

> OSM users conceive that the extent of various segments of land cover are greater than OSMM would suggest.

Variance owing to incomplete OSMM data is unlikely due to OS's stringent data collection methodologies (Harding, 2011). The other possibilities could however provide the answer and require further analysis of the data. The nature of the features analysed could also account for completeness issues. Mapping vegetation and waterways is very difficult as there are rarely sharp transitions (Peters et al. 2006) from one area to the next with ecotonal areas clouding the perceptions of mapping boundaries. It is thus possible for disagreement over what one OSM user conceptualises a forest boundary (Fisher et al. 2006) despite guidelines being distributed via the wiki. The point at which a patch of grass becomes an area of parkland or a pond a lake is highly contested within official organisations, signifying that amateur mappers are unlikely to be able make similar land classification judgments (Mooney et al. 2010). Evidence of this was clear in the visual analysis; OSM was missing many of the smaller polygonal features which could denote that users may experience difficulty determining the scale at which objects should be defined and classified.

Figure 3.9: Extent comparisons between OSM and OSMM



3.6.1.3 Comparison of extent accuracy by Land Cover Type

The polygon area of each land cover type was calculated for where both OSMM & OSM had an attributed land cover type in order to establish whether the interpretation of the extent of a land cover feature in OSM varied between feaure type. Table 3.1 & Figure 3.10 below show the area of OSM and OSMM by type and the percentage difference between them. Also shown is the percentage of OSM polygons that have a greater area than OSMM and the percentage of polygons that have a smaller area than OSMM for each polygon type.

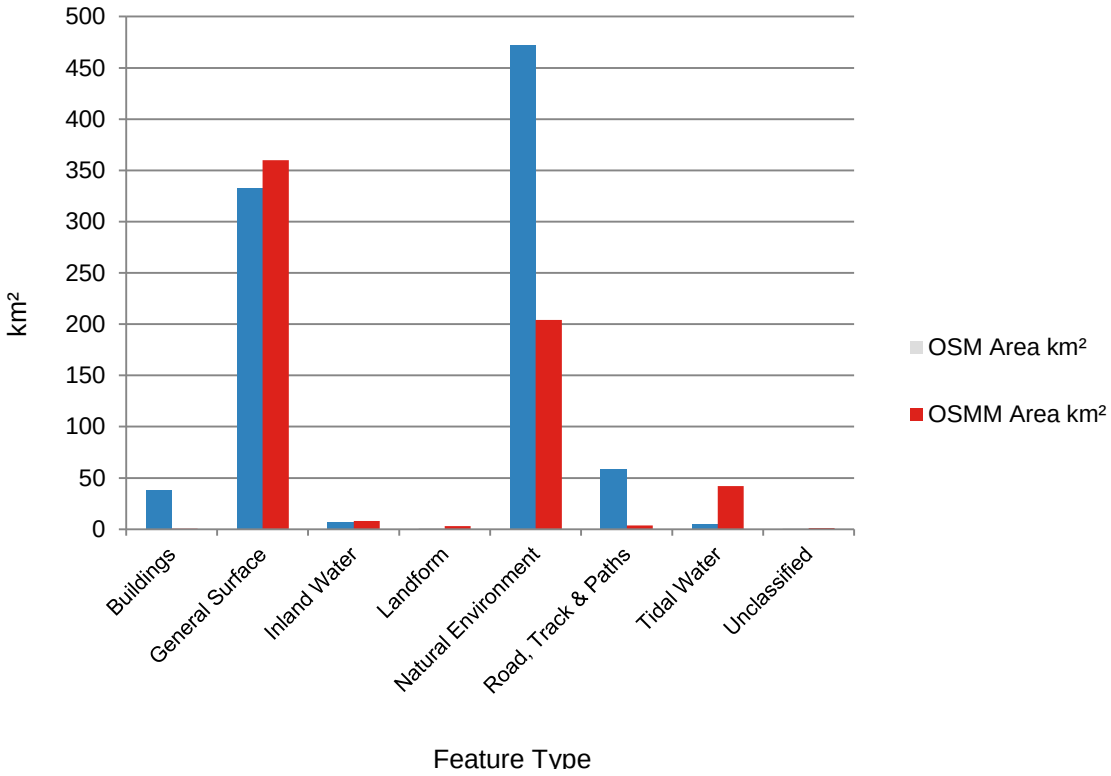
Table 3.1 Variance in polygon extents by OSMM featute type

Type	OSM Area km^2	OSMM Area km^2	Percentage Difference	OSM polygons > OSMM (%)	OSM Polygons < OSMM (%)
Buildings	38.42	0.5	97.45	93.77	6.23
General Surface	332.28	360.05	-4.01	40.92	59.08
Inland Water	6.68	8.12	-9.74	46.27	53.73
Landform	1.1	3.17	-48.53	42.00	58.00
Natural Environment	471.80	204.05	39.62	49.96	50.04
Road, Track & Paths	58.54	3.53	88.62	78.55	21.45
Tidal Water	5.27	41.95	-77.67	20.69	79.31
Unclassified	0.8	1.19	-19.48	35.00	65.00

The Buildings and the Road categories have a significant number of polygons which exceed the area of their associated OSMM polygons with 93.7% and 78.5% of polygons respectively exceeding the area of their OSMM counterpart. This is largely explained by the variance in data collection; OSM has separate tagging ontologies for buildings and roads whereas OSMM includes the land cover data for them as part of Topo. Thus, each building and road within OSMM would have been attributed to a polygon covering the area around the building and thus being associated with a much larger area. Overall OSM and OSMM are relatively similar in terms of the extents covered by “General Surface” and “Inland Water” (with only a 4.01 and 9.74 percent respective difference between the areas of the two types. What is certain is that there are major differences between the data sets in terms of the extent of different land

cover; a Pearson correlation coefficient performed indicates a significance value of Pearson: 0.0132 which indicates that there is no statistical relationship between the areas of equivalent natural area polygons between the datasets.

Figure 3.10 Comparison of Equivalent Polygon Extents by OSMM Featute Type



3.6.2 Spatial Distribution of Geograph Data:

The distribution of VGI submissions traditionally suffers from spatial disparity; people are more likely to take photographs of and describe areas that are attractive to or are of importance to them. Given that the objective of Geograph is to photograph every single UK grid square there is an unusually even spatial distribution of data. An overlain 1 km² grid permits the visualisation of point density; the number of landscape interpretations per grid square varies between 1 (68 cases) and 1932 photographs showing much heterogeneity in distribution. A quantile gradient is used to display this variance (Figure 3.11) and 15% of the tested grid squares have more than 50 photographs. The analysis with Flickr (Estima et a. 2013) saw a point density of less than one photograph per km² showing Geograph to have a relatively good distribution with a density of 33 per 1 km².

Geograph offers a distinct advantage over other photograph sharing sites such as Flickr in this regard as the objective facilitates spread of coverage whereas Flickr's distribution is dependent upon population variability, photogenicity of landscape and the accessibility of an area. The fact that grid squares have multiple descriptions increases the viability of use for alternate purposes as the data can be cross validated.

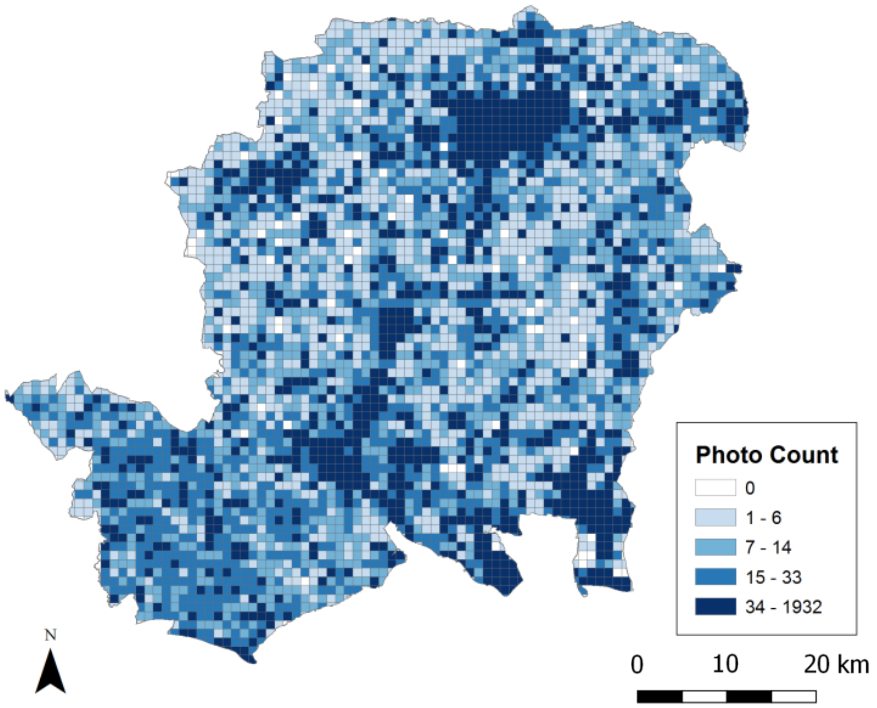


Figure 3.11 Geograph Coverage by Grid Square

3.6.3 Temporal Distribution of Geograph Data

The rate of submissions and the time at which content is geo-located has great implications for the utility of crowd generated content; if the rate of submissions is slowing there is little futurity in its relevance as reference data for land cover classification. If the interpretations are not made at an appropriate time of year they are little use as ground truth in aiding the classification of satellite imagery which is often flown over the months of May-September. Figure 3.12 indicates the temporal proximity of submissions by year and Figure 3.13 shows the number of images per grid square that have been taken at a time of year that is appropriate for use as a ground truth resource; 86.8% of grid squares do contain a photograph taken over the course of May-September. 56.8% of Geograph points have been captured and

uploaded since January 2011 with 81.7% of grid squares attributed a maximum photo temporal proximity of 2012. Only 1.8% of grid squares have a maximum temporal proximity value below that of the most recently available CEH survey date. The high temporal appropriateness of data signifies that Geograph has the potential to be a useful source of ground truth information where official survey data is sparse or too expensive to obtain.

Despite adequate temporal coverage, the rate of Geograph uploads is decreasing - submissions peaked in 2011 with 18.2% of points were uploaded; in 2012 2% fewer uploads were made. Initiatives such as Geograph are likely to be at or nearing their peak of temporal utility – users are unlikely to photograph and label their environment as profusely once every grid square has been mapped, ebbing motivation will need to be addressed to avoid such sources becoming relics of past environments. Without a directed approach analysers of VGI are faced with attempting to bridge the holes of data in sparsely populated areas and those where no volunteers want to venture. Determining and finding mechanisms of ensuring VGI projects are sustainable is a prominent research challenge.

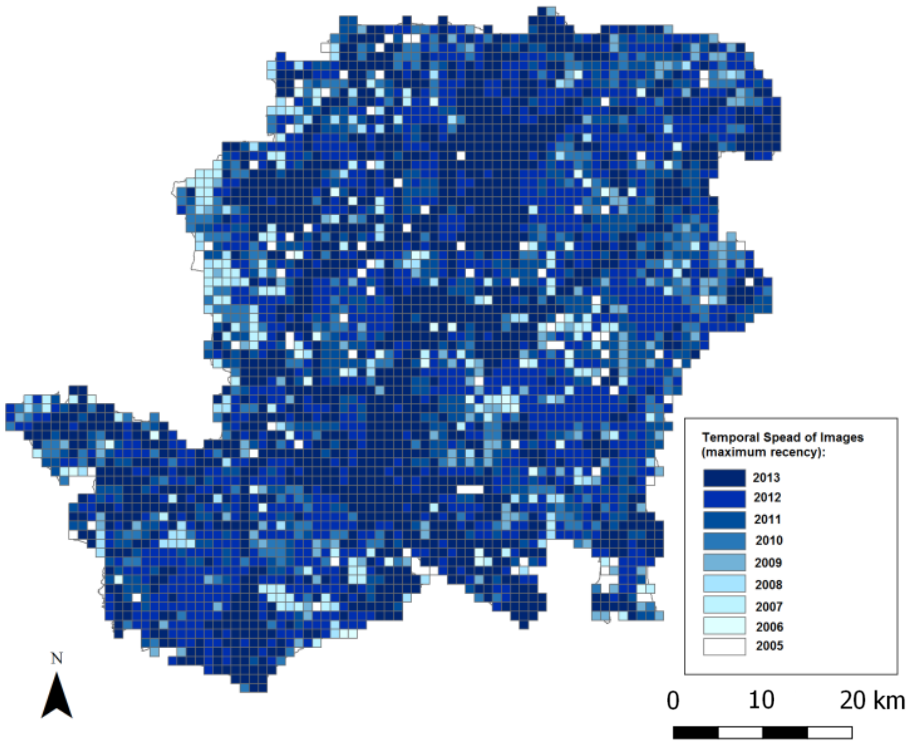


Figure 3.12 Temporal Spread of Geograph Images

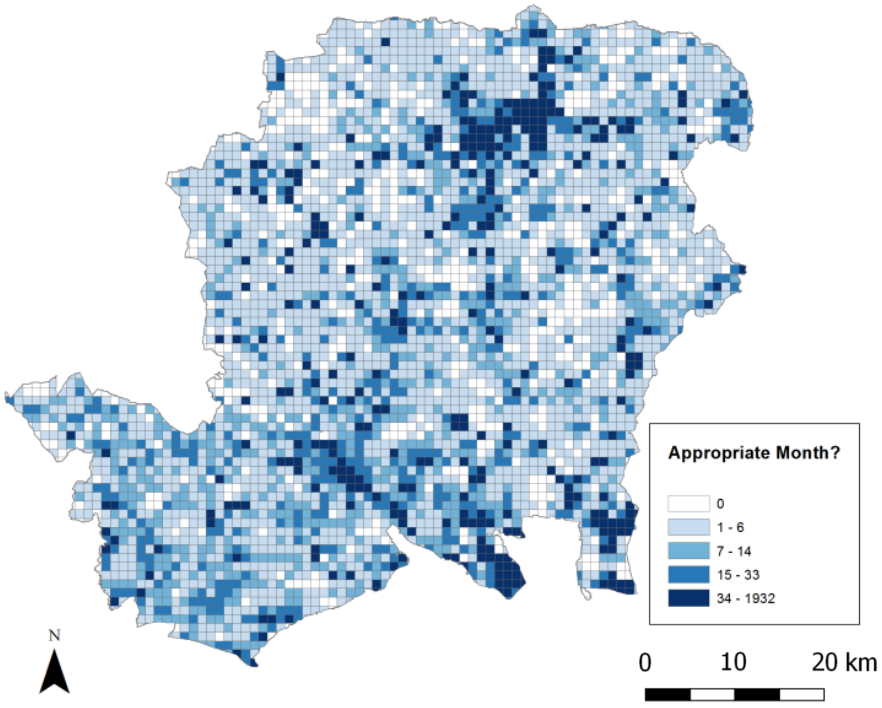


Figure 3.13: Spread of Geograph Images with Appropriate Temporal Attributes

3.6.4 Accuracy of Attribute Data and Spatial Scale of Crowd Interpretation

3.6.4.1 Preparation of Geograph Data Set

As Geograph contains EXIF data which allows the calculation of the direction from which each photograph was taken and the location of the viewpoint it is possible to map the areas that the photographs depict (see Figure 3.14). By buffering each point by the viewpoint distance and repositioning the points by the direction it is possible to calculate what the land cover in the photograph is according to the reference data. It is then possible to determine whether the depictions of Geograph users correlate with official land cover classifications using spatial joins in GIS software packages. Firstly, the Geograph points had to be clipped to the extent of Land Cover Map 2007 data and OS MasterMap in order for the join to complete successfully. As a spatial join with OSMM would exceed the 2GB limit on Arc GIS analyst operations the Geograph points and OSMM had to be split and re-clipped to avoid Null values for points exceeding the file size limit. The two clips were then spatially joined to each OSMM segment and merged at the end of the operations to produce one file. The

unnecessary data within Land cover Map (problist, construct and field code) were removed enabling the spatial join to be performed in one operation. As can be seen in Figure 3.14, coverage of the Geograph data was particularly good with very few sparsely covered areas.

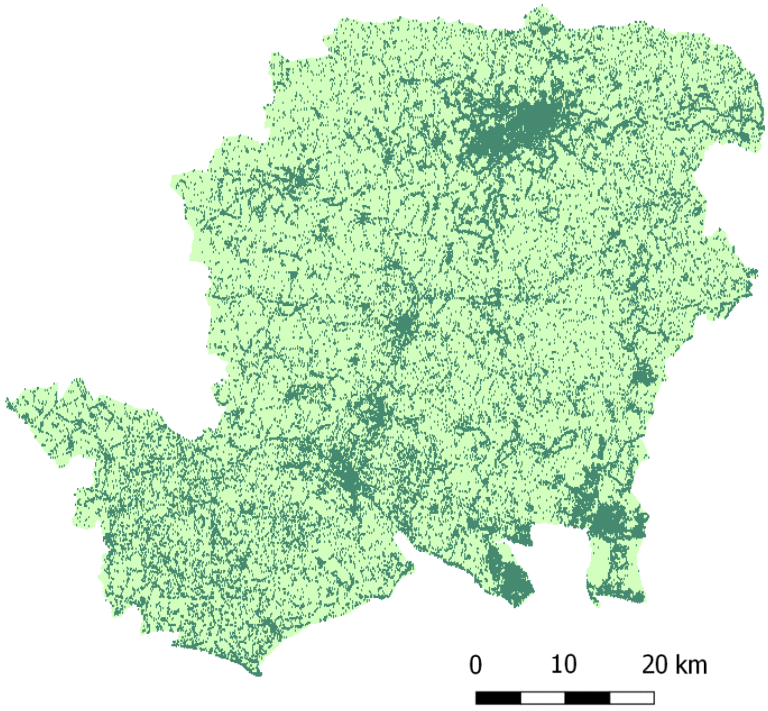


Figure 3.14: The Distribution of the 129,253 Geograph Points within Hampshire

3.6.4.2 Word Counting Script: Geograph

The relevant textual data from the Geograph and OSM data was converted into .txt format in order to analyse the descriptions employed by people to describe their photographs and uploads. In order to determine the terms that people are using to describe land cover it was necessary to create a Python script to read and sort the relevant data so that duplicates could be grouped. Code to remove all the unnecessary punctuation and common words such as ‘and’, ‘is’ and ‘a’ was implemented to reduce the amount of output data. The output was then opened within Microsoft Excel and sifted for land cover terminology. A key limitation of this technique is that the words are without context signifying that words could have been incorrectly included or excluded for example context is likely to help determine whether a term is in fact depicting land cover or the name of a street. Another limitation is also present due to the physical sifting of the data; human judgement as to what was or was not land cover could have implications on the words included within the study.

In total 14,208 unique words were found within the titles which were reduced to 369 specifically related to land cover following the sift. A further sift was undertaken to account for pluralisms and spelling mistakes e.g. alley and alleyway which both indicate the presence of a narrow artificial surface. A total of 64,811 land cover terms (155 unique) were discovered in the Geograph photo title analysis (Table 3.2) signifying that a great deal of land cover information is associated with many of the photographs. Geograph users describe the landscape at a variety of scales ranging from very broad terminology such as “Natural” and “Woodland” down to more species specific land cover information such as “Hogweed” and “Bracken”.

The most common terms used are road, lane, farmland, forest, river and park which one would traditionally find in land cover classification systems. Also worthy of note is that the context may be required to verify some of the terms as land cover – “lava” is present within the word count results but it is extremely unlikely that this refers to current land cover within Hampshire. Upon closer inspection of the dataset it appears that the term was indeed suggesting land cover but the user goes on to correct their initial judgement:

“Lava flow near Chilton Candover! Dumped on a crop of oilseed rape, on closer inspection it was not lava but I suspect dredging from a pond. It had a most odd rubbery texture.”

Table 3.2 Unique terms extracted from Geograph

Count	Term	Level	Count	Term	Level
12802	Road	2	22	Marina	3
6540	Lane	2	21	Bluebells	3
4892	Farmland	2	21	Scrubland	3
3149	Forest	1	18	Slope	2
3121	Hill	2	18	Tarn	2
3114	Park	2	18	Slipway	3
2793	Way	1	17	Seafront	2
2599	Footpath	2	17	Weir	3
2537	Street	2	16	Lavender	3
2374	Field	2	16	Logs	3
2281	Woodland	1	15	Saltmarsh	2
1702	Copse	2	13	Foreshore	2
1457	Track	2	12	Turf	3
1361	Path	2	12	Thicket	3
1162	River	2	12	Bracken	3

972	Pond	2	11	Hawthorn	3
922	Glass	3	10	Sand	3
806	Trees	2	9	Water meadows	1
806	Churchyard	3	9	Vegetation	1
669	Railway	2	9	Foxgloves	3
659	Mill	3	9	Landfill	3
642	Houses	3	8	Pebble	3
628	Downs	1	8	Vineyard	3
545	Common	2	8	Rubble	3
506	Lake	2	7	Agricultural	1
501	Canal	2	5	Floodplain	1
454	Water	3	5	Soil	2
393	snow	2	5	Brambles	3
291	Building	3	5	Shrubs	3
220	Plantation	2	4	Foliage	1
199	Wheat	3	4	Hazel copse	3
170	Heathland	2	4	Thyme	3
161	Land	1	4	Wetlands	3
153	Beech	3	3	Willows	1
139	Garden	2	3	Spoil	2
134	Pasture	2	3	Ash	3
127	Oilseed	3	3	Cropland	3
122	Yew	3	3	Buttercups	3
111	Bushes	2	3	Daffodils	3
107	Marsh	2	3	Greenspace	3
100	Orchard	3	2	Mountain	1
99	Grassland	2	2	Hay	1
94	Yard	3	2	Mounds	2
88	Coppice	2	2	Plant	2
88	Floral	2	2	Reed beds	3
88	Moor	2	2	Bulrushes	3
87	Alleyway	3	2	Broadleaved	3
85	Birch	3	2	Grapevine	3
82	River	2	2	Ivy	3
70	Natural	1	2	Knapweed	3
60	Oaks	3	2	Nettle bed	3
60	Tidal	3	2	Rhododendron	3
56	Hedgerow	2	2	Greenery	3
53	ditch	2	1	Ground	1
52	Industrial	1	1	Lowlands	1
50	Coastal	1	1	Open land	1
48	Elm	3	1	Water ditch	1
47	Flooded	2	1	Riverbank	2
45	Bog	2	1	Alpine	2
45	meadow	2	1	Sludge	3
45	Conifer	3	1	Lava	3

45	Reservoir	3	1	Loch	3
44	Lawn	2	1	Blossom	3
43	Arable	1	1	Cloverleaf	3
42	Concrete	3	1	Cobbles	3
42	Pines	3	1	Corn	3
39	Mud	3	1	Daisy	3
37	countryside	1	1	Dandelions	3
33	Wild	1	1	Fir	3
33	Clay	3	1	Gooseberry	3
32	Clearing	1	1	Hogweed	3
32	Wood hay	3	1	Leaf	3
32	Gorse	3	1	Mulch	3
29	Paddock	3	1	Nutmeg	3
28	Reservoir	2	1	Poplars	3
27	Fungus	3			
24	Island	1			
22	Clifftop	2			
22	Shingle	2			
22	thatched	3			

3.6.4.3 Word Counting Script: OpenStreetMap

The Python word counting script shows that very few words are used within OpenStreetMap to describe land cover (Table 3.3). 10,800 terms were extracted spanning only sixteen different land cover tags for Hampshire, showing far less landscape heterogeneity in comparison to the 155 unique terms obtained from the Geograph data. There were also no terms at the level 3 hierarchy (species specific level) indicating that OSM does not offer anything above and beyond what LCM2007 can offer in terms of the level of detail at which land cover is recorded. OSM is however very good for identifying areas of woodland (68.96% of the extracted terms indicated an area of woodland), recording 7,448 instances opposed to only 2,281 in Geograph (assuming that a mention in a Geograph description indicates the predominate cover). This suggests that OSM is useful for extracting the extent of woodland information, however there is a lack of specification as to what the variety of woodland is present with Geograph more likely to provide detail on whether the woodland is broadleaved or coniferous or even the species of tree present in a given area.

Table 3.3 Unique terms extracted from OpenStreetMap

Count	Term	Level
7448	Natural – Wood	1
999	Landuse – Residential	2
952	Natural – Water	1
424	Landuse – Park	2
285	Landuse – Forest	1
197	Natural – Heath	2
133	Landuse – Common	2
127	Landuse – Garden	2
118	Natural – Scrub	2
57	Natural – Coastline	1
22	Natural – Mud	2
21	Natural – Land	1
13	Natural – Marsh	2
2	Natural – Scree	2
2	Landuse – Basin	1

3.6.4.4 Spatial Correlation of Terms

A script was written using Python programming language to compare certain terms within each of the 129,253 points of study following the aforementioned spatial joins with OSMM, OpenStreetMap and Land Cover Map. The script compares a specified number of strings with another or specified set of strings and prints out the terms and associated reference numbers of cases where the strings are the same within both of the input conditions to a text file.

3.6.4.5 The Geograph Text Comparison

Most of the photographs within Geograph contain a great deal of textual information including terms referring to the land cover of the area of interest. A basic title is provided along with a description which in some cases exceeds 250 characters providing a detailed overview of what is in each photograph. The title, description, and contexts were deemed the most appropriate to analyse owing to the high frequency of land cover terms within the columns and were thus inputted into the python script where they were combined into a single string for comparison with LCM 2007.

As ArcGIS can only store strings with a maximum of 254 characters, it was necessary to truncate some of the Geograph comments using Excel's LEFT function so that the first 250 characters could be shown and further analysed. The LCM2007 categories containing land cover information (Broad habitat and Broad habitat Sub-Class) were combined and used as the string for comparison with the LCM2007 data.

3.6.4.6 OpenStreetMap text comparison

As with Geograph, the Python script was used to compare the terms being used within OpenStreetMap to the terms used within LCM 2007 for each of the locations at which there was a Geograph photo. OpenStreetMap contains a feature name, feature category, a theme and any tags that users of OpenStreetMap deem appropriate for use. Whilst there are guidelines available on the OpenStreetMap Wiki describing the terminology and suggesting appropriate tag ontology systems, the tags attributed by users do not have to conform to this, signifying that the systems are often very loosely utilised. The tags of Land cover, Land Use and Natural have been included within this analysis as although they have very little data within them they are the most frequent tag categories.

3.6.4.7 Data included within the python script:

The OSM Name, Feature, Theme and Natural tags were included within the text matching analysis to elicit the number of cases where the OSM categorisation matches that of the LCM2007 and OSMM classification. The two LCM2007 categories (Broad habitat and Broad habitat Sub-Class) were combined and used as the string for comparison with the LCM 2007 data.

3.6.4.8 Outputs of Python Script

Agreement with Land Cover Map Data

The output of the python script yielded 6,566 Geograph matches with the LCM data (after the removal of duplicate matches within one record) showing that some Geograph users are using terminology consistent with official land cover classifications – equating to 5.08% of the 129,253 photograph points for Hampshire (Table 3.4). The points that agree with the LCM classification are shown in Figures 3.15 and 3.16 and the frequency of the terms is shown within Figure 3.17.

OpenStreetMap displayed a very poor correlation with the Land Cover Map 2007 attribute data; there was only a match in 0.54% of terms for the 11868 OpenStreetMap polygons. Only three unique terms were matched (land, industrial and heath – Figure 3.17) this shows some consistency within these terms; OSM does not portray much heterogeneity in terms of the language used to describe land cover. This could suggest that people are using alternative terms to the official classification language.

Figure 3.15 Geograph Points that match LCM Data

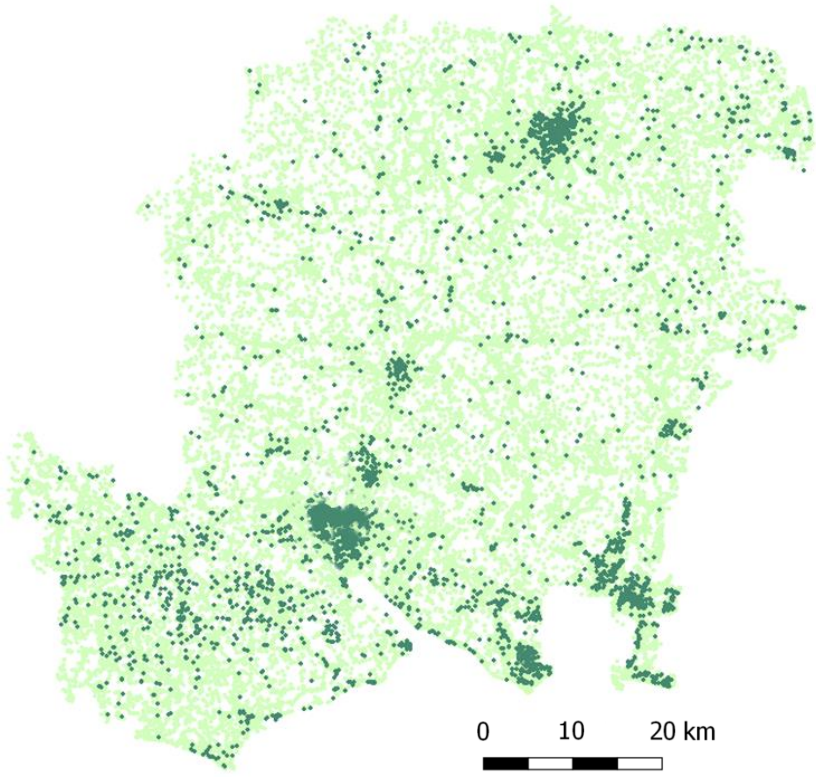


Figure 3.16 OSM Points that match LCM Data

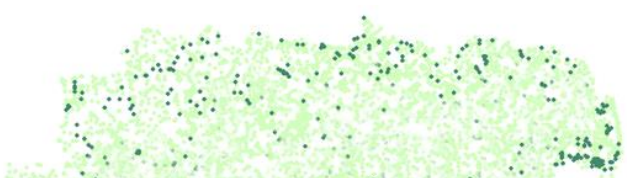


Figure 3.17 OSM & Geograph terms that Spatially Correlate with LCM

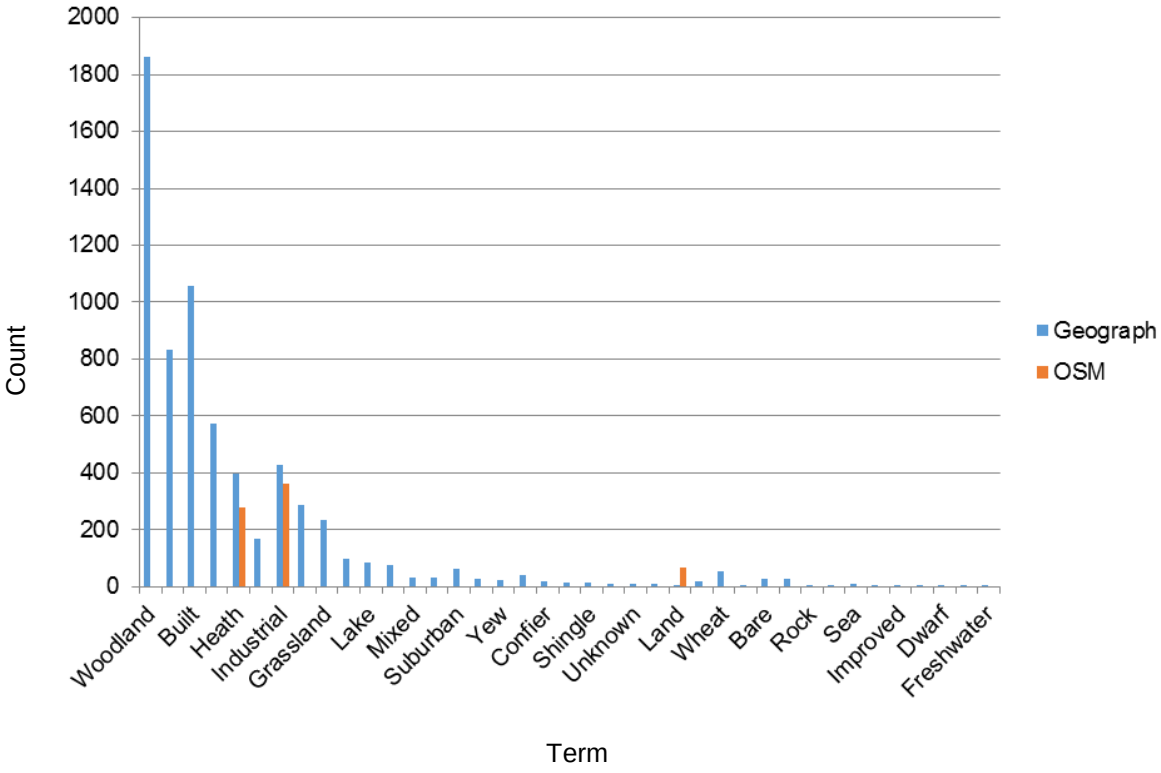


Table 3.4 Counts of OSM & Geograph Terms that match with LCM

Term	Geograph	OSM

Woodland	1863	
Gardens	832	
Built	1056	
Arable	573	
Heath	398	278
Urban	169	
Industrial	426	361
Heather	289	
Grassland	235	
Areas	96	
Lake	85	
Water	78	
Mixed	34	
Broad	31	
Suburban	63	
Deciduous	29	
Yew	24	
Grassland	42	
Conifer	17	
Coniferous	14	
Shingle	14	
Mud	11	
Unknown	9	
Estuary	11	
Land	6	67
River	18	
Wheat	52	
Sand	5	
Bare	27	
Saltmarsh	26	
Rock	3	
Rough	3	
Sea	8	
Salt	3	
Improved	2	
Shrub	2	
Dwarf	4	
Felled	3	
Freshwater	5	
Total	**6566**	**706**

3.6.5 Scale of Land Cover Attribution

In order to determine the hierarchical scales at which Geograph and OpenStreetMap users are classifying land cover it was necessary to attribute a hierarchical level to each of the terms obtained from the Python word counting script (results shown in Table 3.5 and Figure 3.18). This was based upon the Biodiversity Action Plan Broad Level and Sub Class classifications used for the LCM2007 data (Maddock, 2008)

which classify terms such as ‘Urban’ and ‘Woodland’ as broad themes. Where a term was deemed to be more specific than the LCM2007 classification system catered for, a level of 3 was used to denote a species or material specific classification.

Table 3.5 Proportions of Data at the Various Hierarchical Scales

Hierarchical Level	Terms at Level (Unique)	Terms at Level (all)
1	28 (7.59%)	9,398 (14.5%)
2	246 (66.7%)	49,836 (76.89%)
3	95 (25.75)	5,599 (8.64%)

Figure 3.18 Proportions of Data at the Various Hierarchical Scales

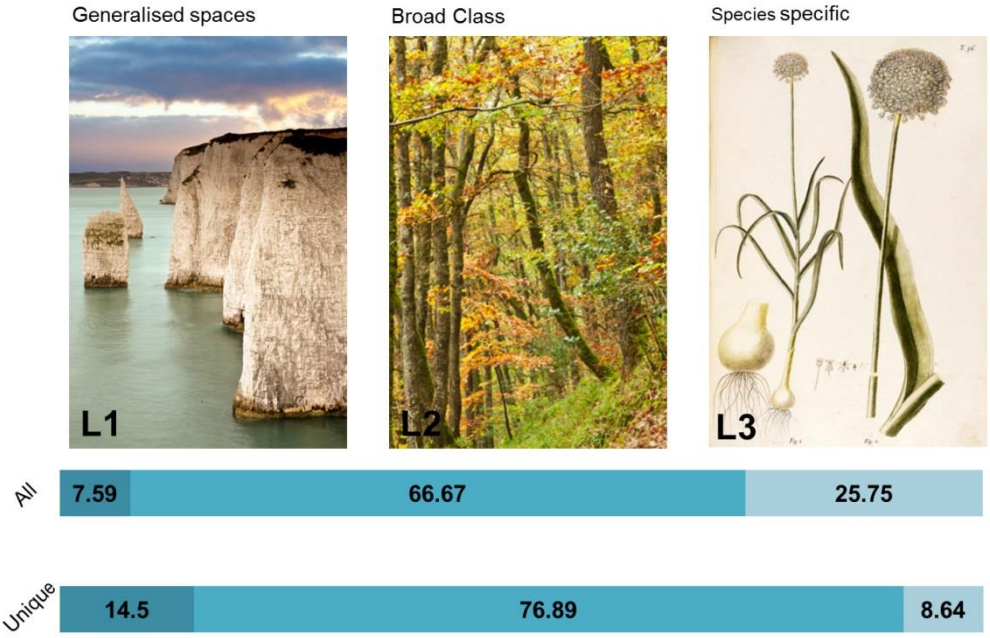


Table 3.5 shows the results of the allocation of a land cover hierarchy to each of the land cover terms. The categorisation ranged from 1 (non-specific terms that were deemed to be particularly broad in their spatial scope e.g. coastal) to 3 (species of material specific terms that were deemed to be particularly narrow in their spatial scope). The majority of the terms were at level 2 (76.89%) followed by the more generalised broad level land cover terms making up 14.5% of the total terms used. The unique terms display similar results but with greater proportions of level 3 terms (25.75%) in existence. The word counting analysis shows that people are more inclined to use specific terms to describe places than broad ones and also that there is a great deal of level 3 species specific data (8.64% of terms) that goes above and beyond the capture specification of traditional classification systems such as LCM2007.

3.6.6 Can Accurate Contributors be Identified?

3.6.6.1 User Analysis LCM

The 6,566 Geograph matches with LCM data were then analysed in terms of user contribution; the top five contributors were selected and exported as a layer to enable the visualisation of the 187 contributors that submitted a description that correlated with the LCM data (Figure 3.19). Whilst there are thousands of submissions within the Geograph Hampshire data set there is evidently a core group of users that are particularly good at defining land cover. Whilst many of the 187 contributors have only one or a few matches (and could thus have by chance defined an area very accurately) the top 5 contributors account for 59.44% of the matching submissions (Table 3.6).

3.6.6.2 User Analysis – OSMM

The top five users that utilised terms that matched the OSMM classification were exported and plotted as shown in Figure 3.20. Again there appears to be a core group of users whom define land cover in line with the classifications. The top 5 contributors (out of 200) using terminology that matched OSMM cover descriptives accounted for 50.86% of the total number of matching terms (Figure 3.20). Four of the users are the same users whose terms matched with Geograph but there is a completely different user present (18677) who had the second best match rate (but only the 16th best in Geograph) showing there to be at least some variance in the calculated accuracy of a user's terminology when different classifications are tested. This indicates that data from certain contributors may be better suited to using in certain instances than in others.

Table 3.6 Table to Show Top Contributors with Spatially Correlated Matches

User ID	Number of Matches	% of Total
17822	840	12.79
18677	802	12.21
1086	560	8.53
10124	443	6.75
1533	371	5.65

Figure 3.19: Top 5 Users with Matches with OSMM

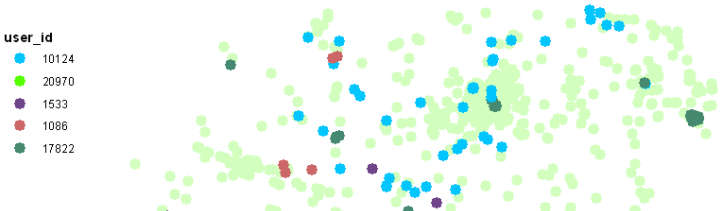
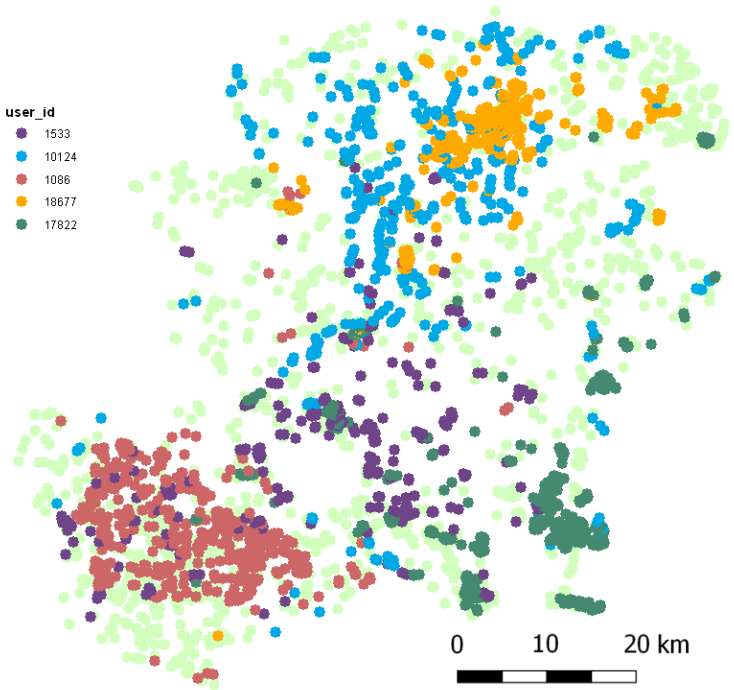


Figure 3.20: Top 5 Users with Matches with LCM



3.7 Discussion

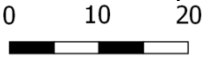
3.7.1 Accuracy of Open Web Data

This study investigates the viability of using open web data to derive land cover information using large samples from two frequently used user generated content

sites. Comparison with authoritative LCM2007 data has provided an understanding of how accurate these sources can be and provided insight into where there is room for improvement.

The fact that there are only 5.62% of explicit matches of OSM and Geograph data with the authoritative set does not necessarily render them inaccurate sources of information. A similar study (Arsanjani et al. 2013) performed at a smaller scale for land use data gave OSM a Kappa agreement of 64.79% however the authors mapped each of the present OSM tags onto the Global Monitoring for Environment and Security Urban Atlas (GMESUA) Land Use ontology so that it was a matter of whether the attributes matched spatially rather than in addition to ontologically as has been performed in this study. Work into matching informal and authoritative geospatial ontologies, such as a semi-automatic method trialled for matching OSM with OSGB (Du et al. 2013) is essential to ensure that the crowd data is logically consistent and coherent. Developing a shared conceptualisation of land surface terminology between open web sources and official land cover would make the data much more usable. The fact that there are any explicit matches shows promise for these sources.

OSM coverage is poor with only 9.83% land surface tags covering the whole area of Hampshire. Where both an OSM and OSMM polygon is present however OSM spans 59.54% of the data which is encouraging. Mooney et al. (2010) also depict how natural features are under-represented within OSM rendering it “unsuitable” for use of ground-truthing remotely sensed imagery where a great deal of randomly sampled points are required to train image classifiers (Baraldi et al. 2005). Geograph coverage has a much appropriate spatio-temporal coverage with all grid squares having at least photograph (density of 33 per km²) and only 1.8% of grid squares have a maximum temporal proximity value below that of the most recently available CEH survey date. This makes Geograph very useful as a source of ground truth information for validation regardless of whether the attribute information is actually correct. Indeed, the advancement of computer vision techniques signifies automated methods can reach accuracies of 90-6% when identifying vegetation (Luscier et al. 2006). If such methods were applied to the Geograph data set, a very accurate map of ground cover could be achieved and refined using the Geograph user annotations.

Some photo descriptors are highly detailed whilst some photographs have very little or no associated description. The extent of description ranges from the very simple such as “wood” to detailed descriptio

garlic (ransomes) in flower at the foot of the beech trees” for the same location. Also evident is the fact that users

interpret the same area differently. In the example in Figure 3.21 two photographers have recorded the same area and categorised the photographs differently. One interprets the area as “path” whereas the other determines it to be “heath”. Whilst both seem to be accurate there is certainly an issue of interpretation scale with one user choosing the predominate land cover in the area as the categorisation whilst the other user chooses to define the categorisation of what is immediately in front of them. Neither is wrong but the inconsistency in categorisation scale creates another potential problem if the data were to be used to complement official land cover data. The Geowiki project (see et al. 2012) has also experienced similar issues with scale mismatches.

Figure 3.21 Grid Square 3208 – Variance Between Photographer’s Interpretations of Land Cover



Estima et al. (2014) shed some insight into how difficult it is to provide a single tag for an area of land cover, particularly within photographs; the authors give a class of “Not Clear” to their photographs to indicate where more than one class is present not clear which one, if any, is predominant. This acknowledges that not all of the images are necessarily suited to having one tag and thus may not be useful even if classified correctly by a crowd contributor. Some of the inaccuracy may be due to contributors naming multiple classes in one area; this indicates that there is complexity in the landscape and that the contributor may not necessarily be wrong.

With regards to the precision of the data all geographs are georeferenced to at least 1km precision. At some locations the crowd classification may be correct but appear incorrect owing to the mapping units of the authoritative data. LCM2007 map has a minimum mapping unit of 0.5ha, which means that the class descriptive chosen by a volunteer mapper may be correct at that specific location (calculated form the photograph EXIF data) but not in terms of the larger vicinity. Technical problems can also influence the positional accuracy of data with Estima et al. (2014) noting that the

geotagging on some devices was not always indicative of the real world location. This could be relevant in the case of Geograph where amateur devices such as GPS enabled cameras may not necessarily contain the best GPS unit.

Positively, all of the Geograph photos and OSM tags are geographically relevant, Leung and Newsham (2012) depict how Flickr data can sometimes contain too much noise to be of use in detecting land use rendering pre-processing techniques as necessity. The requirement to sift the data is common with using such data to complement official land cover information. It shows a lack of standardisation and human error which should be expected and accounted for with all crowdsourced data sets.

3.7.2 Usability of Open Web Data

The scale at which land cover data is attributed in user generated content may not be appropriate for some users. Insurance and risk companies are not likely to want species specific data but ecologists are likely to want specific level classifications which could perhaps indicate species presence. There are very few images with broad tags (7.59 %) however these can easily be inferred from more detailed tags if a hierarchical land cover ontology was mapped onto the user generated terminology. The majority of unique terms used were at level 2 (76.89%) which is the extent of detail provided by LCM2007. 25.75% of the terms were deemed to be at the 3rd level of detail indicating the presence of a surface descriptor that went above and beyond the specification of LCM2007. The coverage of these terms is probably insufficient for an organisation requiring data of that level of detail however the fact that users are labelling things to this degree suggests that given more structure and direction, open web users are capable of classifying places to this level of detail in a more systematic way.

Geograph could be a fruitful source of crowd generated data given that there is the potential to validate. It does however have drawbacks as people are encouraged to take a picture of what is geographically interesting; the data may not be fully indicative of the ground truth in areas with users being perhaps more tempted to photograph a geographic phenomenon such as rock strata than a general view of the ground surface that is perhaps more representative of the area in the photograph. This significantly reduces the usability of the data however, moderator validation and user change suggestions will certainly reduce the possibility of irrelevant tags and descriptions (Fonte et al. 2015). The community validation aspect of OSM (Palen et

al. 2015) also denotes that data downloaded from OSM will more than likely have been validated to some extent; the presence of community created content guidelines also increases the likelihood of the data's usability. Whilst guidelines do exist on the creation of land cover it is important to note that they are used infrequently and with little regulation meaning that the quality of land cover data is likely to be heterogeneous (Mooney et al. 2010).

3.7.3 Proposed Methods for Enhancing Open Web Data Usability

It is important that prospective users of open web data engage with the producers of the information at an early stage, particularly with regards to projects like OSM where there is a history of volunteers uploading data for a specific purpose. The Humanitarian OpenStreetMap team continually work with aid agencies and the appropriate authorities to ensure that they are providing accurate and usable data at times of humanitarian crisis (Palen et al. 2015). There is sufficient diversity in the user base of OSM that it is likely that for a given project someone will be interested and motivated sufficiently to want to meet the specifications of a scientific data collection target. In the case of land cover, a proposal on the OSM Wiki already acknowledges the insufficiency of land surface tagging and seeks to improve the consistency by which people map land cover on OSM. Developing a consistent tagging system would benefit greatly from input from those who might want to use the data and provide a definite motivation to create it.

The presence of a core group of users producing data with high attributional accuracy within the open web data is interesting as it allows for experimentation with contributor weighting. If a particular contributor has historically proven to be a trustworthy labeller of an image it follows that greater onus can be placed upon any further classifications they make. If for a given parcel of land there are multiple, differing contributions, the trusted contributor is likely to have provided the most appropriate classification. Nowak and Rüger (2010) look at inter annotator accuracy and develop methods for filtering noisy, non-expert judgements that cloud the overall veracity of the classification. Raykar and Yu (2011) have also developed an annotator ranking system to detect low quality annotators which could prove useful for Geograph where there are consistently multiple contributions for the same area.

The following excerpt from the OSM Wiki page further exemplifies how some authoritative consensus on how cover should be mapped would be really helpful to the project:

“One of the most difficult cases for landcover mapping is woodland/forests; There is not currently a good tag to describe a landcover of trees as opposed to a landuse of timber production for which landuse=forest is appropriate or natural=wood for primary unmanaged woodland. In addition, landuse and landcover are often confused”.

Additional structure is required to ensure that information is collected in the correct way – instances where multiple land covers are described because of landscape heterogeneity can easily be identified if associated tag information is stored in an appropriate way. For example, a tag could be introduced in OSM that marks ecotonal areas where there are multiple types of legitimate land cover. Tags could make users input a hierarchy so that a broad and specific tag could be recorded and stored. Greater specifications and guidance materials could also be used more effectively to ensure that users are comfortable in tagging the land surface and so that it could also be used for other purposes such as environmental monitoring. This is a form of ‘directed crowdsourcing’ – ensuring that volunteer efforts meet the requirements of those using the data following collection.

3.8 Study Critique

3.8.1 Spatial scale of investigation

Whilst the results of this investigation are true of Hampshire, they may not be indicative of the appropriateness of crowdsourced datasets for the whole of the UK. It is necessary to conduct a full scale analysis of the UK to comprehensively assess the viability of such methods for sourcing land cover information. It may also be worthwhile investigating the difference in accuracy of land cover attributes and extents between rural and urban areas – investigating Hampshire as a whole (as has been done in this project) is not very indicative of any spatial heterogeneity there is in the accuracy of crowdsourced input within the different cover areas within Hampshire. Different altitudes, slope, land cover and degree of rurality could cause disparity in the crowd’s interpretation of data; it is therefore important to look at other areas of the UK, biomes and parts of the world in future studies.

3.8.2 Choice of Data Used

As the OSM XML data for Hampshire is particularly large and difficult to extract it was only possible to extract the land cover, natural, surface and land use tags. Other

information relevant to land cover could be available but placed under a different tag. It may be prudent to investigate the content of other tag sets when doing a full scale analysis. Only the most recent change set is included with OSM; people can edit anything that is on OSM and only the latest version is saved. It would be interesting to look at history of objects as depicted by Mooney (2011) to discover how different people have identified the same object and look for terminological changes to land cover. The temporal nature of the historical data would also enable an investigation into whether OSM could be used for change detection.

3.8.3 Methods Used to Analyse Data

The Geograph comments used within the word matching analysis had to be truncated to 250 characters to enable them to be stored within Excel and spatially joined within Arc GIS. Whilst in some cases this cuts out a lot of information it is unlikely that the land cover is not mentioned within the first 250 characters if it is mentioned at all. This signifies that the found agreement with LCM data can be seen as a minimum level of agreement.

The Python text matching code signifies that terms are analysed and grouped without any consideration for the context of the word. Owing to the scale of the Geograph data set it was not possible to go through all the descriptions, titles and tags individually without the aid of a script which pulls out any words that match. This does however remove any associated context from the sentences leaving a word which could have been used as a name or to describe things other than land cover. An example of such a word is ‘lava’, which was attributed to an area in which the land cover looked like lava but was in fact dredging from a pond.

3.9 Conclusions

1. **Can indirect sources of land cover information be used to provide an accurate land cover class?**

The analyses undertaken suggest that the approach of incorporating VGI with authoritative land cover information provides a promising avenue to providing *in situ* ground truth required to generate high quality land cover maps. Owing to the adequate coverage they could be useful as reference data where existing methods misinterpret imagery and the temporal relevance could prove useful in flagging up landscape changes. In terms of attributes there is some very detailed and potentially minable information.

2. **Do open web data producers describe features at an appropriate scale?**

The scale at which open web contributors have labelled the geographic features within OSM and Geograph has been analysed to ascertain the extent to which open web sources can provide anything above and beyond the scope of authoritative land cover data sets. There is some very detailed, species specific information in Geograph which could be very useful to those requiring land cover information at a high level of detail.

3. **Can core producers of accurate open web data be identified?**

This study has ascertained that core users are present within both OpenStreetMap and Geograph that use terminology that is in concordance with authoritative land cover ontologies. The fact that different core users were found when comparing against OSMM as opposed to LCM2007 suggests that different users may be more naturally inclined to using the terms within different classification systems. This shows the importance of understanding the contributor base. There is potential to use the knowledge of the contributor base to enhance the usability of the data; taking only contributions from trusted users or weighting their contributions so they have more gravitas when taking into account the descriptions from multiple contributors may provide greater confidence in use of open web data.

4. **Is it viable to use indirectly sourced crowd generated content to enhance existing authoritative data sets? i.e. What limitations must we be aware of when working with such data?**

The lack of standardisation in labelling is currently prohibitive to its unedited use. Without rigid ontology rules, which constrain the fluidity of crowdsourcing projects, substantial work is required to index ad hoc terminology within a nationally applicable standard. There are also some substantial barriers from a business perspective in terms of quality assurance and the current lack of interoperability with other data

sources. It is clear that crowd generated content cannot be used without a great deal of pruning which could remove the efficiency in the updating of topographic databases. As such it is necessary to experiment with techniques of generating a refined dataset first hand so that usage can be as near to ‘off the shelf’ as possible.

Chapter 4 | Refinement of Approaches to Directed Crowdsourcing

	Research Objectives Addressed In This Chapter:
1	Establish the degree of accuracy required from ecological datasets and any concerns that the potential use of crowd generated content presents to users of ecological datasets.
2	Obtain open web data and establish whether it could provide any contribution over and above that of the authoritative data collection and classification systems.

3	Understand how the crowd identifies ecological features field *in situ* and develop the best practises for developing a web based crowdsourcing platform.
4	Develop a system for harvesting crowd contributions in a directed manner in the context of Phase 1 Habitat Mapping and quality assessment of crowd contributions.

4.1 Introduction

In order to develop an effective directed crowdsourcing platform whereby users are led to complete a specific crowdsourcing task (as discussed further within Section 5.1) it is important to understand how the user interacts with the platform, classification system and tasks. This chapter details two preliminary studies which aim to determine how crowd annotators classify ecological features within the field and any problems associated with them doing so. It is hoped that the second of the preliminary studies can be used inform the initial design of a crowdsourcing website designed to collect ecological feature data. A brief usability study of the web page is conducted to further inform the design of the crowd annotation harvesting site used within Chapter 5.

4.2 Aims and Objectives

- To gain insight into any barriers presented to people with non-professional knowledge of land cover in how they interpret and use land cover classifications *in situ* (4.3).
- To test the usability of an initial website designed to collect land cover classifications via the web (4.4).

4.3 *In Situ* Land Cover Classification

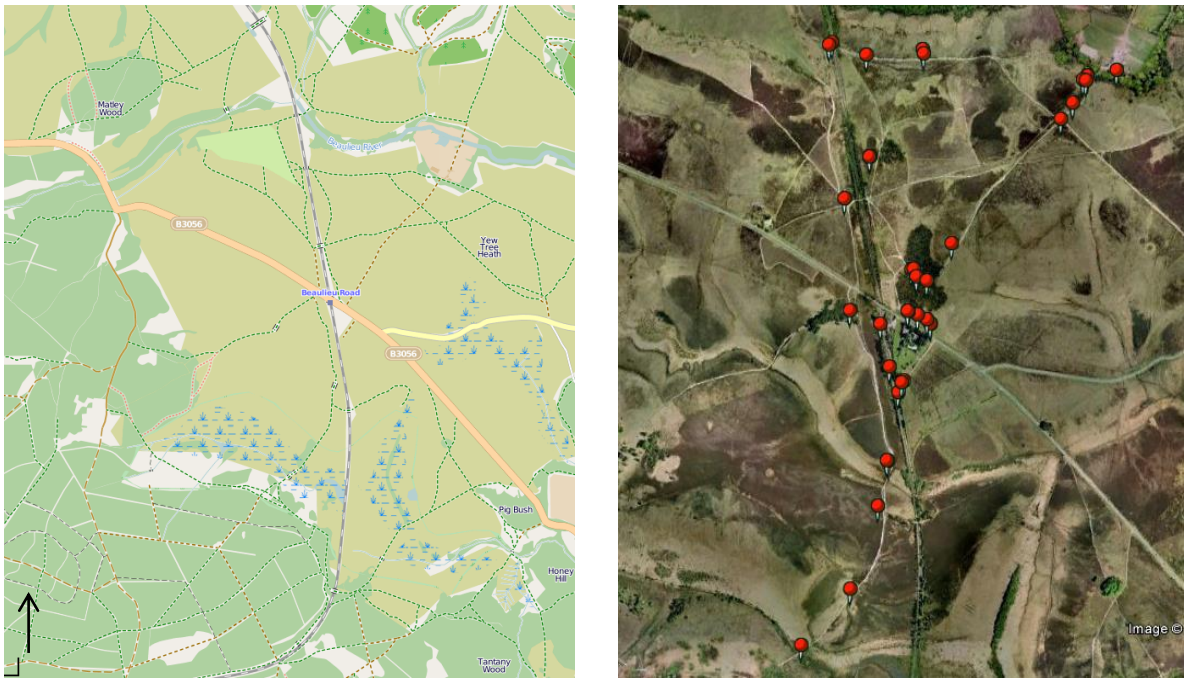
4.3.1 Objectives

- To determine any the barriers to the successful attribution of land cover.
- To determine any difference in classification accuracy between different land cover classification schemes.
- To observe the spatial scale at which non-professional are able to and choose to identify and distinguish land cover.

4.3.2 Preliminary Study Location

The fieldwork was conducted at Beaulieu Road Station (New Forest) owing to the moderate variance in land cover types (Figure 4.1). The rural nature of the area signified that there were more land cover types than one would find in an urban area. Also the land cover classifications have very few urban classifications which would have signified a lack of diversity in the survey results for the parts using the official classifications.

Figure 4.1: Fieldwork Location



Left OSM, New Forest, Right locations visited mapped on Google Earth.

4.3.2 Methodology

A participant with no professional experience of land cover classification was recruited and taken to the study site where they were then asked to mention each time they transitioned into a different area of land cover. At these points a photo was taken of the new land cover (using a camera with geo-positioning) and the participant was asked to classify the new area. This transect walkthrough method enables the observation of the ease and feasibility of non-professionals mapping land cover. Land Cover Map 2007 (CEH), the National Land Use Database Land Cover Classification (contained within OSMM) and the OpenStreetMap Classification guidelines were used for each parcel of land cover investigated to allow a comparison between their

ease of use. The participant was also asked to devise their own classification of each area prior to choosing one of the official categorisations. As the participant was using the classification systems their decision were noted using think aloud protocol (Tirkkonen-Condit, 1990) alongside any other comments that were indicative to the usefulness and practicality of the guidelines.

Whilst we can only draw limited conclusions from a study with one participant, the performance of this walkthrough transect allowed the identification of potential difficulties with a non-professional environmental enthusiast identifying land transitions in the field with professional ontologies.

4.3.4 Observations

In total 31 land cover classifications were made with the three classification guidelines and a separate individual interpretation conducted prior to assigning official classifications. Following the analysis of the post field day interview and transcription of the comments made whilst undertaking the fieldwork, a number of issues can be raised with using the existing classification systems and with giving the public free reign over the terms that they can use to classify an area (as shown in Tables 4.1-4.3).

4.3.4.1 Spatial Scale

One of the more significant problems was the spatial scale at which to classify an area. As the user was told to mention each time that they thought they were in a new area of land cover, the scale at which land cover varied was very much at the disposal of the person classifying the data. Specific examples are shown in Table 4.1 below.

Table 4.1 Issues Identifying Classification due to Spatial Scale

Classification System	Chosen Land cover Class	Issue with classification	Photograph
Land Cover Map 2007	Urban	Does not seem appropriate for a tarmac track. The only alternative is bare but that is not right. Definitely a spatial scale issue.	

Land Cover Map 2007	Water Lake	There is nothing to choose at a smaller scale apart from bog but that would not be appropriate.	
Land Cover Map 2007	Freshwater	The river tag is too big, I have used freshwater as it does not necessarily indicate a scale – even though it should be a smaller scale.	

4.3.4.2 Classification Appropriateness

The classification systems being used whilst participants undertook the land cover classification fieldwork were not appropriate for the land cover encountered at times, despite them being designed for use in the United Kingdom (Table 4.2). Often a similar yet not entirely appropriate classification had to be used instead with there being one case where no classification was deemed appropriate.

Table 4.2 Issues Identifying Classification due to Classification Inappropriateness

Classification System	Chosen Land cover Class	Issue with classification	Photograph

Land Cover Map 2007	Bare	Super category of urban is incorrect – you can have bare soil in a rural area.	
Land Cover Map 2007	N/A	No appropriate classification for temporary storage of felled trees.	
NLUD	Shrub	Reluctantly put shrub down, I think scrub would be more appropriate.	
NLUD	Heathland & Shrub	Scrub is more appropriate than shrub.	
OSM	Land use = woodland	Land use doesn't imply land cover – it should have a natural / land cover tag.	

4.3.4.3 Insufficient knowledge to use classification system

Certain terminology prohibited the certainty of appropriate classification of land cover. Instead an educated guess was relied upon to correctly classify the areas (Table 4.3).

Table 4.3 Issues Identifying Classification due Insufficient Knowledge

Classification System	Chosen Land cover Class	Issue with classification	Photograph

Land Cover Map 2007	Grassland (Acid / Neutral)	Cannot tell the acidity of grassland by looking.	
NLUD	Metalled roadway	I am not sure what metalled roadway means but it's a roadway so it seems appropriate.	

4.3.4.5 Virtues of Classification Systems

Whilst the user proclaimed the free classification system to have certain merits, they preferred to use an official system as it provided a standardised system by which to classify things. The free classification system was particularly useful as it permitted the use of adjectives and descriptors which helped to further specify the type of land cover. The participant regularly noted things such as the quality of grassland, the width of paths and the density of woodland which provides important information regarding the land cover rarely captured with official classification systems. Particularly relevant is the descriptions used to convey the temporality of water bodies – current land cover classification systems do not incorporate any regard for temporality despite the importance of seasonal alterations and land cover change. Also noteworthy was the fact that the participant used spatial operators to describe areas with more than one type of land cover of significance; "trees within grassland" describes an area of mixed land cover which is predominantly grassland with some trees within the land segment. Official land cover classifications do not offer the flexibility to incorporate such spatial relations despite the fact they would more accurately portray an area.

The participant was given the option to choose either one predominant land cover or several classifications if it was felt that the mixture of land cover present deemed several classifications to be the most appropriate option. The participant chose to use more than one classification on several occasions and it is worth noting that traditionally only one would be used. The results were subsequently compared with the Land Cover Map classification for each point to give an accuracy score:

Open Classification: 6/31 with OSMM
2/31 with LCM

LCM Classification: 6/31

The highest match with Land Cover Map was when the participant used the LCM classification system; however there is still very little correlation with what the participant perceives and what the classification attributes each area. LCM classifies the 31 test areas with just four LCM categories (suggesting that the area had low land cover heterogeneity) whereas the participant chooses to use 15 (suggesting high spatial heterogeneity). This indicates that the participant is identifying features at a more detailed scale than the LCM data which takes a more generalised view of the land parcel.

4.4 Development of a Land Cover Crowdsourcing Web Page

4.4.1 Objectives

- Create a crowdsourcing web page based upon findings from the *in situ* observations.
- Recruit some crowd contributor participants to test the crowdsourcing platform.
- Determine the usability of the website and classification systems.

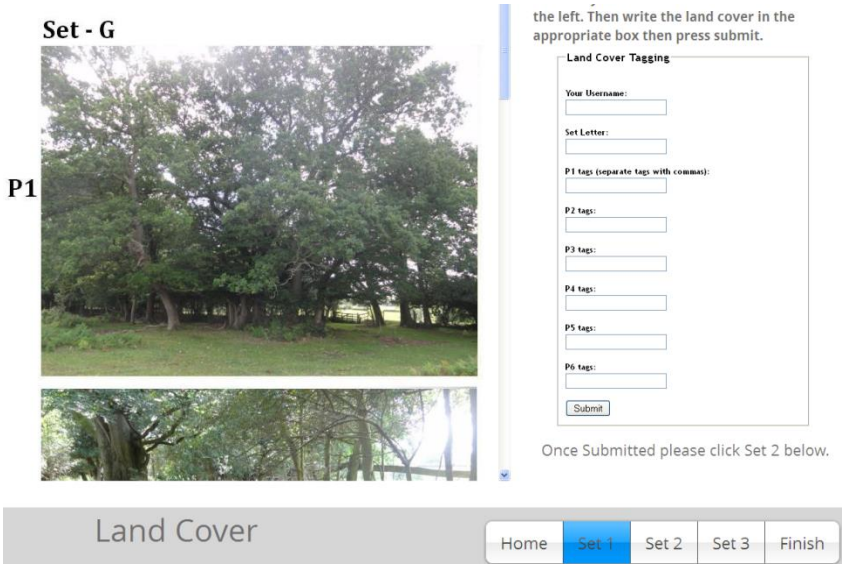
4.4.2 Methodology

A web page has been designed and created that fulfils the purpose of investigating how successfully groups of people can identify land cover and discover the spatial extents at which they do so. The primary aim of this study is to evaluate the usability of the website with the intention of informing the development of a directed crowdsourcing webpage (Chapter 5). In this small experiment user results will be compared to a known official standard (LCM2007) to determine the extent to which people can choose appropriate land cover types. The usability of the various land cover classification systems in a web labelling based context are also evaluated. Nine sets of six georeferenced photos from Beaulieu, New forest (preliminary fieldwork study) and Beaulieu Road Station area have been uploaded to a server created to host the crowdsourcing site which has script enabling the display of a random set of photos (to ensure any variances in classification difficulty do not affect the results).

4.4.2.1 Land Cover Crowdsourcing Web Page Development

A web page has been created (using HTML, Javascript and a freely available server) to facilitate the submission of land cover tags for pre uploaded photographs. The user is shown a random set of photographs and instructed to tag them in three stages. Several screen shots are presented and explained within Figure 4.2.

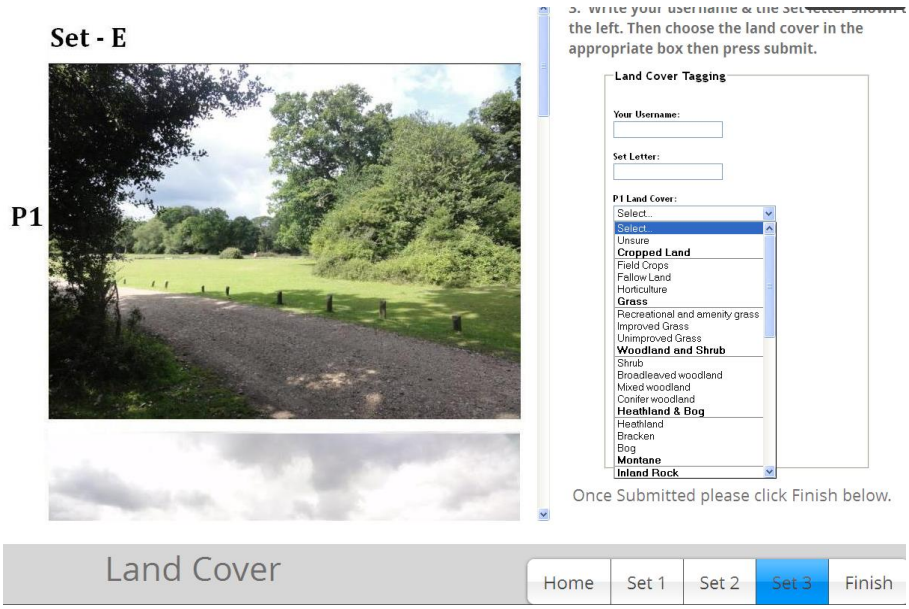
Figure 4.2: Crowdsourcing Webpage Screen Captures



Stage 1: The user is asked to openly tag the photographs i.e. tag the photograph with whatever the user deems most appropriate.



Stage 2: Users are then randomly assigned with another set of photographs for which the user has to choose a land cover tag from a pull down box containing Land Cover Map 2007 ontology. Users can choose the hierarchy level by choosing a generalised or detailed term.



Stage 3: Users are then randomly assigned with another set of photographs for which the user has to choose a land cover tag from a pull down box containing the National Land Use Database land cover ontology. Users can choose the hierarchy level by choosing a generalised or detailed term.



Stage 4: A few questions are then asked to enable the understanding of which classification or tagging system was most appropriate for users and to also find out which users thought produced the most accurate representation of the land cover within the photographs. The level of familiarity with land cover terms and the outdoors is also requested so as to indicate certain groups of people that are particularly good at identifying certain land cover types.

4.4.2.2 Recruitment of Participants

Links were posted on twitter and emailed to colleagues and friends. A member of OpenStreetMap also posted it to the OSM mailing list with a note requesting that if it was to be filled in that they indicate their OSM background on the form. Participants had a range of previous land cover knowledge with the majority having “low” or “mediocre” knowledge as shown in Table 4.4. The responses from 45 participants were used in the analysis.

Table 4.4: Breakdown of Participants

Level of Knowledge	No of Participants
High	5
Mediocre	16
Low	18
None	6

The photographs utilised within the study were plotted within ArcGIS and spatially joined to the Land Cover Map data to provide a set of “accurate” classifications for each of the points. The responses of the participants for the five images looked at

were then compared to this standard in order to determine the accuracy of classification for each participant and for each land cover type (with each participant allocated one point for each correct classification).

4.4.3 Results of Web Classification Experiment

4.4.3.1 Classification Accuracies

Table 4.5: Average and Total Scores by Classification

	Open	LCM
Average	1.96	0.71
Total	86.96	30.71

Table 4.6: Average and Total Scores by Knowledge

	Low / No	High / Good
Average	2.75	3.57
Total	66	54

The average score for people with Low / No knowledge was 2.75 which was higher than the average score for people with High / Good knowledge at 3.57. This is surprising and could indicate that previous knowledge is not a necessity for identifying land cover. As the difference between the averages and total scores are negligible any distinct advantage of land cover knowledge can be ruled out. The varying levels of modesty employed by the participants could also be at play with people underestimating their knowledge of land cover terminology.

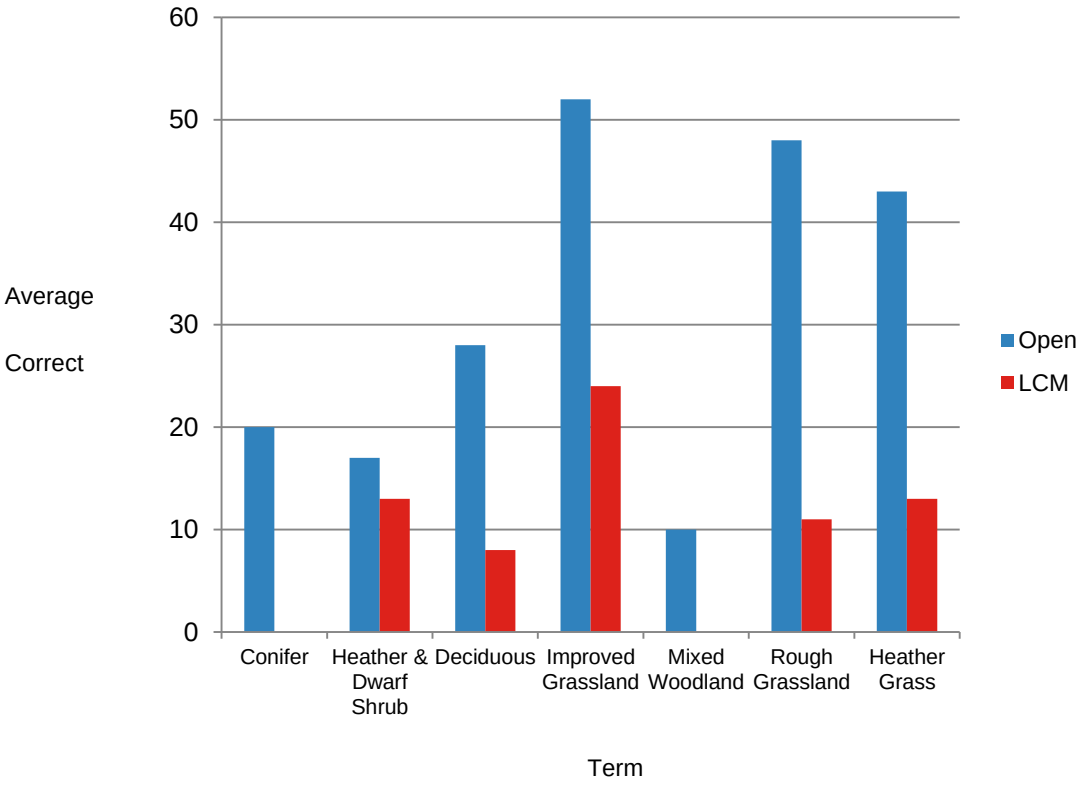
People scored much better with the open classification (66 matching inputs) compared with the LCM classification system (54 matching inputs). The average score when given the open classification photo set was 1.96 and with Land Cover Map 0.71. This signifies that people were slightly more inaccurate without the constraints of a classification system, though people were able to choose more than one land cover type with the open classification system which would have influenced the results. With LCM it is only possible to attribute one classification to a parcel of land.

4.4.3.2 Problems with classifying photographs

Through looking at the results by type of land cover (Graph 4.3) it was possible to

understand where participants went wrong in their classifications. Participants were often referring to things at a lower spatial scale than what the official data caters for - for example in areas of rough grassland with a small path in the middle, participants often referred to the path as the main type of land cover choosing “bare” or “urban” when in fact Land Cover Map does not classify parcels of land at such a small scale. One very common error was the use of scrub instead of shrub – as the terms are quite similar it may be the result of the participants’ non expert understanding of the intricacies between the two terms. The main finding of this experiment (as with the fieldwork) is that people are not good for thinking in generalised terms – they classify what is immediately around them. People consistently add more heterogeneity which could be good for creating high resolution land cover data. It does not however occur at a consistent enough spatial scale to make the data attained from crowds immediately usable.

Figure 4.3 Average Percentage of Photos Correctly Classified by Category



4.4.3.3. User Preference for classification methodology

A brief questionnaire was developed to evaluate the usability of the pilot version of the land cover classification web page. This was divided into two sections which

evaluated comprehension of the classification systems and the perceived accuracy of annotator classifications using each system. Qualitative statements surrounding usability were also collected and grouped into relevant themes and counted to enable understanding via a thematic content analysis as shown in Tables 4.7 and 4.8 below.

Most participants preferred the NLUD classification system (44%) followed by the open system (38%) and LCM (18%). The open system was however deemed most accurate by 42% of participants owing to the flexibility offered and the fact that they were able to describe multiple land cover areas, with 31% and 27% perceiving NLUD and LCM as most accurate respectively. LCM was critiqued for being too lengthy and too hard to understand suggesting that it is not a viable classification system for crowdsourcing. The NLUD system was however found to be of appropriate difficulty as it featured understandable terms. The open classification system would generate too many unique responses to be viable for use within authoritative data, though can provide some species level small scale data.

Table 4.7: Positive Comments Regarding Classification Methodology

Positive comment	No of mentions
Open Classification	
Could write the cover using my own words	5
More Flexible	4
Easiest to Understand	4
Could describe multiple areas of land cover	1
It was quicker to fill out	1
Land Cover Map 2007	
Offers the most choice	6
Appropriate choices are available	1
Deals with varying land cover types better	1
Less effort than having to construct class – easier	1
Rigid categorisation, more defined	1
NLUD Land Cover	
More understandable categories	14
Fewer categories than LCM	5
Better for me	2
Unsure	1

Table 4.8: Negative Comments Regarding Classification Methodology

Negative Comment	No of mentions
Open Classification	
Creates too many different answers	1
Too difficult if you have no knowledge	9
Land Cover Map 2007	
Had duplicate terms and too many terms	8
Has too many classes	6
Terms seem unfamiliar	4
NLUD Land Cover	
Pre-defined system is more restrictive	3

4.5 Summary and Recommendations for Next Iteration of Crowdsourcing Site

Two preliminary studies have been undertaken that will inform the design of the next iteration of a crowdsourcing platform designed to generate habitat data that could potentially be used as training data for remote sensing classifications.

4.5.1 Barriers to the Successful Attribution of Land Cover

The spatial scale of analysis was consistently an issue when using each of the classification systems in the field. The crowd surveyor wanted to identify features at a more specific level than the classification systems would permit, such as small temporary ponds which could only be classified as ‘freshwater’ given the ‘river’ tag was not appropriate. Ensuring that annotators submit a classification at more generalised levels with example training resources will help ensure that they feel more comfortable with assigning a generalised cover term and thinking in terms of the

overriding cover within the immediate landscape rather than the immediate cover in the small scale vicinity.

Ensuring that the classification is appropriate for the features within the landscape is also important; using the NLUD system it was necessary to identify a broad class of ‘urban’ before using the class of ‘bare soil; as a sub class. This implies any bare soil recorded within the predominately rural areas of classification will be recorded as ‘urban’ at the upper hierarchical level, misrepresenting the landscape. Following this initial study a classification schema will be chosen that is deemed appropriate for the test environment.

Insufficient knowledge of the participant was also a problem with the participant unable to determine the acidity of grassland in the field. Whilst it is possible to record this from the presence of indicator species the contributor in question did not have sufficient training in ecology to identify this. Galloway et al. (2009) compare novice observations to those of observers with more project specific training and discovered they are comparable with slight differences. It could thus be prudent to look at the effect of different groups and the extent to which annotator membership impacts upon classification accuracy.

4.5.2 Usability of the Crowdsourcing Site & Classification Systems

The comprehensibility of the NLUD classification terms over the LCM system was consistently highlighted as a positive point of the classification system (N=14). The choice and flexibility of a given classification system were also highlighted as positive aspects of the classification systems which should be incorporated into the next iteration of the page design perhaps through the allowance of multiple sub classes.

The perceived accuracy of crowd classifications using the various classification systems is useful in that it gives an indication of how well the contributor thought they were able to map the classification system on to reality, however it would be useful to ascertain this information for each classification in order to correlate perceived accuracy with the accuracy of classification. Zhai et al. (2014) depict how quality control is always pivotal to achieving trustworthy results in projects tapping in to online participants and use crowd consensus self-check and statistical pruning of likely inaccurate results to achieve a trustworthy dataset. Contributor confidence scores could be used to prune the dataset so that only recordings likely to be of high quality are used.

Chapter 5 | Assessing the Utility of Directed Crowdsourcing in Habitat Classification

	Research Objectives Addressed In This Chapter:
1	Establish the degree of accuracy required from ecological datasets and any concerns that the potential use of crowd generated content presents to users of ecological datasets.

2	Obtain open web data and establish whether it could provide any contribution over and above that of the authoritative data collection and classification systems.
3	Understand how the crowd identifies ecological features field in situ and develop the best practises for developing a web based crowdsourcing platform.
4	Develop and assess a system for harvesting crowd contributions in a directed manner in the context of Phase 1 Habitat Mapping and perform a quality assessment of crowd contributions.

5.1 Introduction

Producing accurate maps of habitat loss and fragmentation is integral to understanding and managing the environment. Traditionally land use and land cover inventories and maps are generated through ground surveys and through the processing of remotely sensed imagery. These techniques do however require significant resourcing in terms of surveying time and processing costs which is why researching innovative methods of sourcing such information is imperative. Many authoritative habitat and land cover maps could benefit from additional ground truthing to enable improved identification of species distributions and transition gradients (Dickinson et al. 2010). VGI presents opportunities for improving current practices of data collection and annotation, providing extensive sample coverage alongside viewpoints from multiple 'sensors' that would not otherwise be economically feasible to collect. Minimal quality control practices do however prompt concerns over accuracy and completeness (Heipke, 2010); VGI typically exhibits considerable quality variance ranging from imprecise digitisation to poor label accuracy making usage for planning and scientific purposes challenging from a data quality perspective.

Crowdsourcing is an umbrella term for the array of different information sourcing activities carried out by citizens. Data can be passively submitted and collected by an interested party (e.g. strings mined from location tagged sources such as Twitter and Flickr) or be actively collected for a specific purpose, directed by an interested party. The research community has developed some insight into how accurate VGI can be and how it can be utilised for validating existing datasets (Goodchild and Li, 2012); however, there is scope for VGI and the more directed approach of citizen science to be used in image classification. With active or directed crowdsourcing volunteers are steered into submitting geospatial information according to a pre-defined scope and specification. It differs from passively sensed data as information is knowingly

submitted for an often pre-defined research hypothesis to meet distinct goals that are explained to the crowd to encourage volunteers to submit information to a particular scientific cause. Directed crowdsourcing involves forcing structure onto a data source that traditionally has no or little structure and is thus a potentially useful approach when ascertaining information for the remote sensing training process; the sampling for which must be regimented and be as scientifically credible as possible.

In combination with machine learning, focused sampling has the potential to produce training samples which, with appropriate weighting, (such as down-weighting less accurate contributors and emphasising the contributions of accurate annotators) could yield timely outputs of appropriate thematic accuracy. The parameters under which crowdsourced annotations should be collated and weighted by accuracy to reduce label noise have been explored in machine learning (Lease, 2011) yet are rarely applied in the domain of VGI. This research aims to assess the potential value of directed VGI in habitat classification with the hope of informing authoritative data collection methods.

5.2 Aims

This study aims to achieve an understanding of whether directed crowdsourced GI can be useful in identifying ecological phenomena and to determine whether any traits within the crowd itself and the data can be seen as predictors of quality and accuracy in an ecological data context. This included the understanding and use of fuzzy accuracy assessment techniques which are integral when utilising information with uncertainties in both the crowd observations and the assumed ground truth. It tests the variability between crowd and professional interpreters when faced with identical classification tasks. Three objectives were defined to explore the study aim:

1. Can scientific sampling procedures be pushed onto volunteer data?

2. Can the crowd produce training samples which are as credible as those provided by a professional ecologist and can any variance in accuracy between feature types be identified?

3. Can predictors of crowd accuracy be established to enable the elimination of information that is detrimental to classification accuracy?

As there a very few cases of VGI being used within the classification of remotely sensed imagery, this study aimed to provide insight into how directed VGI could be useful to ecologists and to investigate approaches to ensuring that VGI is fit for the intended purpose of use.

5.3 Study Rationale

The overall rationale of this study was to find out the extent to which the crowd could be relied upon to produce accurate training samples for the classification of remotely sensed imagery through an experiment designed to explore the competencies of the crowd in classifying ecological phenomena. VGI classifications of ground-based images were gathered at multiple hierarchical classification levels using a web application. Interpretations were gathered alongside associated social information and metadata to identify and determine the strength of any relationships which could be seen as significant predictors of annotation accuracy. This section describes the parameters chosen in conducting the experiment and provides the justification behind their usage.

5.3.1 Choice of Research Approach

As this study primarily tested the accuracy of the crowd in using a pre-defined classification system, quantitative analysis was crucial in determining accuracies against a perceived truth. There was also an aspect of inductive research; qualitative feedback regarding the usability of the classification system and web application was of great importance and offered insight into how the system could perhaps be adjusted to obtain more usable annotations.

5.3.2 Selection of Classification System

A plethora of classification systems exist within ecology which are used to identify, characterise and map ecosystems. They enable ecological phenomena to be recorded in a standardised and thorough manner which facilitates interpretations of the interrelationships between resources and processes. They are often hierarchical systems which include the classification of land cover and assigning hierarchies to the geographical extent of biomes and ecosystems. Given the findings from Chapter 4 and the need to engage with participants with multiple degrees of pre-existing knowledge and skills, it was necessary to choose a classification system that was simple to use and applicable to a variety of landscapes. Also the indication that there

was a mismatch between the scale of land units and the scale at which volunteers perceive land units meant that a hierarchical and flexible system was required.

The Joint Nature Conservation Committee (JNCC) Phase 1 Habitat Classification system was chosen for this experiment as it presents a manageable system (with 155 classes over three hierarchical levels) widely used as an in-field classification system. It provides a standardised system to record semi-natural vegetation and natural wildlife habitats at a land parcel scale. It is often used within ecological impact statements which makes it an attractive system for usage with volunteers as the recordings are useful to ecological monitoring (it provides a motivation). The Phase 1 handbook (JNCC, 2010) also suggests that volunteers may be trained to conduct the survey under supervision suggesting that the survey is accessible to some degree to non-professionals.

Habitat Surveys are useful for a variety of reasons:

- Setting targets for habitats in Biodiversity Action Plans
- Identifying appropriate places for habitat restoration
- Responding to planning applications
- Ensuring policy directives are followed e.g. Habitats Directive

VGI has the potential to plug the data gap for regulatory bodies which struggle to maintain the funding for this important source of data.

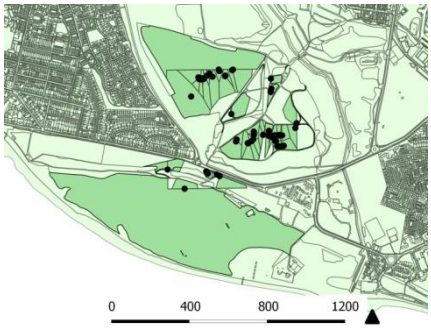
5.3.3 Selection of Study Location

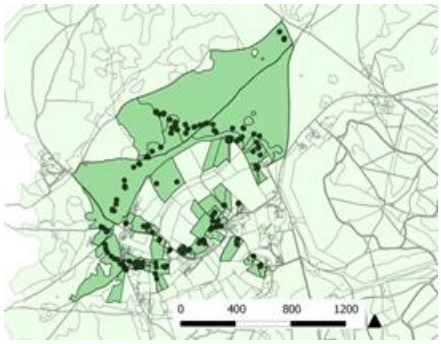
An area with a high diversity of habitats was required to attain sufficient understanding of the variance in crowd interpretation within the constituents of the classification system. Hampshire has a particularly high diversity of flora and fauna with the greatest species diversity of any county in England (Hampshire Biodiversity Partnership, 2013). Habitats in Hampshire include ancient woodlands, heathlands, chalk downland, river valleys and coastal habitats. Three sites (Figure 5.1) were chosen within Hampshire based upon habitat diversity and the availability of ecologist habitat classifications and ground photography provided by Ordnance Survey Research:

- Browndown Nature Reserve
- Titchfield Haven Nature Reserve
- Fritham – New Forest

Figure 5.1 Sites of Study







Three sites (in Hampshire, UK) were chosen based upon ground truth availability and the temporal appropriateness of the available remotely sensed imagery. The sites of interest and associated land parcels are shown. Top left – Titchfield Haven, top right –

5.3.4 Selection of Classification Tasks

150 ground based photographs were chosen from a pool of over 500 photographs taken by and classified by a professional ecologist. The images are shown in Figures 5.3-5.5. A number of factors determined the inclusion of each photograph:

5.3.4.1 Clarity of Photograph

Any photographs that were deemed unclear were removed from the sample set as they were difficult to classify for reasons beyond ecological classification variables such as the mosaicked nature of a habitat. Clarity was judged upon factors influencing the ability to classify, for example, heavy shadowing preventing the determination of habitats and poor or bright lighting within the photograph rendering a habitat abnormally represented.

5.3.4.2 Temporality

As it was intended that the crowd classifications be used as reference information during image classification it was important to only use the images that were temporally relevant for image classification. The date at which information is geo-located has great implications for the utility of crowd generated content; if the interpretations are not made at an appropriate time of year they are of little use as

reference data for aerial imagery which is often flown over May-September when vegetation is ‘leaf-on’ (Gladstone et al. 2012).

5.3.4.3 Location

To ensure an appropriate geographical distribution of samples the X Y coordinates of the photographs were plotted in Arc Map. This enabled the elimination of any photographs of the same habitat type within the same parcel of land according to the Ordnance Survey MasterMap® data for the same locations. In any future study it may be interesting to include multiple images of the same area to understand how the style (e.g. focus) of the photograph influences classification).

5.3.4.4 Habitat Type

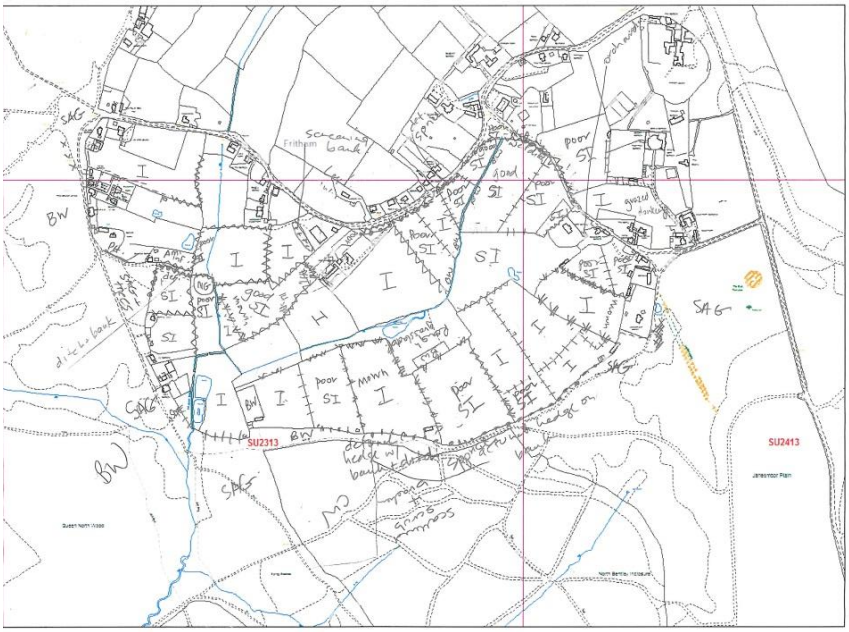
The photographs were sorted into categories according to the predominate habitat type listed by the ecologist. This was to ensure an even distribution of each habitat class within the experiment so that each habitat was represented appropriately and so that inferences could be drawn about how well the crowd could classify each habitat. Where certain classes were under represented (e.g. Open Water) all the available photographs (meeting the listed conditions) were used.

5.3.4.5 Classification Difficulty / Uncertainty

Lastly the photographs were sorted according to their classification difficulty. This was determined by two factors:

- An uncertainty rating provided by the ecologist in field notes (Figure 5.2). If an ecologist was uncertain (UC) of the classification the photograph would be given a difficulty classification category of ‘3’ or difficult. If the ecologist was fairly certain (FC) the photograph would be given a rating of ‘2’ or intermediate and if the ecologist was certain (C) the photograph was given a rating of ‘1’ or easy.

Figure 5.2 Example of Ecologist Field Notes



- Images with ecotones and mosaicked habitats were assigned a greater difficulty level. The ecologist had specified all habitat classes present within the image alongside certainty ratings which permitted categorisation. The conditions by which the images were assigned difficulty levels are listed in Table 5.1

Table 5.1 - Combined Image Difficulty Level Conditions

Condition	Difficulty Rating
C & ≤ 2 habitat types	Easy (1)
C & 3 habitat types or > 1 FC	Intermediate (2)
FC & >3 habitat types or UC	Difficult (3)

Figure 5.3 Classification Images Set 1 (Easy)

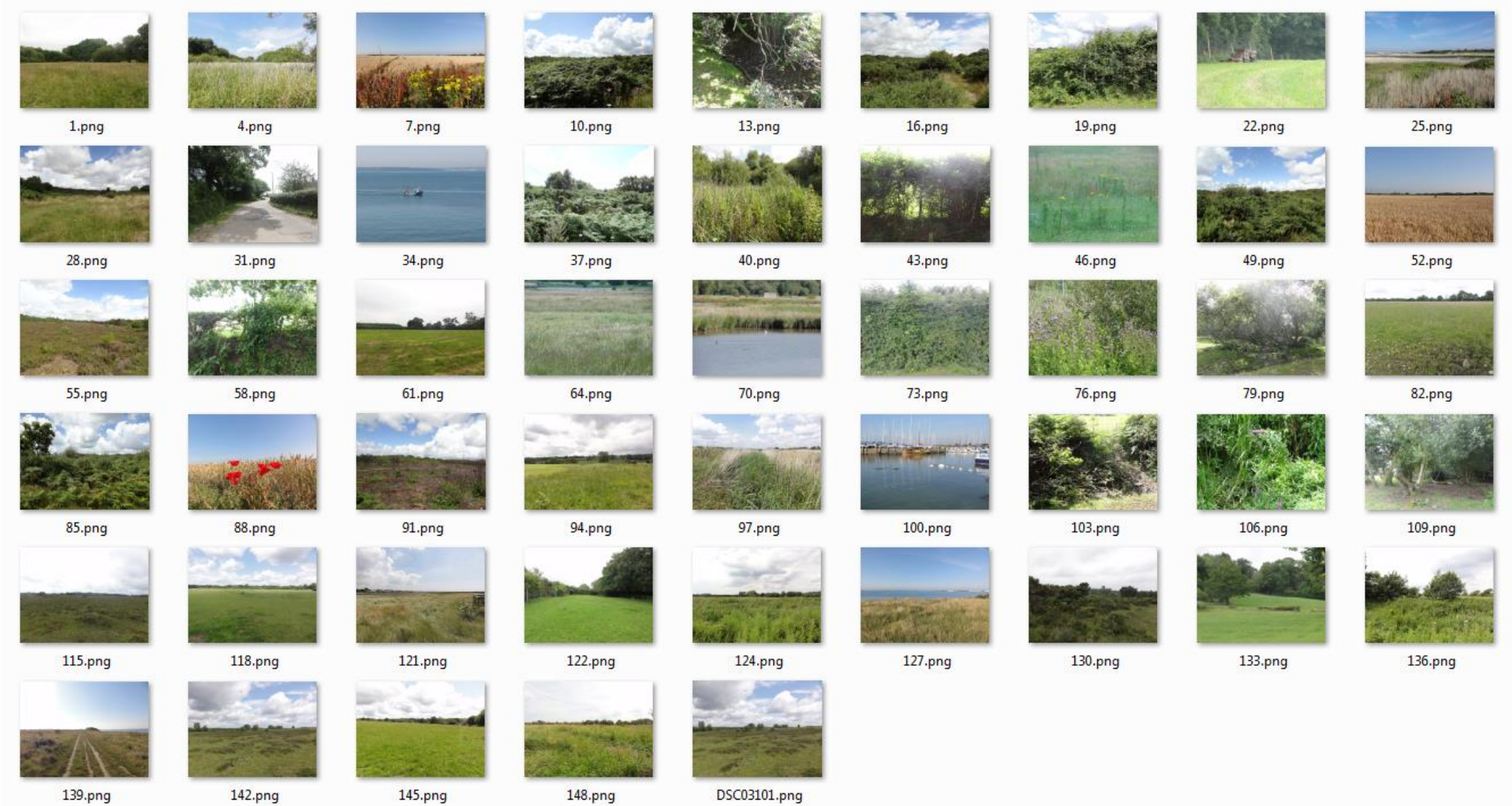


Figure 5.4 Classification Images Set 2 (Intermediate)

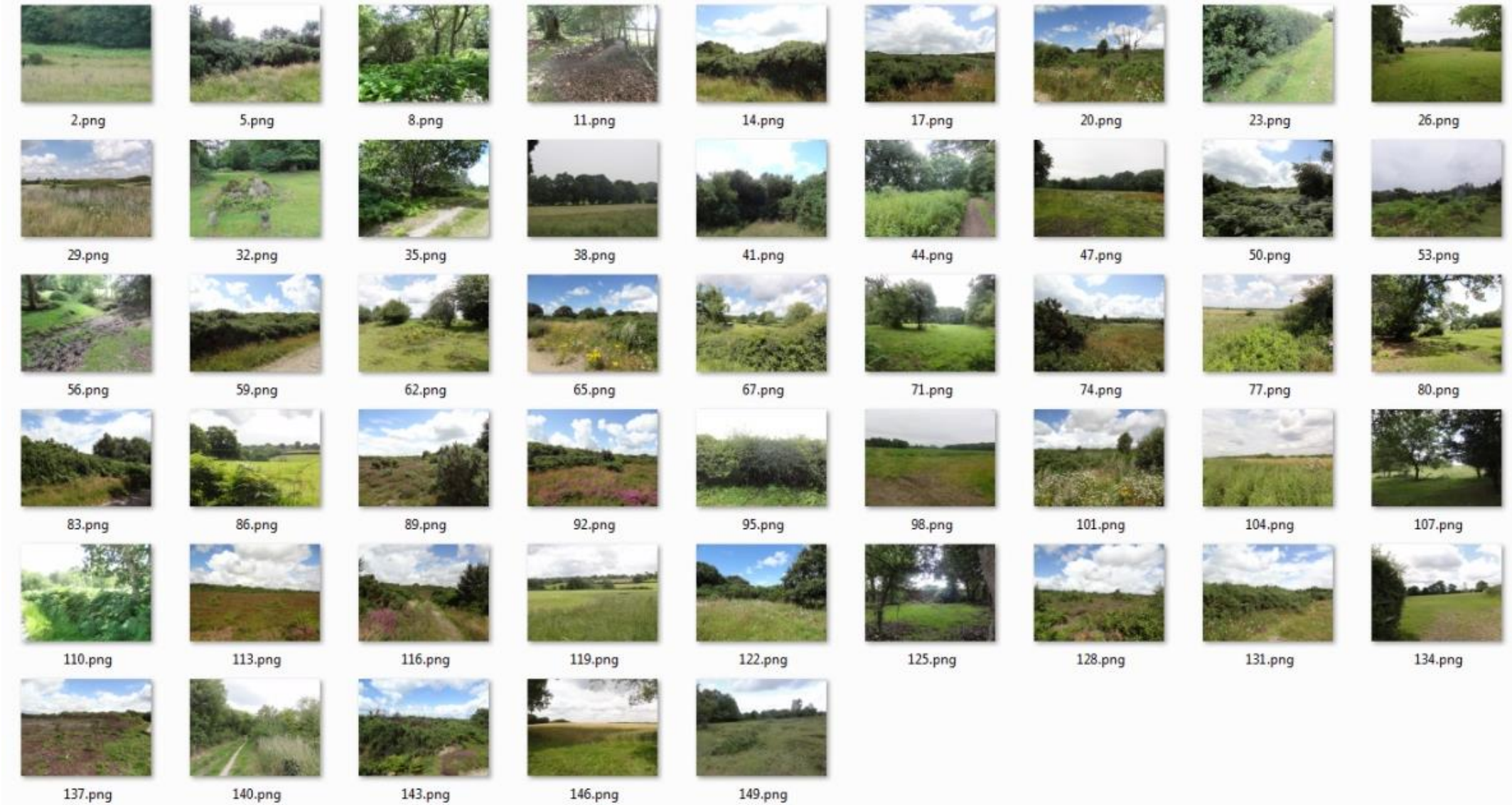
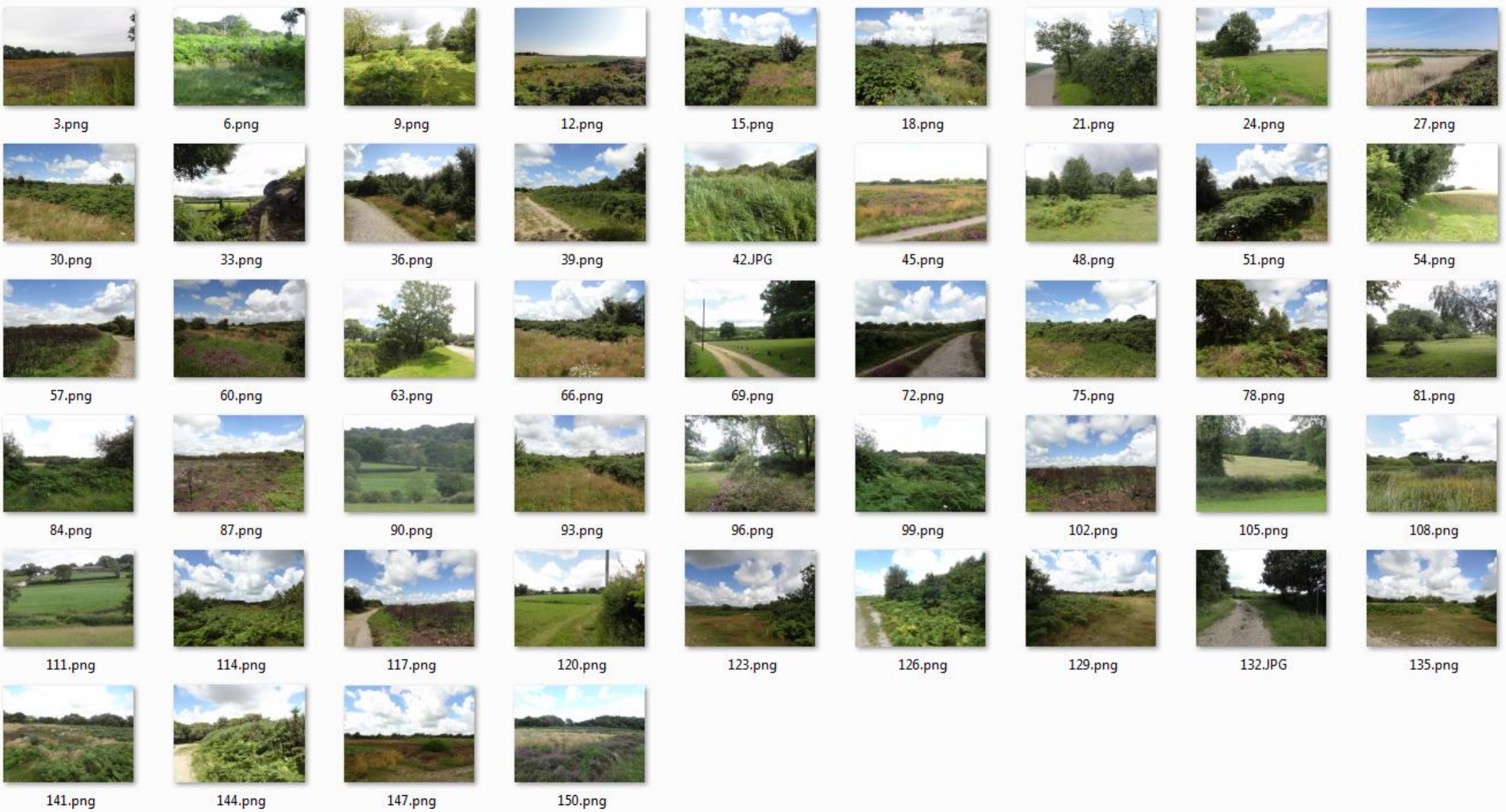


Figure 5.5 Classification Images Set 3 (Difficult)



5.3.5 Selection of Participants

It was important that the crowd represented a common ‘snapshot’ of VGI participants with varied abilities, enthusiasm and engagement with the experiment tasks. In order to attract a selection of volunteers that were representative of the ‘crowd’ it was important to market the experiment as a regular crowdsourcing activity wherever possible. It was hoped that marketing the experiment as one with scientific rationale and as one whose results would be used to inform scientific research would provide an authentic cross section of the crowd from which reliable conclusions could be drawn. There were no restrictions upon recruitment in order to reflect non-uniformity within the conventional crowd.

In the case of OpenStreetMap it takes approximately five users per grid square to map an area to an appropriate quality (Haklay, 2010) with subsequent users providing no significant improvements in quality. In the case of habitat classification there are greater uncertainties with regards to what is correct – it is more subjective than the geo-coordinates of a road signifying that the strength of consensus is of importance. It was thus necessary to attain as many participants as possible for the directed study, which would provide significant groupings to investigate in terms of inter-group accuracy. A GeoWiki study focusing upon land cover classification using satellite imagery recruited 65 volunteers (Foody et al. 2013) which suggests that the obtained unique participant count of 82 volunteers is sufficient for purpose.

Whilst demographic information such as age and gender was recorded it was not deemed necessary to recruit demographic categories proportionally. There is a gender imbalance in the creation of VGI – in OpenStreetMap only 3% of contributors are women (Budhathoki, 2013) and whilst the gender ratio could vary between projects it seemed misrepresentative to actively target a particular demographic.

5.4 Methods

5.4.1 Participant Recruitment

It was necessary to collect crowd observations for the study given the directed nature of the investigation. There was no restriction on participation (as discussed previously) and the experiment was marketed (as much as possible) as a regular crowdsourcing activity to promote the participation of an ‘authentic’ crowd.

5.4.1.1 Recruitment Methods

The Centre for Ecology and Hydrology's 'Citizen Science Guide' (Tweddle et al. 2012) was used as a best practice guide for participant recruitment. The guide, based upon reviews of 234 citizen science projects recommends engaging interested parties through local subject related webpages, emailing newsgroups of associated parties, distributing posters in public places, the use of Twitter and social media to access a broad range of participants and launch events in the local community.

Following ethical approval, participants were recruited as recommended via the following recruitment methods:

- Contact with groups deemed to have potential interest around the subject of habitat mapping and inventory such as the Hampshire Ornithological Society Google group and the Nottingham Rambling Society sought to recruit special interest groups. Email contact with the British Botanical Society resulted in a link to the experiment being placed within their email newsletter for December 2013.

- Engagement with citizen sensing projects focusing upon ecology such as 'Bioblitz' and the Open University's 'iSpot' included communications through Twitter and other social media outlets. Following an email a group for land cover crowdsourcing enthusiasts, Geo-Wiki, posted about the experiment on the social media outlet, Facebook.

- An OpenStreetMap meeting was also attended in Nottingham (November, 2012) where it was possible to explain the crowdsourcing experiment to those interested in the mapping of natural features.

- A Twitter account (@HabitatStudy) was created to promote the experiment, explain the rationale behind it and to speculate on the potential use of the collected data (Figure 5.6).



Figure 5.6 Twitter Account used to Promote the Experiment

- Posters were distributed around the University of Nottingham campus in highly frequented places such as the Students' Union building. Emails were also sent to university students both studying subjects involving ecology modules (via the school of Biology) and those without any background in ecological science.

- Information about the study was also distributed on social media sites in an un-focussed manner – permitting those with no ecological expertise to find out about the study.

5.4.1.2 Participant Groupings

It was hoped that groupings of participants would naturally emerge within the results which could then be used to investigate inter-group accuracy variance. Coote and Rackham (2008) categorised users of VGI but despite the differing motivations behind participating in VGI projects there is no such assessment of who creates VGI. In order to determine any groupings within the participant set, participants were asked whether they had any professional experience in ecology and land cover classification (through careers and special interest groups) and whether they had any hobbies that they considered to be relevant to ecology (such as rambling, botany or bird watching) through a questionnaire. It was then possible to group participants according to any significant commonalities within the participant pool. As in Raddick et al. (2010) the participants were independently grouped by two researchers to ensure the groupings

were as appropriate as possible. It was assumed that those with a keen interest or with a professional background in ecology would respond to the classification tasks with a greater degree of accuracy. To make a distinction between professional ecologists and interested amateurs it was assumed that those with a Master's degree or equivalent and or professional experience of more than one year would constitute a grouping of experienced ecologists. Those educated in ecology to Bachelor's level or an equivalent were identified as 'students' and participants that expressed an interest in ecology or botany were grouped as 'interested amateurs'. Those without any background or interest in ecology and ornithologists were also grouped separately.

When split into two groupings, there were 42 untrained participants and 40 participants trained in ecology (including some students with Masters qualifications). The breakdown for other smaller groupings present within the participant pool are shown in Table 5.2.

Table 5.2 Breakdown of Participants

Group	Total Participants
Ecologists	24
Ornithologists	11
Students	16+
Interested Amateur	14
No Interest	17

5.4.2 Data Collection

5.4.2.1 Task distribution

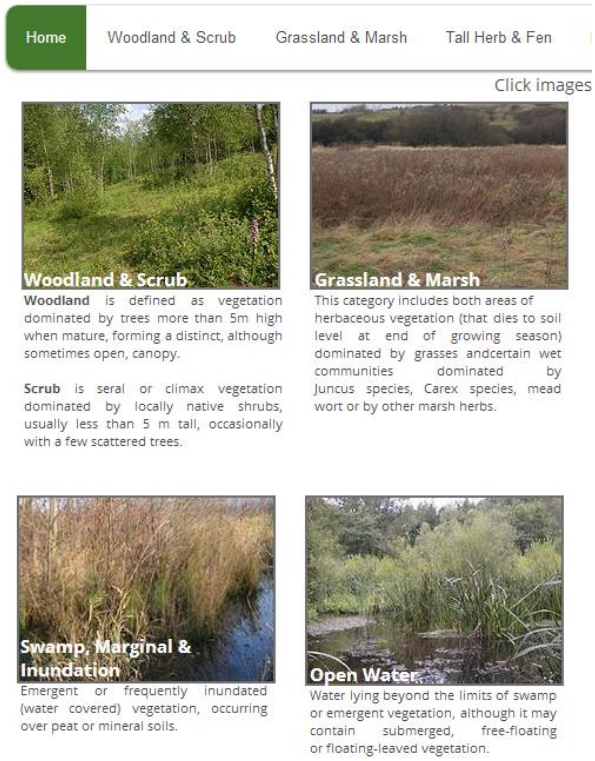
The crowdsourcing tasks were distributed as a simulated citizen science (CS) application which was designed to look as similar to existing CS applications as possible. Qualtrics survey software was used to design the task presenter which used a new page for every classification image. The task presenter was primarily distributed via the web to reach a maximum audience, though was not suitable for mobile use. To mimic a genuine crowdsourcing application, participants could spend as little or as much time classifying habitats as they desired. Whilst this resulted in incomplete data sets it does permit the analysis of the impact of missing cases on the overall

classification accuracy. The study was open and publicised for three months to ensure that a sufficient number of participants could be attained.

5.4.2.2 Training

It is good practice for high skill demanding crowdsourcing initiatives to have some associated training materials or supporting documents to promote accuracy and ensure consistency of semantic interpretation between annotators. A training web page was developed to enable participants to classify each image with some degree of accuracy. Each broad classification option had an example image with associated text describing the habitat and suggesting species which may assist in identifying it. Each broad category could be clicked upon to show all possible sub-class options, again alongside example images and descriptive text. It was hoped that participants would be able to use the website to help inform their classification choices. The text was adapted from the JNCC Phase 1 Handbook (JNCC, 2010) to provide short, comprehensive descriptors which were deemed as accessible to a range of volunteers. A screenshot of the web page is shown in Figure 5.7

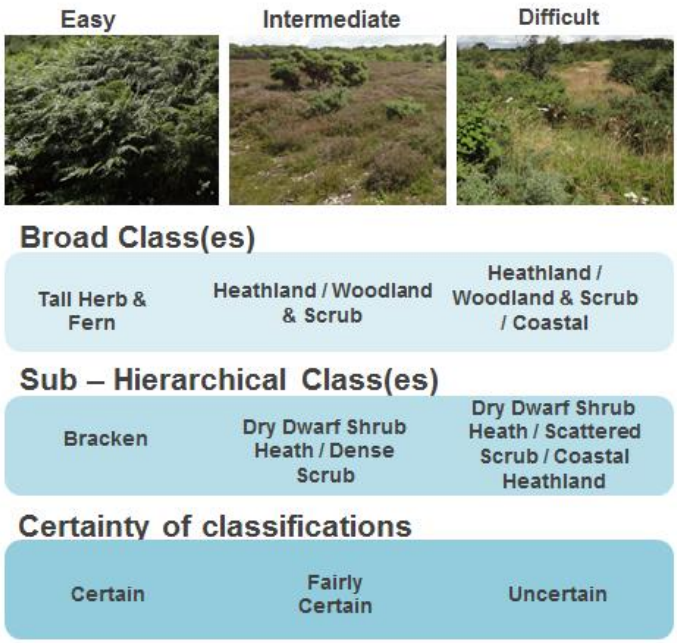
Figure 5.7 User Guidance Web Page



5.4.2.3 Classifications

Participants were asked to select the predominate broad habitat type from within each presented image using a drop-down selection. Participants were then instructed to select an appropriate sub-level classification from a drop-down selection box whose contents depended upon the previous selection. Further sub-classifications could be chosen for instances where more than one habitat type was present within an image. Participants were also able to write any words as free text where they felt that the classification system was insufficient to describe the image presented. Figure 5.8 shows an example of a recorded classification. Images were presented to participants at random to ensure that each photograph had an appropriate sample of classifications but all participants were shown the first six images as a training period which fed-back the appropriate classes (according to the ecologist) following classification.

Figure 5.8 Example Classifications



5.4.2.4 Task metrics

Several pieces of additional information were gathered alongside the habitat classifications in order to look for any significant associations with classification

accuracy beyond the difficulty of the classification task and the skill group to which the contributor was allocated. The collected task metrics are described and explained below:

- ***User certainty -*** The certainty of the participant in classifying each broad and sub-level classification was recorded via a drag and drop Likert scale ranging from 0 (completely uncertain) to 100 (completely certain). Attaining numeric certainty levels permits an understanding of the confidence each participant has in their classification being correct and could thus be used to infer accuracy.

- ***Timing -*** The time taken for each participant to complete each classification task was recorded alongside the number of clicks. This information was used to infer the level attention paid to each task and to suggest where participants found classifications to be easy (taking a short time and few clicks) or difficult (indecision causing a large number of clicks and lengthy classification time).

- ***Screen recording -*** A subset of ten participants were screen recorded with open source software 'CamStudio' whilst they completed the classification task. This involved recording both the mouse movements and the screen itself and enabled an investigation into how participants used the training materials and the identification of cases where participants were particularly confused about how a habitat should be classified.

5.4.2.5 Evaluation

A short qualitative evaluation of the tasks and classification system was performed by each participant to provide valuable insight into any problems which may have affected the accuracy of classifications. Participants were asked whether they used the training materials, to what extent and to provide any other qualitative feedback on the tasks they had completed regarding the following:

- How easy the classification system was to use at a broad level and at the detailed level.
- Whether participants found certain habitat types easier to distinguish than others.
- Any other comments.

5.4.3 Analysis

The crowd classifications were initially assessed using a standard confusion matrix to calculate the overall habitat classification accuracy and the Kappa coefficient (Congalton, 1991). The producer's and user's accuracy were also derived from the confusion matrix in order to evaluate the omission and commission errors for each habitat class (Congalton, 1991). Annotator accuracy and per-term consistency are assessed in terms of the annotator's expertise background.

Cohen's Kappa statistics are used to investigate the reliability of inter-annotator agreement. If p_a = the proportion of observations in agreement and p_ε = the proportion in agreement due to chance, then Cohen's kappa is:

$$\kappa = \frac{p_a - p_\varepsilon}{1 - p_\varepsilon}$$

5.4.4 Adaptation of Woodcock & Gopal Method

Given the differing degrees of correctness within the interpretation of a habitat (owing to hierarchy within the classification system and inherent uncertainties from multi-class ownership), a linguistic scale (Woodcock and Gopal, 2000) is used to assess annotation accuracy in addition to regular confusion matrices. The scale makes use of linguistic constructs to define how accurate a map or classification is; the membership value of each map class at a sample site is estimated which allows experts to acknowledge any heterogeneity in the vegetation cover in the site or ambiguity regarding map classes (Woodcock and Gopal, 2000). Relative similarity scale code scores based upon Woodcock and Gopal's (2000) linguistic scales are thus attributed to each classification based on the hierarchical nature of the Phase 1 Habitat classification system and the identification of habitats that are commonly juxtaposed (Table 5.3).

Table 5.3 Woodcock and Gopal Analysis

Scale Code	Woodcock and Gopal (2000) Linguistic Scale Description	Phase 1 Habitat Classification
5	Absolutely right: No doubt about the match. Perfect.	Correct Sub Class
4	Good answer: Would be happy to find this answer given on the map.	Correct Broad Incorrect Sub Class Related

3	Reasonable or acceptable answer: Maybe not the best possible answer but it is acceptable; this answer does not pose a problem to the user if it is seen on the map.	Correct Broad Incorrect Sub Class un-related
2	Understandable but wrong: Not a good answer. There is something about the site that makes the answer understandable but there is clearly a better answer. This answer is a problem.	Incorrect Broad and Sub Class Broad Class Related
1	Absolutely wrong: This answer is absolutely unacceptable and completely wrong.	Incorrect Broad and Sub Class Completely unrelated

5.5 Results and Analysis

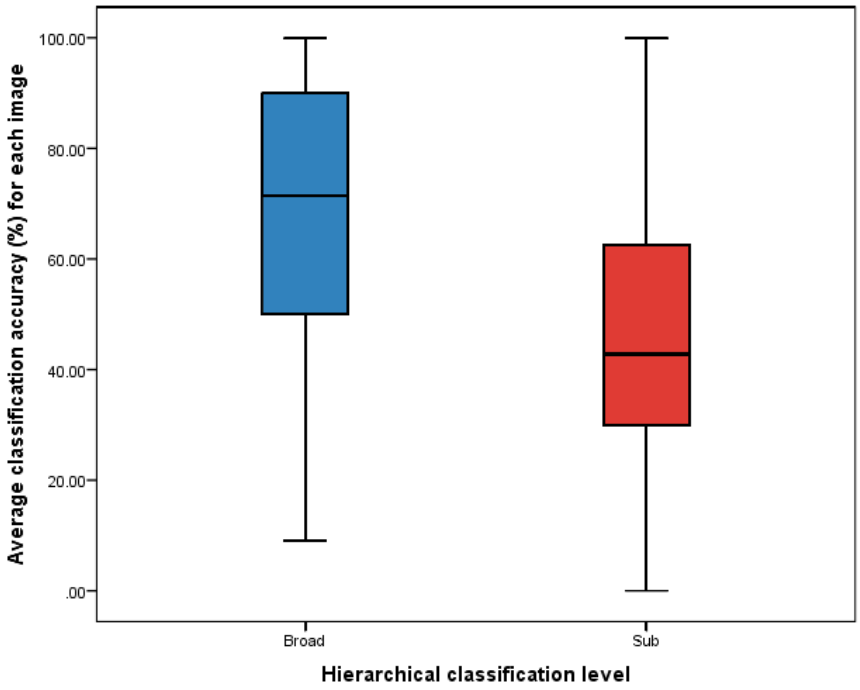
5.5.1 Overall Accuracies

Overall,1,795 crowd habitat classifications of 150 georeferenced, ground based photographs taken within the New Forest have been gathered via a web application. An example of the results (for the 1st experiment image is provided in Appendix 1). Broad and sub-level habitat classifications were sourced from 82 participants ranging from the amateur-enthusiast to professional ecologists who were allocated an experience category following a short questionnaire.

When all of the classifications are taken into consideration the broad level accuracy is 64.40% (overall Cohen's Kappa agreement: 0.55, p=<0.01) and the sub level accuracy is 30.08% (overall Cohen's Kappa agreement: 0.31 p <0.0005). When solely the modal classification for each image is used, the accuracies rise to 77.33% and 47.33% respectively. Given that Remote Sensing classifications are typically deemed to be of suitable accuracy when classification accuracies of 85% are consistently seen across multiple classes (Foody, 2002); the sub level classifications are very poor and the broad level classifications are below adequate.

The boxplot (Figure 5.9) shows the distribution of overall classification accuracies per image at both the broad and sub hierarchical classification levels. The variance in accuracy is very wide at both levels though at the broad level the overall accuracy does not fall below 9.09%. The interquartile range is higher at the broad classification level with an average accuracy across all images of 69.43% at the broad level and 45.14% at the sub classification level.

Figure 5.9 Boxplot to Show the Variance in Classification Accuracy per Image by Classification Level



A Fisher's exact test was used to test for independence of the modal crowd classification and the authoritative ecologist classification as four cells in the contingency table had a value of five or less. The test yielded a p value of 0.98 (p=>0.01) signifying there is no statistically significant difference in classification between the two groups at the broad classification level. At the sub-level a p value of 0.066 was attained when conducting Fisher's exact test also showing no significant difference between the crowd and authoritative interpretations.

5.5.2 Variance between Participant Groups

5.5.2.1 Analysis of Variance between Prominent Groupings

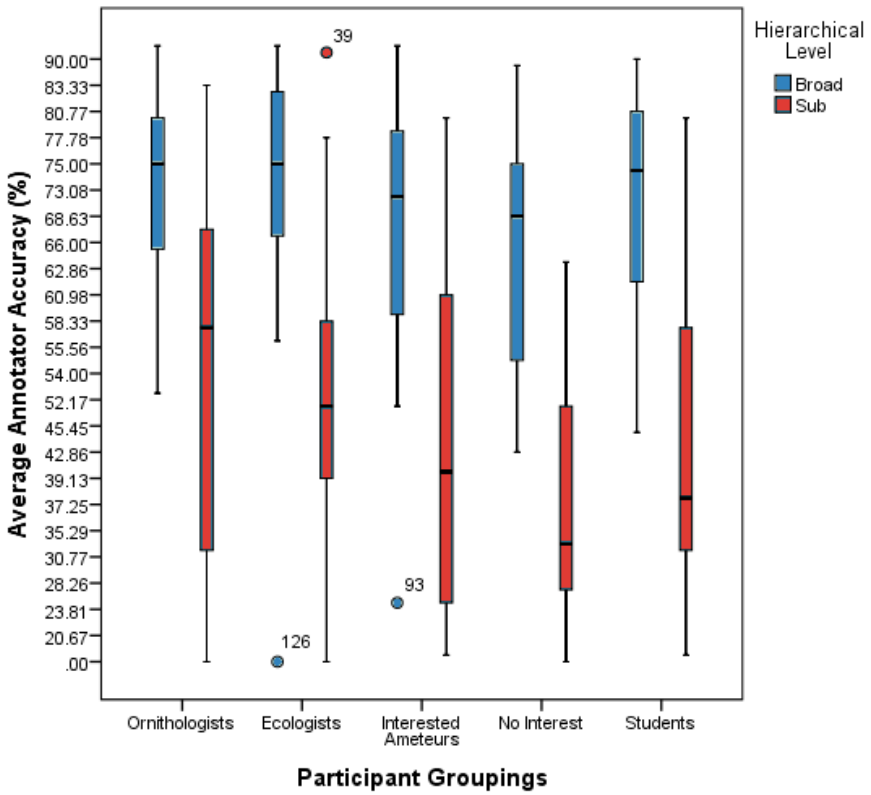
82 participants participated in the web classification study and a short screening questionnaire at the start of the study was used to attribute each participant to a user group depending on their level of previous experience with ecology, interests and hobbies and qualification level. The distribution of participants belonging to each grouping is shown in Figure 5.10.

Figure 5.10 Group membership of Participants



The boxplot (Figure 5.11) shows the distribution of overall classification accuracies per participant at both the broad and sub hierarchical classification levels by participant grouping. The variance in accuracy is similar in all cases though the interquartile ranges and means (Table 5.4) appear to be affected by participant group membership. The mean accuracy and interquartile range is higher at the broad classification level in all cases. At both levels the interquartile range is greater for the sub classification level suggesting that there is a greater level of variability and uncertainty when classifying at this level.

Figure 5.11 Boxplot to Show the Variance in Classification Accuracy per Annotator by Classification Level and Participant Grouping



The mean participant annotation accuracy is lowest for those with no interest in ecology or botany at the broad and sub hierarchical levels (66.18% and 36.15% respectively). Ornithologists were seen to be the most accurate annotators with an accuracy of 73.93% at the broad-level and a sub-level accuracy of 50.17%. Some outliers are present within the datasets; however, these will be preserved given that they reflect the non-uniformity of crowdsourced information.

Table 5.4 Mean Accuracy per Annotator Across all Images

Participant Group	Broad Accuracy (%)	Standard Deviation	Sub Accuracy (%)	Standard Deviation
(1) Ornithologists	73.93	12.92	50.17	25.10
(2) Ecologists	73.36	11.52	48.10	18.43
(3) Interested A's	69.98	15.64	43.25	21.36
(4) No Interest	66.18	14.07	36.15	16.09
(5) Students	71.77	14.57	45.18	20.3

Overall Inter-group broad scale accuracies vary by up to 12.91% with ecologists & ornithologists as the most accurate annotators. There is significant variance between ecologists, ornithologists and the group of participants with no background or interest in ecology (observed Z values of 2.55 and 2.03 respectively) showing that filtering crowd annotations by group membership could improve the accuracy of crowd derived classifications (Table 5.5).

Table 5.5: Cohen's Kappa Analysis

Group	Kappa	Variance	Z statistic	Overall Accuracy
Interested Amateur	.592	0.0035	10.01	65.63%
Ecologists	.652	0.0024	13.31	74.45%
No Interest	.468	0.0028	8.84	61.54%
Ornithologists	.613	0.0023	12.78	70.00%
Students	.600	0.0025	12.00	68.79%

5.5.2.2 Significance of Variance between Participant Groups

Significant normality levels determined by the Sharpio-Wilk test of normality denoted that a one-way between-groups analysis of variance (ANOVA) was used to determine any statistically significant variability between participant groupings for each factor (or hierarchical level). Table 5.6 shows the results of the ANOVA which show significant

variance between the overall participant accuracy across all images between participant groupings.

Table 5.6: ANOVA – Participant Groupings

Participant Accuracy – average accuracy per annotator across all images

Hierarchical Level	ANOVA F Value	Significance
Broad	1.467	.221
Sub	1.422	.235

Participant Accuracy – average annotator accuracy per image

Broad	4.49	.01
Sub	13.573	.00

There is a significant effect for participant groupings when the ANOVA was conducted to compare the groupings performance on a per-image basis. At the broad-level there is a significant difference in average accuracy at the $p < 0.5$ level for the 5 participant groupings $F (4, 680) = 4.49$, $p = 0.01$. The Eta squared value (0.026) shows that the effect size of participant grouping on the accuracy of classification is quite low at the broad-level (Cohen, 1988). Post-hoc comparisons using the Tukey test (Table 5.7) indicated that there are significant differences between groups 1 ($M = 75.43$, $SD = 31.98$) and 4 ($M = 59.22$, $SD = 37$) and groups 2 ($M = 73.72$, $SD = 36.34$) and 4.

The sub-level ANOVA also presented significant differences in accuracy between participant groupings $F (4, 676) = 13.573$, $p = 0.00$ with a medium effect of grouping on participant accuracy (Eta squared = 0.074). The Tukey test (Table 5.7) showed significant variances in the mean accuracy between many participant groupings at the $p < 0.05$ level.

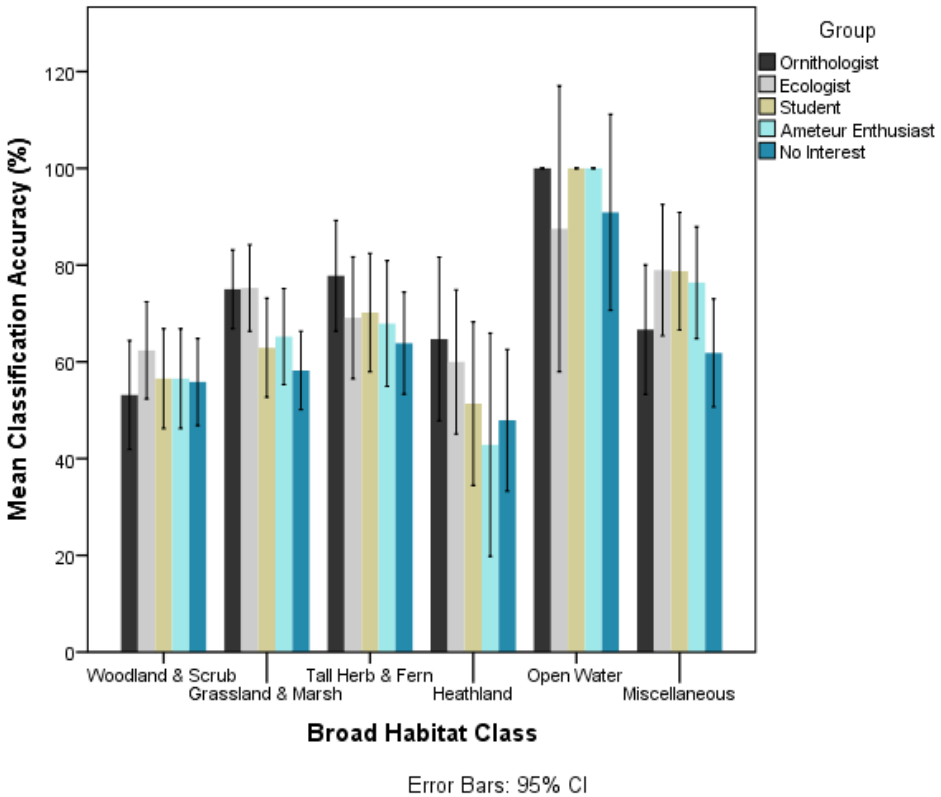
Table 5.7 Post Hoc Test – Tukey (Broad & Sub Levels)

Participant Group	Mean Difference	Std. Error	Significance
1 & 4 Broad	16.21	4.12	.001
2 & 4 Broad	14.5	4.29	.007
1 & 4 Sub	26.76	4.38	.000
2 & 3 Sub	14.49	4.92	.027
2 & 4 Sub	29.02	4.49	.000
2 & 5 Sub	13.37	4.56	.029
3 & 4 Sub	14.52	4.8	.022
4 & 5 Sub	-15.64	4.44	.004

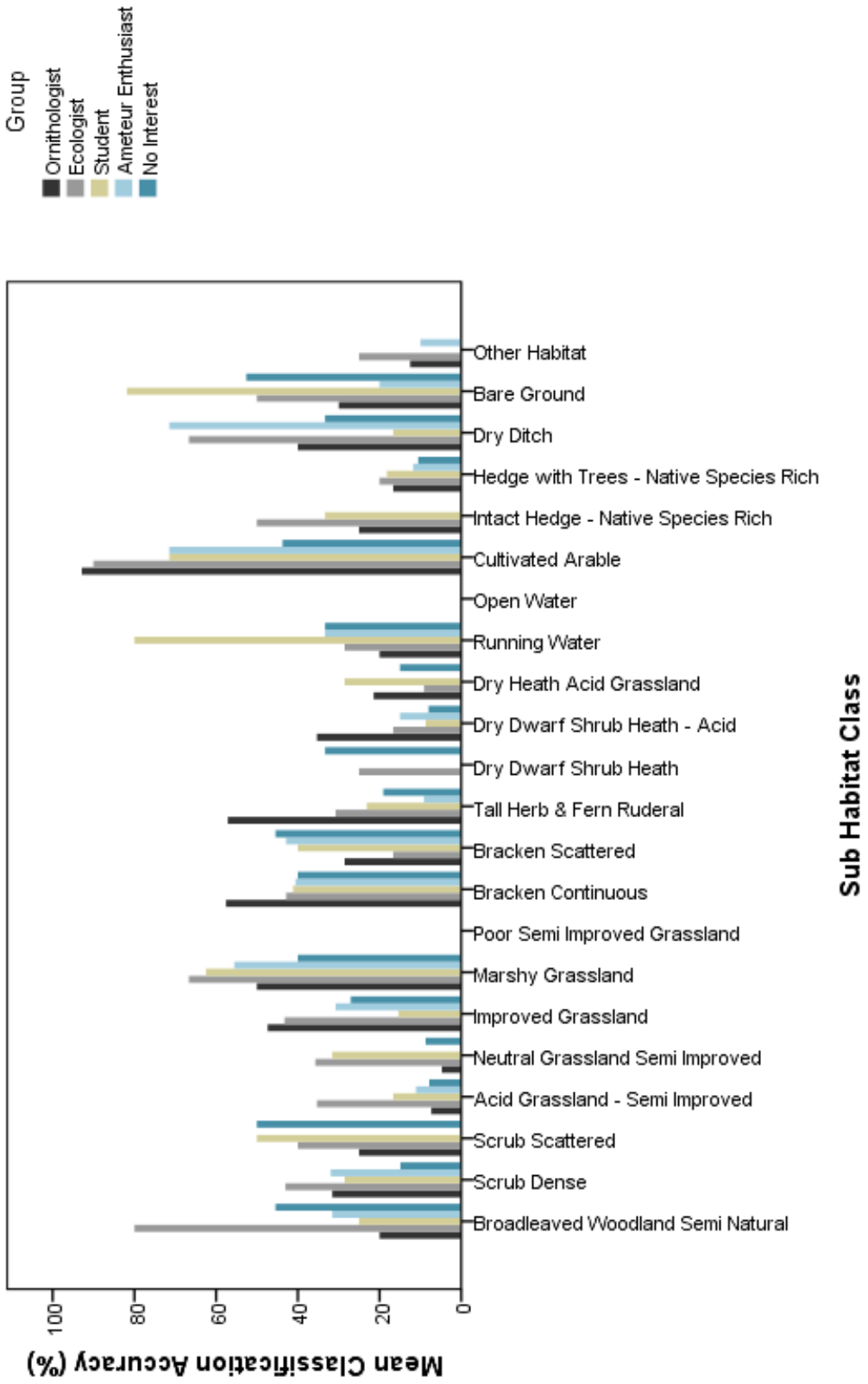
5.5.2.3 Habitat Classification Accuracy by Participant Grouping

Figure 5.12 and Figure 5.13 show how the classification accuracy varies between participant groupings at the Broad and Sub hierarchical levels. There is no statistically significant variance between any of the groups at the broad classification level. Accuracy at the sub level has greater inter-group variation than at the broad classification level. Ecologists are significantly better than any other group with Broad Leaved Woodland and Acidic & Marshy Grassland varieties.

Figure 5.12: Graph to Show Broad Habitat Classification Accuracy by Participant Grouping



Figur **pant**


Group
Ornithologist
Ecologist
Student
Ameteur Enthusiast
No Interest
Other Habitat
Bare Ground
Dry Ditch
Hedge with Trees - Native Species Rich
Intact Hedge - Native Species Rich
Cultivated Arable
Open Water
Running Water
Dry Heath Acid Grassland
Dry Dwarf Shrub Heath - Acid
Dry Dwarf Shrub Heath
Tall Herb & Fern Ruderal
Bracken Scattered
Bracken Continuous
Poor Semi Improved Grassland
Marshy Grassland
Improved Grassland
Neutral Grassland Semi Improved
Acid Grassland - Semi Improved
Scrub Scattered
Scrub Dense
Broadleaved Woodland Semi Natural
Sub Habitat Class
Mean Classification Accuracy (%)
0
20
40
60
80
100

5.5.3 Grouping by Level of Training

An independent samples t-test was conducted to compare the difference in annotation accuracy between participants trained in ecology and those without any professional ecology experience. There was no significant difference at the broad level between trained participants (M = 68.03, SD =36.23) and untrained annotators (M = 71.65, SD = 37.25) at the broad level; t (-1.27, p=.205, two-tailed). The magnitude of differences in the means is very low with an Eta squared value of 0.002.

In contrast there is a very small statistically significant difference (Eta squared = 0.01) in interpretation accuracy at the sub level between participants trained in the field of ecology (M = 42.96, SD =37.61) and untrained participants (M = 51.66, SD =41.11); t =-2.803 p =0.005, two tailed.

5.5.4 Variance in Inter-Class Accuracies

Broad scale inter-class overall accuracy varies by up to 39.89% (Table 5.8) suggesting that the crowd could be better suited to identifying certain habitat types. At the broad classification level 'Woodland and Scrub' and 'Heathland' are the most commonly misclassified habitat type (43.07% and 45.95% of classifications were incorrect respectively). The crowd were most successful at identifying the 'Open Water' habitat followed by the 'Tall Herb and Fern' category. Analysis of the sub level accuracies shows further disparity in inter-class accuracy with certain classes such as arable land being consistently identified by the crowd (72.06% accuracy), whereas for more indiscriminate classes such as acid and neutral grassland varieties the crowd classified the images inadequately (14.64% and 15.38% respectively).

Table 5.8: Broad Class Accuracy

Group	User's Accuracy	Producer's Accuracy	Std. Dev.	Overall Accuracy
Woodland & Scrub	61.79%	59.58%	49.55	56.93%
Grassland & Marsh	73.31%	68.79%	47.16	66.73%
Tall Herb & Fern	66.45%	69.10%	46.24	69.21%
Heathland	65.27%	57.98%	49.97	54.05%
Open Water	93.94%	93.94%	24.23	93.94%
Miscellaneous	65.29%	71.43%	45.38	71.16%

5.5.4.1: Confusion Matrix Analysis

The confusion matrix (Table 5.9) enables the determination of which classes the contributors are selecting in lieu of the 'ideal' ecologist labelled class. At the broad hierarchical level the producer's accuracy shows that Woodland & Scrub and Heathland are the least accurately labelled classes. Woodland & Scrub is most commonly classified as Grassland and Marsh with the 'Miscellaneous' and Tall Herb and Fern categories also causing confusion. Heathland is most commonly misclassified as Woodland & Scrub with annotators probably mistaking areas of heath for areas of scrubland.

Table 5.9: Confusion Matrix for Broad Level Classifications

		Reference Data (Ecologist)										Row Total	User's Accuracy
		A	B	C	D	E	F	G	H	I	J		
Classified Data (Crowd)	A	283	76	39	25	0	0	0	0	0	35	458	61.79%
	B	34	368	32	49	0	0	0	0	0	19	502	73.31%
	C	75	20	208	4	0	0	0	0	0	6	313	66.45%
	D	35	2	12	109	0	0	0	0	0	9	167	65.27%
	E	0	1	1	0	0	0	0	0	0	1	3	0.00%
	F	0	12	4	0	0	0	0	0	0	0	16	0.00%
	G	0	2	0	0	0	0	31	0	0	0	33	93.94%
	H	1	2	0	0	0	0	2	0	0	0	5	0.00%
	I	0	3	0	1	0	0	0	0	0	6	10	0.00%
	J	47	49	5	0	0	0	0	0	0	190	291	65.29%
Column Total		475	535	301	188	0	0	33	0	0	266	1798	
Producer's Accuracy		59.58%	68.79%	69.10%	57.98%			93.94%			71.43%		

Habitat Categories

A = Woodland & Scrub **B= Grassland & Marsh** **C= Tall Herb & Fern**

D = Heathland **E = Mire** **F = Swamp, Marginal & Inundation**

G = Open Water **H = Coastland** **I = Exposure & Waste**

J = Miscellaneous

Overall Accuracy: 66.13%

Kappa Coefficient: .57 p <0.0005 (moderate agreement)

Table 5.10: Confusion Matrix for Sub Hierarchical Level

Classified Data (Crowd)	Reference Data (Ecologist)																						Total	Uaer's Accuracy
		A1	A2	A9	A10	B3	B6	B10	B11	B12	C2	C3	C6	C8	D1	D6	G2	J1	J7	J12	J15	J18	J19		
	A1	0	8	5	0	0	1	2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	19	0.00%
	A2	0	36	9	0	3	0	4	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	56	64.29%
	A3	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0.00%
	A4	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.00%
	A5	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.00%
	A6	0	1	18	0	2	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	25	0.00%
	A7	0	2	6	0	0	2	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	15	0.00%
	A8	0	2	16	0	1	2	2	0	0	2	2	1	0	0	3	0	0	0	1	3	3	0	38	0.00%
	A9	0	1	117	0	10	11	0	0	0	7	0	7	0	0	6	0	0	2	2	0	1	1	165	70.91%
	A10	0	1	18	6	2	8	1	0	0	4	1	2	0	2	5	0	1	0	0	0	11	2	64	9.38%
	A11	0	3	5	0	5	0	7	0	1	0	1	0	0	0	1	0	0	0	0	1	1	0	25	0.00%
	A12	0	3	2	0	6	0	4	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	19	0.00%
	A13	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2	0.00%
	A14	0	1	0	0	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	6	0.00%
	A15	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.00%
	A16	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.00%
	A17	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0.00%
	B1	0	0	0	0	5	1	0	2	0	1	0	0	0	0	3	0	0	0	0	0	0	0	12	0.00%
	B2	0	0	3	0	7	4	1	1	0	1	0	1	0	1	2	0	0	0	0	0	0	0	21	0.00%
	B3	0	2	3	0	43	12	7	2	1	2	0	3	0	1	5	0	1	0	3	0	0	0	85	50.59%
	B4	0	1	0	0	8	7	14	5	1	4	0	3	0	1	7	0	1	0	0	0	0	1	53	0.00%
	B5	0	0	2	0	10	2	1	1	0	2	0	2	0	0	5	0	0	0	0	0	0	0	25	0.00%
	B6	0	0	9	0	25	21	11	1	1	2	0	1	0	0	8	0	0	0	0	1	0	0	80	26.25%
	B7	0	0	0	0	3	1	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0.00%
	B8	0	0	0	0	0	1	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	4	0.00%
	B9	0	0	1	0	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0.00%
	B10	0	0	6	0	20	2	79	2	0	2	0	0	0	1	1	0	9	0	6	1	0	2	131	60.31%
	B11	0	0	0	0	8	1	4	27	0	7	0	2	0	0	1	0	1	0	0	0	0	0	51	52.94%
	B12	0	1	0	2	4	3	2	0	0	1	0	0	0	2	8	0	0	0	0	0	1	0	24	0.00%
	C1	0	2	8	0	1	1	0	0	0	6	2	1	0	0	0	0	0	0	2	0	0	0	23	0.00%
	C2	0	1	29	0	1	1	0	0	0	88	8	0	0	0	0	0	0	1	0	0	0	0	129	68.22%
	C3	0	1	2	2	2	2	1	0	0	18	15	0	0	0	0	0	0	0	0	0	0	1	44	34.09%
	C4	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.00%
	C5	0	0	13	2	4	3	0	0	0	23	2	16	0	1	0	0	0	0	0	0	0	1	65	0.00%
	C6	0	0	5	0	2	1	0	1	0	5	0	21	0	0	0	0	0	0	0	0	0	0	35	60.00%
	C7	0	0	3	0	1	0	0	0	0	2	0	5	0	3	0	0	0	0	1	0	0	0	15	0.00%
	C8	0	0	17	0	0	1	0	0	0	5	0	1	2	11	13	0	0	0	0	0	3	0	53	3.77%
	D1	0	0	5	0	0	0	0	0	0	0	0	0	0	15	18	0	0	0	0	0	1	0	39	38.46%
	D2	0	0	1	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	4	0.00%
	D3	0	0	5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	6	0.00%
	D4	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	4	0	8	0.00%
	D5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.00%
	D6	0	0	4	0	0	1	0	0	0	1	1	1	4	4	31	0	0	0	0	0	0	0	47	65.96%
	D7	0	0	2	0	0	0	0	0	0	1	0	0	1	0	2	0	0	0	0	0	0	0	6	0.00%
	E1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0.00%
	E2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0.00%
	E3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0.00%
	F1	0	0	0	0	0	0	3	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	7	0.00%
	F2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.00%
	F3	0	0	0	0	1	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0.00%
	F4	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3	0.00%
	G1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	14	0.00%
	G2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	14	100.00%
	H1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0.00%
	H2	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0.00%
	H3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.00%
	I1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	4	0.00%
	I2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.00%
	I3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	4	0.00%
	I4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0.00%
	J1	0	0	2	0	15	0	3	0	0	0	0	1	0	0	0	0	52	0	0	0	0	0	73	71.23%

	J2	0	0	1	0	0	2	4	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	9	**0.00%**
	J3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	4	**0.00%**
	J4	0	0	1	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	5	**0.00%**
	J5	0	2	7	0	1	0	0	0	0	1	0	0	0	0	0	0	0	2	5	2	0	1	21	**0.00%**
	J6	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	10	**0.00%**
	J7	0	1	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**5**	11	0	0	0	24	**20.83%**
	J8	0	2	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	7	0	0	0	18	**0.00%**
	J9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	4	**0.00%**
	J10	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	**0.00%**
	J11	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	1	8	**0.00%**
	J12	0	0	2	0	0	0	4	0	0	0	0	0	0	0	0	0	0	3	**14**	0	0	0	23	**60.87%**
	J13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	3	**0.00%**
	J14	0	1	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	4	11	**0.00%**
	J15	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**12**	0	1	16	**75.00%**
	J16	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	10	**0.00%**
	J17	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	**0.00%**
	J18	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	**24**	4	32	**75.00%**
	J19	0	1	1	0	1	0	4	0	0	0	0	0	0	0	0	0	0	1	0	1	2	**3**	14	**21.43%**
Total		0	86	357	16	204	97	176	48	10	196	41	72	10	44	125	28	68	18	69	27	54	38	1789	
Producer's Accuracy		**0.00%**	**41.86%**	**32.77%**	**37.50%**	**21.08%**	**21.65%**	**44.89%**	**56.25%**	**0.00%**	**44.90%**	**36.59%**	**29.17%**	**20.00%**	**34.09%**	**24.80%**	**50.00%**	**76.47%**	**27.78%**	**20.29%**	**44.44%**	**44.44%**	**7.89%**		

Classification Codes (Sub Hierarchical Level):

A1 Woodland
A2 Broadleaved woodland - semi-natural
A3 Broadleaved woodland – plantation
A4 Coniferous woodland
A5 Coniferous woodland - semi-natural
A6 Mixed woodland
A7 Mixed woodland - semi-natural
A8 Scrub
A9 Scrub - dense/continuous
A10 Scrub – scattered
A11 Parkland/scattered trees
A12 Broadleaved Parkland/scattered trees
A13 Coniferous Parkland/scattered trees
A14 Mixed Parkland/scattered trees
A15 Recently felled woodland
A16 Coniferous woodland - recently felled
A17 Mixed woodland - recently felled
B1 Acid grassland
B2 Acid grassland - unimproved
B3 Acid grassland - semi-improved
B4 Neutral grassland
B5 Neutral grassland - unimproved
B6 Neutral grassland - semi-improved
B7 Calcareous grassland
B8 Calcareous grassland - unimproved
B9 Calcareous grassland - semi-improved
B10 Improved grassland
B11 Marsh/marshy grassland
B12 Poor semi-improved grassland
C1 Tall Herb & Fern
C2 Bracken
C3 Bracken – continuous
C4 Bracken - scattered
C5 Upland species-rich ledges
C6 Other tall herb and fern
C7 Other tall herb and fern - ruderal
C8 Other tall herb and fern - non ruderal
D1 Dry dwarf shrub heath
D2 Dry dwarf shrub heath - acid
D3 Dry dwarf shrub heath - basic
D4 Wet dwarf shrub heath
D5 Lichen/bryophyte heath
D6 Dry heath/acid grassland
D7 Wet heath/acid grassland
E1 Bog
E2 Sphagnum bog
E3 Blanket sphagnum bog
F1 Swamp
F2 Marginal & Inundation
F3 Marginal vegetation
F4 Inundation vegetation
G1 Standing Water
G2 Running Water
H1 Intertidal
H2 Coastal Grassland
H3 Coastal Heathland
I1 Natural rock exposure and waste
I2 Other rock exposure
I3 Artificial rock exposure and waste
I4 Spoil
J1Cultivated/disturbed land - arable
J2 Cultivated/disturbed land - amenity grassland
J3 Cultivated/disturbed land - ephemeral/short perennial
J4 Introduced Shrub
J5 Boundaries
J6 Hedges
J7 Intact hedge - native species-rich
J8 Intact hedge - native species-poor
J9 Defunct hedge - native species-rich
J10 Defunct hedge - native species-poor
J11Hedge with trees
J12 Hedge with trees - native species-rich
J13 Hedge with trees - native species-poor
J14 Fence
J15 Dry Ditch
J16 Built-up areas
J17 Caravan Site
J18 Bare Ground
J19 Other Habitat

Overall Accuracy: 34.94%

Kappa Coefficient: .31 p <0.0005 (low agreement)

Analysis of the sub level accuracies, using the sub class confusion matrix (Table 5.10), show how people have classified these more indiscriminate classes. Misclassified grassland is most commonly classified under a different variety of grassland e.g. Acidic Grassland is often labelled as a variety of Neutral Grassland (43 cases), Improved Grassland (20 cases), Calcareous varieties (8 cases) and Marshy Grassland (8 cases) suggesting that the various types are too difficult to identify from a photograph with the available materials. Dry Heathland has a very poor producer's accuracy (24.8%) with this variety of heathland often labelled as 'Other Tall Herb & Fern' or as 'Scrub'. Areas the ecologist classified as 'other habitat' were mostly attributed to other miscellaneous categories giving a low producer's accuracy of 7.89%. 'Other Tall Herb & Fern' was also miscategorised as 'Upland Species Rich Ledges' (16 cases) and Dense / Continuous Scrub (7 cases). Dense / Continuous Scrub is a category more understandably miscategorised with 16 counts of 'Scrub' and 18 counts of 'Scattered Scrub'. It is also most commonly confused with 'Bracken' (29 counts).

5.5.5 Correlations with Ancillary Information

5.5.5.1 Insights from Screen Recording

Ten of the participants were screen recorded during the classification study in order to understand the extent to which the participants used the available training materials. All of the recorded participants were randomly selected from the Student group so it could be assumed that all of the participants had a similar extent of pre-existing habitat knowledge. The video was analysed to measure the time each participant spent using the training guide and the accuracy of the annotations was looked at in relation to the time spent on each classification. The results are presented in Table 5.11

Table 5.11 Analysis of Classification Guide Attentiveness

Participant	Broad Accuracy (%)	Sub Accuracy (%)	Average Time Per Classification
6	73.07	46.15	147.11
11	72.72	45.45	136.78
12	52.17	30.43	72.06
29	56.41	35.89	86.78
41	47.06	39.29	137.50
42	88.88	33.33	75.29
58	80.00	40.00	56.91
64	66.66	33.33	101.89
77	73.08	34.62	49.83
80	80.77	65.38	118.39

A slight positive correlation was found between the amount of time looked at the classification resources and the accuracy of classifications at both the broad and sub hierarchical levels. The broad level accuracy has a statistically insignificant correlation of 1.86 (P=0.605) and the sub level accuracy has a slightly stronger insignificant correlation of 5.68 (P=0.086). The sample size is insufficient to certify the effects of use of the training manual but suggests that use of such materials could result in more accurate classifications. It is impossible to remove the effects of varying levels of familiarity with habitats or existing ecological knowledge so it is important to look at the qualitative responses regarding use of the classification guidance materials.

Following the study the participants were asked the extent to which they found the training materials useful (results presented in Tables 5.12 and 5.13). Most students (N=8) said that the guide was a helpful resource that enhanced the quality of their submissions and most students (N=9) thought that without the guide their classifications would have been different. Four participants noted that the guide could have been more usable given it was very difficult for them to continually flick between the guide and classification tasks however all of the participants (N=10) stated that the resource itself was easy to navigate. All students (N=10) noted that the images within the guide made it much easier to classify the image sets and approximately half of the participants (N=6) found the textual descriptions to be helpful. A few participants (N=3) thought that the text within the descriptions was still too inaccessible for the lay participant.

Table 5.12 Liked Aspects of Help Resources

Content	Presentation	Navigation
• The textual content was helpful (6) • The images used made the text easier to understand (10) • The content helped to enhance the quality of my submissions (8)	• Very clear formatting (8) • Colour scheme was aesthetically pleasing (6)	• Very easy to move between different definitions (10) • The use of clickable images makes it easy to navigate (7)

Table 5.13 Drawbacks of Help Resources

Content	Presentation	Navigation
• The text within the descriptions is inaccessible (3) • There were too many categories to choose from (4)	• The layout was overwhelming for the Grassland category as there were too many classification options (2)	• It was hard to keep going between the guide and classification tasks (4)

5.5.5.2 Click Count and Classification Accuracy

Small though statistically significant negative correlations (p= <0.005) were found with both the number of clicks per classification and time taken to perform each classification at both the broad hierarchical level (a correlation of -2.44: clicks, -2.45: time) and sub level (-3.62: clicks, -1.81: time) as shown in Table 5.14. A greater number of clicks denotes that more time was spent looking at the associated classification definitions webpage. The negative correlations of both the time taken and number of clicks per classification (Figure 5.14) with accuracy are perhaps the result of hesitation and uncertainty indicating a complex classification site.

Figure 5.14: Scatter Chart to Show the Mean Count of Clicks in Relation to the Classification Accuracy per Image at both the Broad and Sub Hierarchical Classification Levels

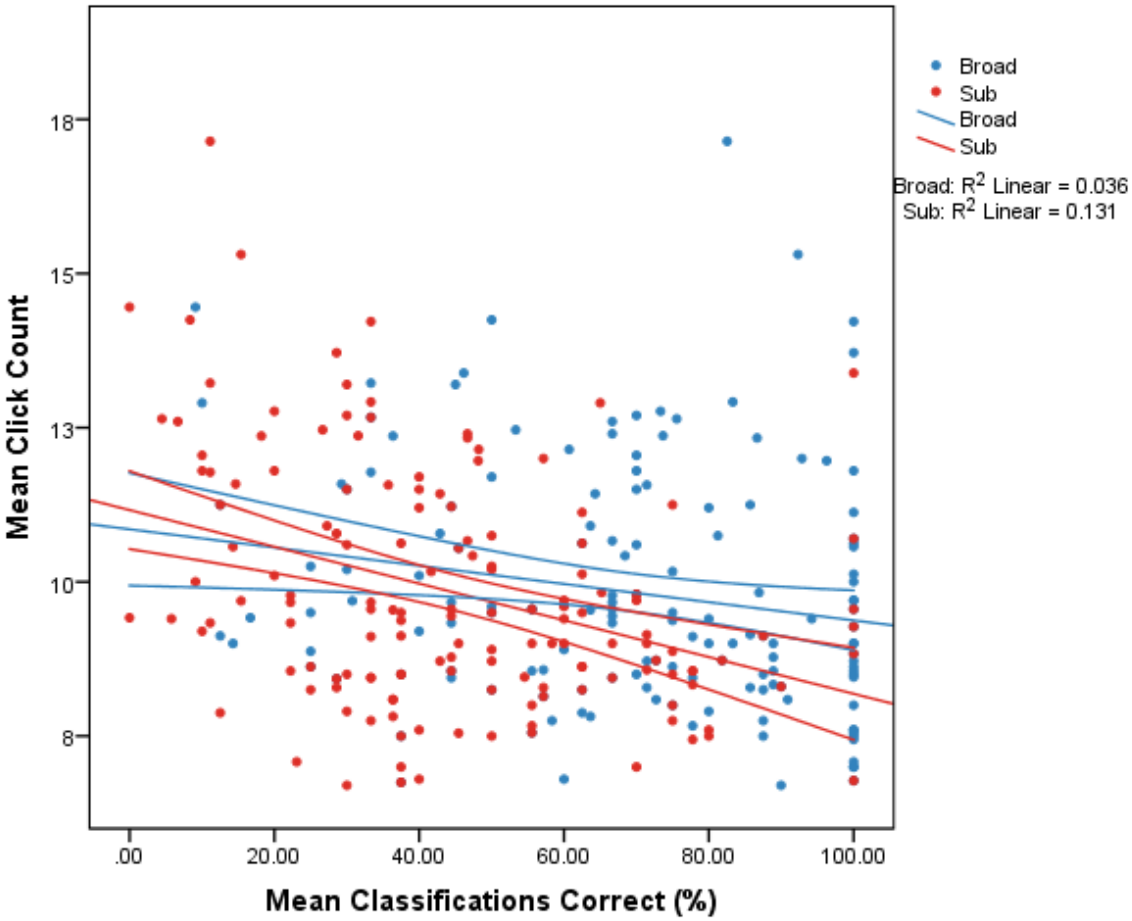


Table 5.14: Pearson's Correlation Coefficient Analysis

Hierarchical Level	Pearson Correlation	Significance
Broad	-2.44	.0001
Sub	-3.62	.0001

5.5.5.3 User Certainty and Classification Accuracy

Pearson's correlation coefficient tests show a medium positive correlation (0.56, p= <0.01) with user-confidence at the broad classification level and a medium positive correlation at the sub hierarchical level (0.366, p= <0.01) (Table 5.15 & Figure 5.15) signifying that user-confidence could be an adequate indicator of the accuracy of crowdsourced habitat annotations. This signifies that the use of crowd annotations

above a certain threshold of certainty could be used to improve the accuracy of the label set.

Figure 5.15: Relationship between Classification Accuracy and Label Confidence

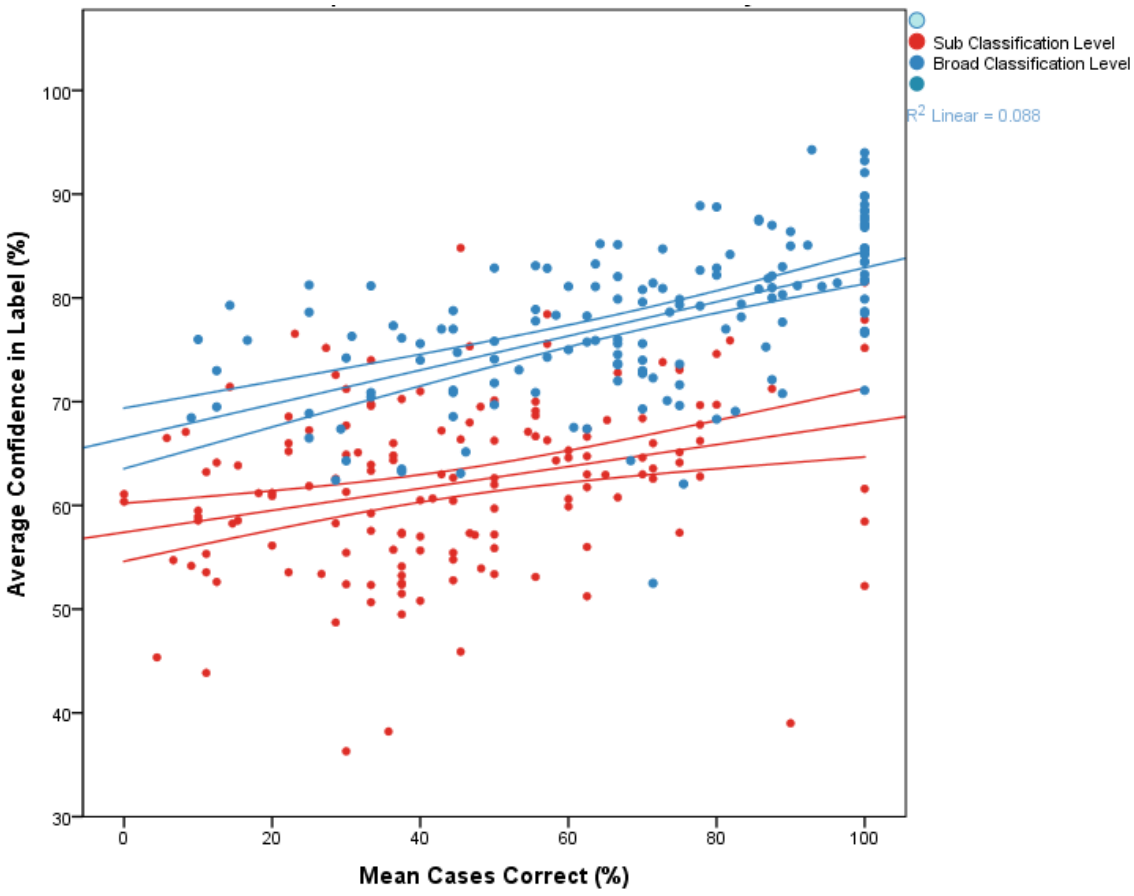


Table 5.15 Pearson's Correlation Coefficient Analysis

Hierarchical Level	Pearson Correlation	Significance
Broad	0.56	0.01
Sub	0.36	0.01

5.5.6 Fuzzy Assessment

Landscapes are not crisply segmented (Rocchini, 2007) and the hierarchical nature of the classification system used also signifies that interpreters can be correct to varying degrees. Relative similarity scores based upon Woodcock and Gopal's (2000) linguistic scales were thus attributed to each classification based on the hierarchical nature of the classification system and the identification of habitats that are commonly

juxtaposed (Table 5.3). When taking fuzzy accuracy methods into account, the percentage accuracies for the sub-classes (aggregated by broad class in Table 5.16) improve significantly with the highest improvement being seen for ‘open water’ and ‘grassland and marsh’ categories which encompass a great number of hard to distinguish sub-categories.

Table 5.16: Percentage Accuracies and Percentage Improvements n.b. headings based on Woodcock and Gopal study

	Matches Using (Cumulative %)			Improvement	
Habitat Class	**Max M**	**Right R**	**Somewhat S**	**(R-M)**	**(S-M)**
Woodland & Scrub	31.7	59.5	87.8	27.8	56.1
Grassland & Marsh	23.3	66.7	67.1	43.4	43.8
Tall Herb & Fern	38.7	69.2	93.7	30.5	55
Heathland	16.8	55.7	98.4	38.9	81.6

5.5.7 Uncertainty in Classification (Qualitative Analysis of Open Text Feedback)

Problems such as the low image resolution highlighted by participants (Table 5.17) could provide an explanation for the low sub-scale accuracy; certainly the sub grassland and hedgerow categories require high resolution imagery to classify them to an appropriate degree of accuracy. Border areas were also highlighted as something which prohibited classification; this information highlighting uncertain areas could however be of use in determining ecotonal boundaries and provide a good indicator as to where further professional sampling effort is required.

The qualitative data also shows interesting interpretation variations owing to the fact that certain international participants were unfamiliar with certain categories. This suggests that familiarity or exposure to a certain habitat could promote accuracy and the presence of between-group variance in case labelling, however, no statistically significant variance in accuracy was found between UK and non-UK participants.

Table 5.17 Comments from Participants

Theme	Participant Comment
Low Image Resolution	Can't see enough detail to be any more specific about the vegetation - I'm making inferences from the Salix in the surrounding area. Certainly not a single-species hedge - diversity looks good but not easy to assess. As it is quite heavily grazed I don't find it possible to tell whether this is dry or humid heath (can't see whether Molinia is present but I suspect not). Also it may have a reasonably high base status but not enough detail in the grassland to be sure.
Border Areas Causing Uncertainty	The image is located at the border between two habitats . The habitat is dominated by two different types of grass, which are grouped together in patches Hard to tell as many different types in same area.
Lack of Context	I have to choose "Intertidal" for coastal open water, although I have no way of knowing. Perspective issue - more bracken in landscape but not in photograph.
Classification System inadequacies	To me, this looks to be heathland overgrown with bracken but the only classification that allows for "bracken" is "tall herb and fern” It's certainly a grassland community but impossible to classify further with any degree of accuracy given that I'm viewing on a phone and have no species list.
Lack of participant knowledge / understanding	Not sure if I'm familiar enough with UK ecosystems to answer. Interesting. I have to look up half the expressions though, like the difference between mire and marsh which is pretty clear to me in German, but not in English.
Easy to identify categories	Bracken is very easy to identify given the distinctive leaf structure.

5.6 Discussion of Results

5.6.1 Directed Crowdsourcing for Habitat Mapping

A habitat classification platform has been developed in order to test the viability of using directed crowd contributions to classify habitats in trial areas of the New Forest. A total of 1,795 crowd habitat classifications from 82 participants of 150 georeferenced photographs have been collected and analysed to determine the viability of a directed crowdsourcing approach with regards to Phase 1 Habitat Mapping data collection. The results suggest great promise in using the crowd to generate training data for remote sensing classifications but the analysis of the qualitative feedback suggests that there are several technical design and usability aspects to address in any future re-development stages which could further improve the accuracy of the crowd annotators.

The modal accuracy of the crowd annotators was 77.33% at the broad hierarchical level and 47.33% for the sub. Whilst there is no pre-defined minimum acceptable level of accuracy Thomlinson et al. (1999) have set a target of an overall accuracy of 85% with no class being less than 70% accurate. This suggests that the dataset obtained from the crowd in this instance is not entirely usable without any filtering. There is however an inherent variability in accuracy when using a supervised classification technique. Human interpretations introduce an element of subjectivity (Verburg et al. 2012) and having used 82 participants within this study the variability due to the subjectivity of various annotators with varying levels of training is likely to be huge. There is even a great deal of variance between what one ecologist states to be correct and the opinion of another with the modal classifications for the ecologist group having a 74.45% match with the ground truth provided by a separate ecologist. This leaves a huge problem of what to do when there is no absolute ground truth.

When taking into account the fuzzy accuracies using the Woodcock and Gopal method the accuracy derived from the crowd classifications are quite high levels of accuracy with an 81.6% improvement for Heathland and a 63.6% improvement for Open Water. The Grassland and Marsh category is the only one remaining with an ‘inaccurate’ (< 85%) level of agreement with the authoritative data at 67.1%. This shows that the crowd annotations can be very helpful for projects that do not require the ultimate level of precision.

The inter group variance seen between groups highlights the importance of targeting a crowd of participants with certain characteristics and attributes. For this study there

was a significant difference in the accuracy of participants that had an ecology background and those with an interest in ornithology in comparison with those who had no interest in ecology and the outdoors (with a variance of up to 12.91% in accuracy). Kazai et al. (2011) have developed a worker typology based upon personality traits and found correlations between certain worker types and the level of accuracy obtained from crowd labels. Methods have also been proposed to determine worker strength from personality and demographic traits (Li et al. 2014, Ipeirotis, 2010 & Ross et al. 2010) This suggests that it may be possible to design citizen science surveys in a way that attracts a particular group of people based on ideal traits and interests via a screening survey or similar. Certainly in this case using the modal values from ecologists and ornithologists yields a higher result than using the modal values of all participants showing that the attracted contributor base is instrumental in determining the quality of labelling.

The results show that some classes are definitely easier for crowd contributors to classify than others; the accuracy of broad level classifications varies up to 39.89% with Woodland and Scrub and Heathland being the categories for which there was the least agreement between the contributors and the ecologist's ground truth interpretations. Estima et al. (2014) found inter class accuracy variances when classifying land cover using Flickr photographs. Arsanjani et al. (2015) also conducted a per class analysis investigating the usability of OSM for Land Use data and found certain classes to be more exploitable than others. In the case of this study Tall Herb and Fern, Open Water and Miscellaneous categories are certainly worth exploring, particularly from certain user groups.

5.6.2 The Importance of Prioritising Data Collection

Recognising the flaws in existing data sets is important; scantly available on ground resources can then be used to prioritise the collection of data accordingly. This experiment has shown how it is very difficult to motivate the crowd to perform this kind of labelling at a mass scale for more than a few images. Most contributors were happy to try a few classification tasks with very few opting to classify over half of all the available tasks. It is thus important to fully focus crowd sampling efforts on the areas of ultimate uncertainty, where automated methods fail or where very high resolution sampling is required for a specific scientific purpose. Grassland areas and forest plantations are currently widely misrepresented, with estimates of global grassland cover deviating widely from various sources (Fritz et al, 2008). Ecotonal areas (transitional zones between adjacent ecosystems and ecological systems) are also

difficult to map using automated systems and as they house species that are at the limits of their physical and competitive tolerances (Wasson et al. 2013) they are very susceptible to environmental changes. Focusing crowd ground observations on such quickly transitioning areas could improve the representation of important land cover metrics which are currently poorly recorded.

5.6.3 Label Weighting

Generally the use of volunteer annotators denotes that the knowledge and motivations of the annotator are usually uncontrollable, but measurable, variables. Ambiguity of data, poor definition of an annotation scheme and mistakes owing to the annotator's lack of motivation or knowledge are said to lead to noisy data (Kilgariff, 1998). Ideally various expert annotators would characterise the same data permitting the inter-annotator agreement to be computed before a consensus based classification is made. This ensures the thematic quality of a label and permits the generation of a degree of certainty to which the end user can trust the label based upon the degree of homogeneity there is in how the data is labelled.

This study has investigated group membership, the time spent classifying each case and the annotators confidence in the label as potential indicators of annotator accuracy. Small significant negative correlations were found with both time and the number of clicks that annotators made when classifying the images (Broad = -2.44, Sub = -2.45). These results contrast to those present in the existing literature (Zhai et al. 2012) that do find a significant positive relationship between classification time and the accuracy of photograph labelling from crowd contributors. This result is attributed to the greater level of care that is perhaps taken if an annotator spends longer on a classification. However, the results in this study can perhaps be attributed to the fact that if a classification takes longer to perform, the annotator is more uncertain and they would spend a greater amount of time deciding or looking at the reference guide to determine a class. The annotator may be more likely to get the label wrong because it is a more difficult to determine case, potentially with several viable options.

A human based approach of determining classification is also trailed in which the contributors were asked to self-rate the confidence they had in the correctness of their classification given the contributor is best placed to evaluate their label quality. A medium positive correlation was found between classification accuracy and the level of confidence contributors had in their classifications (Broad = 0.56, Sub = 0.37). This signifies that self-rated confidence assessments could be used to filter the data to

yield the most accurate response. Oyama et al. (2013) have also found how self-rated confidence scores can improve the accuracy of crowdsourced labels, particularly when using a small number of annotators. The fact that some users are perhaps over confident in their responses whilst some are under confident is problematic (Oyama et al. 2013) but could be alleviated with some kind of user confidence collaboration test (Kato and Zhang, 2010). Tuia et al. (2013) also use confidence scores with success to improve classification accuracy by focusing on ensuring that the contributor is able to provide a reliable label. Contributors can become frustrated when encountering numerous cases they are unable to label to the point at which they might give unreliable information. To alleviate this effect the confidence of the user is assessed at each iteration, so that a similar pixel is only presented to him/her if the confidence about the previous similar pixel exceeds a given threshold θ. Both this and other experiments show how self-rated confidence scores can be very useful when weighting the crowdsourced submissions to improve the accuracy of the final output.

There are a number of arguments as to what the best circumstances are under which to concentrate labelling effort. The accuracy of performance of a learned model is dependent upon the quantity and quality of labels (Sheng et al. 2008); a trade-off is presented between labelling previously unlabelled samples versus attaining 'round robin' repeat samples of a subset to reduce the noise of the sampling (caused by low quality labels). With professional interpreters experienced in the field of classification the level of intra-label noise is likely to be minimal. Linus' law denotes that if enough contributions are made, the correct one will eventually be found by consensus approach, however where the project is not collaborative, a high number of unprofessional contributions will generate a great deal of noise, potentially degrading the quality of the dataset. Foody et al. (2013) link this to the Hughes effect, by which the accuracy of a remote sensing classification can actually decline even with a surplus of high accuracy data. Indeed, another study, (Sheng et al. 2009) showed how the marginal increase in performance for a given change in the dimensions of label quantity and repeat label quantity varied hugely for different combinations of these dimensions. This study has shown how group attribution, classification time and user uncertainty are measurable variables which can be shown to have some bearing on the accuracy of crowd interpretations.

5.6.4 Qualitative Interpretations

Analysis of the annotator comments shows how a number of factors could have contributed to participant inaccuracy; the low resolution of some images was a problem for certain participants with one quoting that the level of diversity of the hedge was not easy to assess from a photograph and another was unable to determine the dry/ humid status of heathland as the photograph was not detailed enough to see any indicator species such as *Molinia.* This is the primary drawback of classifying a landscape from the web rather than *in situ*. In an ideal system there would be multiple photographs of the same location from which contributors can get an enhanced understanding of the habitat. Geo Wiki allows users to state whether or not an image is of high resolution or not which could perhaps be used as an indicator of classification accuracy (Fonte et al. 2015).

Perspective and having a lack of context was also problematic for contributors; an instance where there was more bracken in the landscape but not within the photograph (which was taken from a small patch of grassland) led to a contributor classing the areas as Acid Grassland when the ecologist ground truth stated it to be Dense / Continuous Bracken. Crowd participants were also made uncertain of their classifications by the presence of border areas e.g. the presence of two different types of grass which were clumped together in patches.

The lack of flexibility within the classification system also signifies that the annotators were not necessarily putting the most appropriate answer they were capable of giving. One participant noted that an area was a mixture of Heathland and Bracken but was only able to put one broad class signifying that the noticed degree of heterogeneity could not be recorded. One way around this would be to perhaps allow users to indicate the degree of landscape heterogeneity and label an area with multiple broad classes. It does however start to become a trade-off between obtaining the highest possible quality of label and not over facing the participant with too many tasks.

5.7 Critique of Study

5.7.1 Simulated Crowdsourcing Environment

The study relies on a series of assumptions made about the data collected. It firstly assumes that the participants recruited were representative of the volunteering ‘crowd’ rather than solely friends willing to perform the experiment. Owing to ethical procedure it was impossible to eliminate the fact that the study was indeed a research

focused study which could have influenced the type of contributors, how much effort they put into the classification tasks and how long they spent performing them. Knowledge that the data was unlikely to be used directly within environmental management could have been off-putting for certain participants. Citizen scientists in the domain of ecology are usually particularly keen for the collected data to be open and accessible. Despite this limitation it is considered that sufficient participants were obtained to enable the results obtained within this study to be of significance.

5.7.2 Participant categorisation

Whilst the rationale of participant categorisation is explained it is not necessarily wholly appropriate. A wide spectrum of respondents meant it was difficult to form suitable groupings – those with a MSc qualification and above were thought of as 'expert ecologists' though it was flawed to put an MSc graduate and an ecologist with 25 years of experience in the same grouping. Similarly some botanists and ornithologists from special interest groups were highly skilled in habitat identification but allocated the 'non specialist' grouping owing to their lack of professional qualification. Future work could perhaps delve further into the level of prior knowledge and experience that the crowd has to provide more appropriate inter-group comparison.

5.7.3 Ecologist 'Ground Truth'

The ecologist classifications were assumed as the authoritative ground truth classifications for each of the cases. Whilst this is a reasonable assumption to make owing to the ecologist's strong familiarity with Hampshire and site visits made in concordance with the specification of the Phase 1 Habitat mapping schema, it is possible that a different ecologist would have classified the cases differently. The assumption that the 'real' land cover data is not free from error is commonly incorrect (Congalton, 1991). This is an important factor as it is crucial to adequately assess the accuracy of subsequent digital classifications (Congalton, 1991) and the volunteer contributors. A consensus approach is optimal when obtaining ground truth for remote sensing applications; the Kappa variance of 0.68 between participants within the professional ecologist grouping shows how there was uncertainty in classification that was not necessarily the result of skill variation.

5.8 Summary

Through investigation, this study has addressed the study aims in the following ways:

- Can scientific sampling procedures be pushed onto volunteer data?

  The experiment presented within this chapter has demonstrated that crowd contributors are very much capable of following scientific sampling procedures with regards to the labelling of land parcels with Phase 1 Habitat Data. Some difficulties have been presented with the labelling procedures such as confusion owing to international annotators being unfamiliar with the UK's different habitat categories and with the use of ground based photography such as the poor image resolution and inability to see the context around the feature of classification.

- Can the crowd produce training samples that are as credible as those provided by a professional ecologist and can any variance in accuracy between feature types be identified?

  Whilst the crowd can evidently produce good quality training samples they are not sufficiently accurate for immediate usage. Using a group consensus approach and adaptation of Woodcock and Gopal's linguistic scale to determine the fuzzy accuracy of the labels does improve the thematic accuracy significantly. There is variance in the accuracy of classifications between habitat classes suggesting that crowd annotators are more useful in certain environments than in others. Interestingly the classifications from professional ecologists (included within the crowd annotations) do not have a high agreement with those of the ecologist ground truth 'gold standard' data set. This presents a broader issue of standardising and developing a best practise framework for classifying habitats. Considering the complexity of the Phase 1 Habitat Classification system and the discussed flaws of using ground based photography, directed crowd samples are a promising substitute for authoritative training samples where gold standard data is inaccessible.

- Can predictors of crowd accuracy be established to enable the elimination of information that is detrimental to classification accuracy?

The results show user-confidence and classification time as encouraging indicators of label accuracy. The group to which the participants were assigned also presents significant variances in the degree of label accuracy between Ecologist/Ornithologist group members and those with no interest in ecology. These indicators could perhaps be used to filter the data further and to easily prune datasets for classifications which are very unlikely to be accurate.

Chapter 6 | Synthesis of Findings and Future Research Directions

6.1 Introduction

This chapter details the research findings and draws conclusions from the two main studies presented in this thesis. The outcomes of the research objectives are discussed in turn before possible routes not explored throughout this work are drawn upon to develop suggestions for future research directions in the domain of using directly crowdsourced information and open web data in producing ecological data sets.

6.2 Summary of Findings

6.2.1 Assessing the Potential for Open Web Data in Enhancing the Quality of Authoritative Land Cover Mapping

Work based upon an Ordnance Survey Research internship project (2012) looked into the feasibility of using open web land cover data from Volunteered Geographic Information sources. Geograph (a landscape photo sharing site) and OpenStreetMap were assessed as land cover sources compared with the authoritative Centre for Ecology and Hydrology Land Cover Map data for Hampshire. Grid square comparisons and Python word matching scripts were used to determine the extent to which the data varied in terms of attribute quality and distribution. Spatial word matching of VGI strings with an authoritative dataset indicated that passively contributed land cover open web data could assist in providing an additional source of high quality land cover information. An analysis of terms showed the potential to scrape highly detailed information from open web data sources.

The analyses undertaken within this first study suggest that the approach of using open web VGI when authoritative land cover information is unavailable provides a promising avenue to providing the *in situ* ground truth data required to generate high quality land cover maps. Owing to the adequate coverage, they could be useful as reference data where existing methods misinterpret imagery and the relatively high temporality of the data could prove useful in flagging up landscape changes.

In terms of attributes there is some potentially minable information; there is some very detailed species specific information within Geograph which could be very useful to those requiring land cover information at a high level of detail. The lack of standardisation in labelling though prohibits unedited use. Without rigid ontology rules, that risk constraining the fluidity of crowdsourcing projects, substantial work is

required to index the open web's ad-hoc terminology within a nationally applicable standard. There are also some substantial barriers from a business perspective in terms of assuring the quality of open web data and in mitigating for its current lack of interoperability with other data sources. It is clear that crowd generated content cannot be used without a great deal of pruning which consequently could reduce the potential benefits offered by these datasets. These outcomes suggest the merit of using more directed approaches when obtaining data for analysing ecological features such as land cover parcels which cannot be collected in such an ad-hoc manner.

6.2.2 Assessing the Utility of Directed Crowdsourcing in Habitat Classification

The results of two preliminary studies discussed within Chapter 4 contributed towards the development of a website designed to collect crowd annotations of ground based photography. Directed crowd assisted annotation was trialled as a mechanism of producing accurate thematic habitat maps. Volunteered Geographic Information classifications of ground-based photographs (made using the UK's Phase-1 Habitat Survey nomenclature) were gathered and filtered according to metrics of annotator accuracy and user-confidence in labelling. Interpretations for 150 sites initially classified by a professional ecologist in subsets of varying class distinguishability were then subjected to accuracy assessment techniques to determine their utility as training data for the classification of remotely sensed imagery.

The work has demonstrated that crowd annotators are capable of producing accurate training samples, however, poor image resolution and the inability to see the context around the feature of classification generates confusion leading to discrepancies between the authoritative label and the label perceived to be correct by the crowd. A group consensus approach and a linguistic scale to determine the fuzzy accuracy of the labels has been found to improve the overall thematic accuracy significantly. Inter class variance in the accuracy of the classifications also suggests that crowd annotators are more useful in certain environments than in others.

The results show user-confidence, group attribution and classification time as encouraging indicators of label accuracy. Developing these and other indicators of annotator accuracy as methods for weighting crowd contributor classifications accordingly prior to use in remote sensing data training is likely to be a key future research area (as discussed in section 6.3).

6.3 Next Steps for Future Research

6.3.1 Incorporating Crowd Contributions with Active Learning

Active Learning is a machine learning technique that facilitates interaction between the user and computer in order to improve the efficiency of supervised classification. Where large quantities of adequate training data are not available, the active learning approach can be utilised to iteratively expand the training set, improving the performance of the model (Tuia et al. 2011). This algorithm repeatedly queries available training cases and selects the ones that are necessary for the accurate classification of the image i.e. the pixels that the classifier is most uncertain regarding, according to pre-determined uncertainty criterion. An example criterion would be the distance from the separating hyperplane. The supervisor (user) is then able to label these cases deemed most useful before the process is repeated until an accuracy threshold is met, to optimise the classification. This method ensures that many redundant training cases are not unnecessarily classified by the user which greatly improves efficiency in terms of both time and cost. The method is more computationally efficient than the semi-supervised approach in that the process is not fully automatically iterated; interaction with the user is required (Bruzzone and Persello, 2009).

Directing specific users to specific areas of interest is a concept that lends itself to the active learning method of image classification. If steps can be taken to ensure that the volunteer is competent and confident at classifying the 'uncertain' pixel (by only giving them pixels they are more likely to be able to identify), the classification is likely to have greater reliability. Rebbapragada and Oommen (2011) suggest the improvement of crowdsourced earthquake damage assessment with remotely sensed imagery training, trialling semi automation shifting the focus from the classifying of the entire database to generating reliable training data. Tuia and Munoz-Marí (2012) use optimisation approaches which do not assume uniformity between users classifying pixels. This denotes that there could be scope for user reliability metrics to be incorporated into the classification training to further improve data reliability in a VGI context. This alternative model of crowdsourcing by which fewer areas are classified many times by appropriate users could have significance in improving both the quality and completeness of classifications as local and temporally accurate knowledge is fed into machine learning techniques.

6.3.2 Exploring the Motivations behind Directed Crowdsourcing

Motivation is a key research area in the domain of citizen science and crowdsourcing (Rula et al. 2014). Motivated contributors are more likely to stay with a crowdsourcing project and contribute repeatedly (Raddick et al. 2010), so it is integral to the futurity of citizen science projects that task designers are aware of what best motivates a contributor, enabling the design of tasks that will attract more contributions. Motivation is also something that can be investigated in terms of task performance (Rogstadius et al. 2011). Shaw et al. (2011) discovered that different incentives may elicit higher quality performances than others through undertaking experiments with varying rewards. As such it would be interesting to see whether the motivation behind participation in the classification task (whether it be interest in ecology, altruism or the promise of some form of cash or prize based reward) has any influence on the accuracy of the information submitted by the crowd with regards to land cover and habitat data.

6.3.3 In-Situ Directed Data Collection

A logical progression from the work presented within this thesis would be to trial the directed crowdsourcing method in the field, using mobile devices. The directed crowdsourcing study within this thesis found that the lack of context and resolution issues associated with ground based photography presented strong barriers to successful habitat classification. Repeating with *in situ* data submitted from a mobile device could alleviate these problems and allow annotators to feed in any local knowledge they may have about the landscape.

Location based sensing techniques have been proposed for use in the EU Citizen Observatory Web (COBWEB) project (Meek et al. 2014). A technique developed by Meek et al. (2013) uses smartphone sensors to conduct remote logging and query via line of sight which can be used to direct the user towards parts of a landscape that are of interest to project orchestrators. Directing users on the ground to a specific site of interest can be used to increase the number of observations in areas of low sampling density, partially addressing the lack of coverage that crowdsourced datasets suffer from.

There is potential to use the location based sensing technologies within smartphones in combination with previously discussed (6.3.1) active learning techniques to develop iterative in-field classification. The active learning classification technique is an iterative process; a smartphone device could be used to send crowd contributors to the location representing the next most uncertain pixel to be classified. This technique could reduce uncertainty of an input by collecting related data from more than one source and denotes that fewer locations need to be sampled given the efficiency of sampling using the active learning method (Tang et al. 2011). Notably the poor quality of rural data connectivity would inhibit this potential; it is important that practical difficulties be overcome to advance the utility of crowdsourcing.

6.4 Conclusions

Several experiments have been undertaken where contributors validate land cover maps but this work presents a novel focus on how contributor data can be used in both a directed and mined way to classify land cover where authoritative data is not necessarily available in such detail. The thesis also presents the first known extraction of land cover reference data from OSM (Fonte et al. 2015).

VGI is fundamentally patchy, inconsistent and of a low quality (Coote and Rackham, 2008). Where VGI is present in sufficient quantity and quality, extracting useful information is possible (Foody et al. 2013), but the cost of filtering the data makes this impractical as a widespread solution. Iwao (2006) and See et al. (2013) have shown that annotators can be as good as professionals given training yet this does require a great deal of coordination and a focus upon motivating volunteers to give information in a very specific manner. Whilst using VGI as a validation tool can be useful, this focus is not utilising VGI to its full capacity; the unique advantage of VGI is currency, providing new data about landscapes that can be used to infer change detection and monitor species coverage. There are opportunities to move beyond the use of VGI for validation and to focus on methods of collecting high quality VGI. Filtering VGI according to indicators of accuracy and integrating the annotations within remote sensing classification techniques is likely to yield high frequency land cover maps of a usable quality (for certain applications). This thesis has examined the utility of VGI as ancillary data for remote sensing classifications and has shown how datasets can be synthesised that, whilst not reaching the accuracies of authoritative data, could be seen as an asset to ecological monitoring and inventory.

References

Andelman, S. J. (2011). Conservation science outside the comfort zone. *Nature*, *475* (7356), 290-291.

Anderson, J. R. (1976). *A land use and land cover classification system for use with remote sensor data* (Vol. 964). US Government Printing Office.

Antoniou, V., & Schlieder, C. (2014). *Participation Patterns, VGI and Gamification.* In Proc. 17th AGILE Conference on Geographic Information Science, Castellón, Spain (pp. 3-6).

Aplin, P. (2004). Remote sensing: land cover. *Progress in Physical Geography*, *28* (2), 283-293.

Arnot, C., Fisher, P. F., Wadsworth, R., & Wellens, J. (2004). Landscape metrics with ecotones: pattern under uncertainty. *Landscape Ecology, 19* (2), 181-195.

Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. (2013). Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science, 27* (12), 2264–2278

Baglatzi, A., & Kuhn, W. (2013). *On the formulation of conceptual spaces for land cover classification systems*. In Geographic information science at the heart of Europe (pp. 173-188). Springer International Publishing.

Bakillah, M., Liang, S., Mobasheri, A., Jokar Arsanjani, J., & Zipf, A. (2014). Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science, 28* (9), 1940-1963.

Ban, Y., Gong, P., & Giri, C. (2015). Global land cover mapping using Earth observation satellite data: Recent progresses and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing,* 103, (2015) 1-6.

Baraldi, A., Bruzzone, L. and Blonda, P. (2005). Quality assessment of classification and cluster maps without ground truth knowledge. *IEEE Transactions on Geoscience and Remote Sensing, 43* (4), 857 – 873.

Bigham, J. P., Bernstein, M. S., & Adar, E. (2014). Human-Computer Interaction and Collective Intelligence in *The Collective Intelligence Handbook.* MIT Center for Collective Intelligence.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 65* (1), 2-16.

Blaschke, T., Lang, S., Lorup, E., Strobl, J., & Zeil, P. (2000). Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. *Environmental Information for Planning, Politics and the Public, 2*, (4) 555-570.

Blum, J. R., Greencorn, D. G., & Cooperstock, J. R. (2013). *Smartphone sensor reliability for augmented reality applications. In Mobile and Ubiquitous Systems: Computing, Networking, and Services* (pp. 127-138). Springer Berlin Heidelberg.

Brando, C., & Bucher, B. (2010, May). *Quality in user generated spatial content: A matter of specifications*. In Proceedings of the 13th AGILE International Conference on Geographic Information Science (pp. 11-14). Springer Verlag: Guimar aes, Portugal.

Brody, S., Blessing, R., Sebastian, A., & Bedient, P. (2014). Examining the impact of land use/land cover characteristics on flood losses. *Journal of Environmental Planning and Management, 57* (8), 1252-1265.

Bruns, Axel. "Towards produsage: Futures for user-led content production." (2006): Futures for User Led Content Production. Proceedings: Cultural Attitudes towards Communication and Technology, 275-284.

Bruns, A. (2008). Blogs, Wikipedia, Second Life, and beyond: From production to produsage (Vol. 45). Peter Lang.

Bruzzone, L., & Persello, C. (2009). A novel context-sensitive semi supervised SVM classifier robust to mislabeled training samples. *Geoscience and Remote Sensing, IEEE Transactions on*, *47* (7), 2142-2154.

Budhathoki, N. R., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal, 72* (3-4), 149-160.

Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for open collaboration crowd and community models and the case of OpenStreetMap*. American Behavioral Scientist, 57* (5), 548-575.

Castellote, J., Huerta, J., Pescador, J. and Brown, M. (2013). *Towns conquer: a gamified application to collect geographical names (vernacular names/toponyms)*. In: AGILE 2013, 14-17 May, 2013, Leuven, Belgium

CCS Insight (2013) Mobile Phone Sales Will Hit 1.86 Billion in 2013 as Strong Smartphone Growth Continues [Online] Available at: http://www.ccsinsight.com/press/company-news/1655-mobile-phone-sales-will-hit-186-billion-in-2013-as-strong-smartphone-growth-continues [Accessed: 13/6/13]

Cohen, J. (1988). *Statistical power for the social sciences.* Hillsdale, NJ: Laurence Erlbaum and Associates.

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research, 4* (1), 332-358.

Comber, A., Fisher, P., Wadsworth, R., 2005. What is land cover? *Environment and Planning B 32*, 199-209.

Comber, A., Fisher, P., Brunsdon, C., & Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment*, 127, 237-246.

Comber, A., See, L., Fritz, S., Van der Velde, M., Perger, C., & Foody, G. (2013). Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation, 23*, 37-48.

Congalton, R. G. (1991). Remote sensing and geographic information system data integration: error sources and. *Photogrammetric Engineering & Remote Sensing, 57* (6), 677-687

Cope, M., & Elwood, S. (Eds.). (2009). *Qualitative GIS: a mixed methods approach.* Sage.

Coote, A. and Rackham, L., 2008. *Neogeographic data quality – is it an issue?* Paper delivered at the AGI Geocommunity conference 2008, Consulting Where Ltd.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Dalby, S. (2012). Geo 2.0: digital tools, geographical vision and a changing planet. *The Geographical Journal, 178* (3), 270-274.

DeforestAction (2011) Earth Watchers [Online] Available at: http://dfa.tigweb.org/about/?section=earthwatchers Accessed: 12/4/12

De Longueville, B., Smith, R. S., & Luraschi, G. (2009, November). *Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires.* In Proceedings of the 2009 international workshop on location based social networks (pp. 73-80). ACM.

Devictor, V., Whittaker, R. J., & Beltrame, C. (2010). Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions, 16* (3), 354-362.

Devillers, R., & Jeansoulin, R. (2006). *Fundamentals of Spatial Data Quality* (p. 312). ISTE Publishing Company.

Devillers, R., Bégin, D., & Vandecasteele, A. (2012, September). *Is the rise of Volunteered Geographic Information (VGI) a sign of the end of National Mapping Agencies as we know them.* In: GIScience 2012 workshop "Role of Volunteer Geographic Information in Advancing Science: Quality and Credibility.

Dickinson, J., Zuckerberg, B., & Bonter, N. (2010), Citizen Science as an ecological research tool: challenges and benefits. *Annual review of ecology, evolution, and systematics, 41*, 149-172.

Du, H., Alechina, N., Jackson, M., & Hart, G. (2013*). Matching Formal and Informal Geospatial Ontologies*. In Geographic Information Science at the Heart of Europe (pp. 155-171).

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the association of American geographers, 102* (3), 571-590.

ESA, 2015. [Online] Available at: http://europa.eu/pol/space/index_en.htm [Accessed 07/02/15]

Estima, J., & Painho, M. (2013). *Flickr Geotagged and Publicly Available Photos: Preliminary Study of Its Adequacy for Helping Quality Control of Corine Land Cover.* In: Computational Science and Its Applications – ICCSA 2013. vol. 7974. pp. 205–220.

Estima, J., & Painho, M. (2014). Photo Based Volunteered Geographic Information Initiatives: A Comparative Study of their Suitability for Helping Quality Control of Corine Land Cover. *International Journal of Agricultural and Environmental Information Systems (IJAEIS), 5* (3), 73-89.

Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science, 28* (4), 700-719.

FAO (2000).Land Cover Classification System. [Online] Available at: http://www.fao.org/docrep/003/x0596e/x0596e00.htm Accessed 14/6/13 [Accessed 12/5/14]

Fairbairn, D., & Al-Bakri, M. (2013). Using Geometric Properties to Evaluate Possible Integration of Authoritative and Volunteered Geographic Information. *ISPRS International Journal of Geo-Information, 2* (2), 349-370.

Fisher, P., Arnot, C., Wadsworth, R., & Wellens, J. (2006). Detecting change in vague interpretations of landscapes. *Ecological Informatics, 1* (2), 163-178.

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal, 72* (3-4), 137-148.

Fonte, C.C., Bastin, L., See., Foody, G. and Lupia F. 2015 Usability of VGI for Validation of Land Cover Maps. *International Journal of GIS,* 29:7

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80* (1), 185-201.

Foody, G. M. (2009). The impact of imperfect ground reference data on the accuracy of land cover change estimation. *International Journal of Remote Sensing, 30* (12), 3275-3281.

Foody, G. M. (2013). Rating crowdsourced annotations: evaluating contributions of variable quality and completeness. *International Journal of Digital Earth, 7* (8), 650-670.

Foody, G. M., & Boyd, D. S. (2012). *Exploring the potential role of volunteer citizen sensors in land cover map accuracy assessment.* In: Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Science (Accuracy 2012), Florianopolis, Brazil (pp. 203-208).

Foody, G.M., Boyd, D.S., 2013. Using Volunteered Data in Land Cover Map Validation: Mapping West African Forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6*, 1305–1312.

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Transactions in GIS, 17* (6), 847-860.

Frias-Martinez, V., & Frias-Martinez, E. (2014). Spectral clustering for sensing urban land use using Twitter activity. *Engineering Applications of Artificial Intelligence, 35*, 237-245.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., Obersteiner, M., 2009. Geo-Wiki.Org: the use of crowd-sourcing to improve global land cover. *Remote Sensing 1* (3), 345-354.

Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., Van der Velde, M., & Achard, F. (2011). Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters, 6* (4), 044005.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., & Obersteiner, M. (2012). Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software, 31*, 110-123

Galloway, S. (2009). Simulation techniques to bridge the gap between novice and competent healthcare professionals. *Online Journal of Issues in Nursing*,14 (2).

Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems,* (3), 10-14.

Getzin, S., Wiegand, K., & Schöning, I. (2012). Assessing biodiversity in forests using very high-resolution images and unmanned aerial vehicles. *Methods in Ecology and Evolution, 3* (2), 397-404.

Giri, C., Pengra, B., Long, J., & Loveland, T. R. (2013). Next generation of global land cover characterization, mapping, and monitoring*. International Journal of Applied Earth Observation and Geoinformation*, 25, 30-37.

Girres, J. F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS, 14* (4), 435-459.

Gladstone, C. S., Gardiner, A., & Holland, D. A. (2012). *Semi-automatic method for detecting changes to ordnance survey topographic data in rural environments*. Proceedings of the 4th GEOBIA, São José dos Campos, Brazil, 7-9.

Gong, P. (2012). Remote sensing of environmental change over China: A review. *Chinese Science Bulletin, 57* (22), 2793-2801.

Godet, L., Fournier, J., Toupoint, N., & Olivier, F. (2009). Mapping and monitoring intertidal benthic habitats: a review of techniques and a proposal for a new visual

methodology for the European coasts. *Progress in Physical Geography*, *33* (3), 378-402

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal, 69* (4), 211-221.

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth, 3* (3), 231-241.

Goodchild, M. F., & Hunter, G. J. (1997). A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science, 11* (3), 299-306.

Goodchild, M.F. & Li, L. (2012), Assuring the quality of volunteered geographic information. *Spatial statistics, 1*, 110-120.

Gschwend, C., & Purves, R. (2011*). Comparing Flickr tags to a geomorphometric classification.* In: Proceedings of the Nineteenth Annual GIS Research UK Annual Conference, Portsmouth, United Kingdom (pp. 174-78).

Hagenauer, J., & Helbich, M. (2012). Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. *International Journal of Geographical Information Science, 26* (6), 963-982.

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal, 47* (4), 315-322.

Haklay, M., & Ellul, C. (2010). Completeness in volunteered geographical information–the evolution of OpenStreetMap coverage in England (2008-2009). *Journal of Spatial Information Science.*

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge* (pp. 105-122). Springer Netherlands.

Hampshire Biodiversity Partnership (2013) [Online] Available at: http://www3.hants.gov.uk/biodiversity/hampshire.htm [Accessed 30/6/14]

Hampton, S. E., Strasser, C. A., & Tewksbury, J. J. (2013). Growing Pains for Ecology in the Twenty-First Century. *BioScience, 63* (2), 69-71.

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote sensing of Environment*, 122, 66-74.

Harding, J. (2011). *Usability of geographic information: factors identified from task-focused user interviews*. In Proceedings of the 25th international cartographic conference, Paris (pp. 3-8).

Hassan, M. M. (2005). Arsenic poisoning in Bangladesh: spatial mitigation planning with GIS and public participation. *Health Policy, 74* (3), 247-260.

Heipke, C. (2010), Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing, 65* (6), 550-557.

Hooftman, D. A. P., & Bullock, J. M. (2012). Mapping to inform conservation: A case study of changes in semi-natural habitats and their connectivity over 70years. *Biological Conservation, 145* (1), 30-38.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. G., Zipf, A., and Albuquerque, J. P., (2013) *The use of volunteered geographic information and crowdsourcing in disaster management: A systematic literature review.* In: Proceedings of the Nineteenth Americas Conference on Information Systems, Atlanta, Georgia: 1–10

Huang, Q., & Xiao, Y. (2015). Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information, 4* (3), 1549-1568.

Ipeirotis, P. G., Provost, F., & Wang, J. (2010). *Quality management on amazon mechanical turk*. In Proceedings of the ACM SIGKDD workshop on human computation (pp. 64-67). ACM.

Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development.* Psychology Press.

ISO, 2002. ISO/TC 211, 19113 Geographic Information Quality Principles, International Organisation for Standardisation.

Iwao, K., Nishida, K., Kinoshita, T., & Yamagata, Y. (2006). Validating land cover maps with Degree Confluence Project information. *Geophysical Research Letters, 33* (23).

Jiang, M., Bullock, J. M., & Hooftman, D. A. (2013). Mapping ecosystem service and bio-diversity changes over 70 years in a rural English county. *Journal of Applied Ecology 50* (4), 841-850.

JNCC, (2010). *Handbook for Phase 1 habitat survey -a technique for environmental audit*. ISBN 0861396367

Johansen, K., Arroyo, L. A., Phinn, S., & Witte, C. (2010). Comparison of geo-object based and pixel-based change detection of riparian environments using high spatial resolution multi-spectral imagery. *Photogrammetric Engineering & Remote Sensing, 76* (2), 123-136.

Jones, K. B. (2008). *Importance of land cover and biophysical data in landscape-based environmental assessments*. North America Land Cover Summit. Association of American Geographers, Washington, DC, USA, 215-249.

Kato, K., & Zhang, Y. (2010). *An item response model for probability testing*. In: International Meeting of the Psychometric Society.

Kazai, G. (2011). *In search of quality in crowdsourcing for search engine evaluation. In Advances in information retrieval* (pp. 165-176). Springer Berlin Heidelberg.

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., ... & Foldit Void Crushers Group. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature structural & molecular biology, 18* (10), 1175-1177.

King, R. B. (2002). Land cover mapping principles: a return to interpretation fundamentals. *International Journal of Remote Sensing, 23* (18), 3525-3545.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., ... & Horton, J. (2013, February). *The future of crowd work.* In: Proceedings of the 2013 conference on Computer supported cooperative work (pp. 1301-1318). ACM.

Kodric-Brown, A., & Brown, J. H. (1993). Highly structured fish communities in Australian desert springs. *Ecology*, 1847-1855.

Koukoletsos, T., Haklay, M., & Ellul, C. (2012). Assessing data completeness of VGI through an automated matching procedure for linear data. *Transactions in GIS, 16* (4), 477-498.

Kumar, A., & Lease, M. (2011, February). *Modeling annotator accuracies for supervised learning*. In: WSDM Workshop on Crowdsourcing for Search and Data Mining (pp. 19-22).

Lakhani, K. R., & Von Hippel, E. (2003). How open source software works:“free” user-to-user assistance. *Research policy, 32* (6), 923-943.

Lease, M. (2011), *On quality control and machine learning in crowdsourcing.* Proceedings of the 3rd Human Computation Workshop at AAAI, 97–102

Leidig, M., & Teeuw, R. (2015). Free software: A review, in the context of disaster management. International *Journal of Applied Earth Observation and Geoinformation, 42*, 49-56.

Leung, D., & Newsam, S. (2012). *Can off-the-shelf object detectors be used to extract geographic information from geo-referenced social multimedia?* In Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (pp. 12-15). ACM.

Li, H., Zhao, B., & Fuxman, A. (2014). *The wisdom of minority: Discovering and targeting the right group of workers for crowdsourcing*. In Proceedings of the 23rd international conference on World wide web (pp. 165-176). ACM.

Lowry, C. S., & Fienen, M. N. (2013). CrowdHydrology: Crowdsourcing Hydrologic Data and Engaging Citizen Scientists. *Ground Water, 51* (1), 151-156.

Lucas, R., Medcalf, K., Brown, A., Bunting, P., Breyer, J., Clewley, D., ... & Blackmore, P. (2011). Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(1), 81-102.

Lund, H.G., 2004. Definitions of Forest, Deforestation, Afforestation, and Reforestation. [Online] Available at: http://home.comcast.net/~gyde/DEFpaper.htm. [Available 14/09/15].

Luscier, J. D., Thompson, W. L., Wilson, J. M., Gorham, B. E., & Dragut, L. D. (2006). Using digital photographs and object-based image analysis to estimate percent ground cover in vegetation plots. *Frontiers in Ecology and the Environment, 4* (8), 408-413.

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M., Elston, D. A., Scott, E. M., ... & Watt, A. D. (2010). Long-term datasets in biodiversity research and monitoring:

assessing change in ecological communities through time. *Trends in Ecology & Evolution, 25* (10), 574-582.

Maué, P., & Schade, S. (2008). *Quality of geographic information patchworks*. In: Proceedings of AGILE (pp. 1-8).

Meek, S., Goulding, J., & Priestnall, G. (2013). The Influence of Digital Surface Model Choice on Visibility-based Mobile Geospatial Applications. *Transactions in GIS, 17*(4), 526-543.

Meek, S., Jackson, M. and Leibovici, D., 2014. *A flexible framework for assessing the quality of crowdsourced data*. Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, Spain.

Metcalfe, K., Delavenne, J., Garcia, C., Foveau, A., Dauvin, J. C., Coggan, R. & Smith, R. J. (2013). Impacts of data quality on the setting of conservation planning targets using the species–area relationship. *Diversity and Distributions, 19* (1), 1-13.

Millennium Ecosystem Assessment. (2005). *Ecosystems and Human Well-Being: Synthesis*. Island Press, Washington, DC

Mooney, P., & Corcoran, P. (2011). *Accessing the history of objects in OpenStreetMap*. In: Proceedings AGILE (p. 155).

Mooney, P., Corcoran, P., & Winstanley, A. C. (2010). *Towards quality metrics for OpenStreetMap*. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 514-517). ACM.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., & Simpson, I. (2011). *Final Report for LCM2007-the new UK land cover map*. Countryside Survey Technical Report No 11/07.

Mulligan, M. (2008). RALUCIAPA: Rapid assessment of land use change in and around protected areas (2000-2005). Version 1.0.

Nichols, J. D., & Williams, B. K. (2006). Monitoring for conservation. *Trends in Ecology & Evolution, 21* (12), 668-673.

Nowak, S., & Rüger, S. (2010). *How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation.* In Proceedings of the international conference on Multimedia information retrieval (pp. 557-566). ACM.

Ordnance Survey (2015) Product Accuracy [Online] Available at: http://www.ordnancesurvey.co.uk/support/products-services.html [Accessed: 12/06/15]

O'Reilly, T. (2005). Web 2.0: compact definition. [Online] Available at: http://radar. oreilly. com/archives/2005/10/web_20_compact_definition. html. [Accessed 12/5/13]

OSM Wiki (2011) [Online] Available at: https://wiki.openstreetmap.org/wiki/Main_Page [Accessed 12/05/11]

Oyama, S., Baba, Y., Sakurai, Y., & Kashima, H. (2013). *Accurate integration of crowdsourced labels using workers' self-reported confidence scores.* In: Proceedings of the Twenty-Third international joint conference on Artificial Intelligence (pp. 2554-2560). AAAI Press.

Palen, L., Soden, R., Anderson, T. J., & Barrenechea, M. (2015). *Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events.* In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 4113-4122). ACM.

Parent, J. R., & Volin, J. C. (2016). Validating Landsat-based landscape metrics with fine-grained land cover data. *Ecological Indicators, 60*, 668-677.

Parker, C.J., May, A.J. & Mitchell, V. (2011), *Relevance of Volunteered Geographic Information In A Real World Context*, Proceedings of GIS Research UK 19th Annual Conference, GIS Research UK, University of Portsmouth, April 27-29.

Pawlowicz, S., Leibovici, D., Haines-Young, R., Saull, R., & Jackson, M. (2011, October). *Dynamic Surveying Adjustments for Crowd-sourced Data Observations.* EnviroInfo 2011 Conference Proceedings.

Pérez-Hoyos, A., García-Haro, F. J., & San-Miguel-Ayanz, J. (2012). Conventional and fuzzy comparisons of large scale land cover products: Application to CORINE, GLC2000, MODIS and GlobCover in Europe. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*, 185-201.

Perger, C., Fritz, S., See, L., Schill, C., VAN DER VELDE, M., MCCALLUM, I., & Obersteiner, M. (2012). A campaign to collect volunteered geographic Information on land cover and human impact. *GI Forum*, 83-91.

Peters, D. P., Bestelmeyer, B. T., Herrick, J. E., Fredrickson, E. L., Monger, H. C., & Havstad, K. M. (2006). Disentangling complex landscapes: new insights into arid and semiarid system dynamics. *BioScience, 56* (6), 491-501

Peters, D. P. (2010). Accessible ecology: synthesis of the long, deep, and broad. *Trends in ecology & evolution, 25* (10), 592-601.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., ... & Vandenberg, J. (2010). Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review, 9* (1), 010103.

Ramos, J. M., & Devillers, R. (2013). *Integrating Authoritative and Volunteered Geographic Information: An Ontological Approach.* In P. Mooney, & K. Rehrl (Eds.), International Workshop on Action and Interaction in Volunteered Geographic Information, Leuven, Belgium (pp. 49-54).

Raykar, V. C., Yu, S., Zhao, L. H., Jerebko, A., Florin, C., Valadez, G. H., ... & Moy, L. (2009, June). *Supervised learning from multiple experts: whom to trust when everyone lies a bit.* In Proceedings of the 26th Annual international conference on machine learning (pp. 889-896). ACM.

Raykar, V. C., & Yu, S. (2011). Ranking annotators for crowdsourced labeling tasks. *Advances in neural information processing systems* (pp. 1809-1817)

Rebbapragada, U., & Oommen, T. Integrating Machine Learning into a Crowdsourced Model for Earthquake-Induced Damage Assessment.

Rocchini, D., & Ricotta, C. (2007), Are landscapes as crisp as we may think? *Ecological Modelling,* 204, 535-539.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jimenez-Valverde, A., Ricotta, C., ... & Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography, 35* (2), 211-226.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2013). GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal, 78* (1), 21-40.

Rodrigues, F., Pereira, F., & Ribeiro, B. (2013). Learning from multiple annotators: Distinguishing good from random labelers. *Pattern Recognition Letters, 34* (12), 1428-1436.

Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., and Vukovic, M. *An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets.* Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media: Barcelona, Spain, (2011).

Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). *Who are the crowdworkers?: shifting demographics in mechanical turk.* InCHI'10 Extended Abstracts on Human Factors in Computing Systems (pp. 2863-2872). ACM.

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C., Robinson, L.D., 2012. *Understanding Citizen Science & Environmental Monitoring.* Final Report on behalf of UK-EOF. NERC Centre for Ecology & Hydrology and Natural History Museum.

Rula, J. P., Navda, V., Bustamante, F. E., Bhagwan, R., & Guha, S. (2014, February). *No one-size fits all: Towards a principled approach for incentives in mobile crowdsourcing.* In: Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (p. 3). ACM.

Rural Payments Agency (2012). Annual Reports and Accounts 2011-2012. [Online] Available at: http://www.official-documents.gov.uk/document/hc1213/hc02/0230/0230.pdf Accessed 12/5/13

Saadat, H., Adamowski, J., Bonnell, R., Sharifi, F., Namdar, M., & Ale-Ebrahim, S. (2011). Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 66* (5), 608-619.

Schepaschenko, D., Fritz, S., McCallum, I., Shvidenko, A., Perger, C., Schill, C., See, L., Kraxner, F., Obersteiner, M., 2012. A platform to visualize, analyze and improve biomass datasets: http://biomass.geo-wiki.org. *Folia Forestalia Pol. Ser. A* 54, 137–139.

Schnebele, E., & Cervone, G. (2013). Improving remote sensing flood assessment using volunteered geographical data. *Natural Hazards and Earth System Science, 13* (3),669-677.

Schroeder, P. (1996). *Criteria for the design of a GIS/2*. Specialists' meeting for NCGIA Initiative 19: GIS and society, Summer 1996.

Scottish Natural Heritage, 2012. *Climate Change and Nature in Scotland.* Scottish Natural Heritage, Battleby.

See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., ... & Obersteiner, M. (2013). Comparing the quality of crowdsourced data contributed by expert and non-experts. *PloS one*, 8(7).

Seeger, C. J. (2008). The role of facilitated volunteered geographic information in the landscape planning and site design process. *GeoJournal, 72* (3-4), 199-213.

Shamir, L., Yerby, C., Simpson, R., von Benda-Beckmann, A. M., Tyack, P., Samarra, F., & Wallin, J. (2014). Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *The Journal of the Acoustical Society of America, 135* (2), 953-962.

Shaw, A. D., Horton, J. J., & Chen, D. L. (2011*). Designing incentives for inexpert human raters.* In Proceedings of the ACM 2011 conference on Computer supported cooperative work (pp. 275-284). ACM.

Sheng, V. S., Provost, F., & Ipeirotis, P. G. (2008). *Get another label? improving data quality and data mining using multiple, noisy labelers*. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 614-622). ACM.

Shermer, M. (2002) In Darwin's Shadow: *The Life and Science of Alfred Russel Wallace: A Biographical Study on the Psychology of History, Oxford*: Oxford University Press.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in ecology & evolution, 24* (9), 467-471.

Spyratos, S., & Lutz, M. (2014). *Characteristics of Citizen-contributed Geographic Information*. In: AGILE 2013

Stehman, S. V., Olofsson, P., Woodcock, C. E., Herold, M., & Friedl, M. A. (2012). A global land-cover validation data set, II: Augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing, 33* (22), 6975-6993.

Suau-Sanchez, P., Burghouwt, G., & Pallares-Barbera, M. (2014). An appraisal of the CORINE land cover database in airport catchment area analysis using a GIS approach. Journal of Air Transport Management, 34, 12-16.

Sui, D. (2008). The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems, 32*, 1–5.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., ... & Fink, D. (2014). The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, *169*, 31-40.

Tallis, H., Mooney, H., Andelman, S., Balvanera, P., Cramer, W., Karp, D., ... & Walz, A. (2012). A global system for monitoring ecosystem service change. *Bioscience, 62* (11), 977-986.

Tang, W., Zhuang, H., & Tang, J. (2011). *Learning to infer social ties in large networks. In: Machine Learning and Knowledge Discovery in Databases* (pp. 381-397). Springer Berlin Heidelberg.

Teacher, A. G., Griffiths, D. J., Hodgson, D. J., & Inger, R. (2013). Smartphones in ecology and evolution: a guide for the app-rehensive. *Ecology and evolution, 3* (16), 5268-5278.

Thomlinson, J. R., Bolstad, P. V., & Cohen, W. B. (1999). Coordinating methodologies for scaling land cover classifications from site-specific to global: steps toward validating global map products. *Remote Sensing of Environment, 70,* 16 – 28.

Tirkkonen-Condit, S. (1990). A Think-Aloud Protocol Study. *In Learning, Keeping and Using Language: Selected papers from the Eighth World Congress of Applied Linguistics, Sydney,* 16 21 August 1987 (Vol. 2, p. 381). John Benjamins Publishing.

Torres, M., & Qiu, G. (2014). Automatic habitat classification using image analysis and random forest. *Ecological Informatics, 23,* 126-136.

Townsend, P. A., Lookingbill, T. R., Kingdon, C. C., & Gardner, R. H. (2009). Spatial pattern analysis for monitoring protected areas. *Remote Sensing of Environment, 113* (7), 1410-1420.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., & Wolfe, R. E. (2012). Global characterization and monitoring of forest cover using

Landsat data: opportunities and challenges. *International Journal of Digital Earth, 5* (5), 373-397.

Tuia, D., Muñoz-Marí, J., & Camps-Valls, G. (2012). Remote sensing image segmentation by active queries. *Pattern Recognition, 45* (6), 2180-2192.

Tuia, D., Pasolli, E., & Emery, W. J. (2011). Using active learning to adapt remote sensing image classifiers. *Remote Sensing of Environment, 115* (9), 2232-2242.

Tuia, D., & Munoz-Mari, J. (2013). Learning User's Confidence for Active Learning. *Geoscience and Remote Sensing, IEEE Transactions on, 51* (2), 872-880.

Tsendbazar, N. E., deBruin, S., Defourny, P., & Herold, M. (2014). *Comparative analysis of global land cover datasets.* In: *Book of abstracts of the International Conference Global Vegetation Monitoring and Modelling (GV2M*) (pp. 58-58).

Tweddle, J. C., Robinson, L. D., Pocock, M. J. O., & Roy, H. E. (2012). *Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK*. NERC/Centre for Ecology & Hydrology.

Uden, M., & Zipf, A. (2013). *Open Building Models: Towards a Platform for Crowdsourcing Virtual 3D Cities. In Progress and New Trends in 3D Geoinformation Sciences* (pp. 299-314). Springer Berlin Heidelberg.

UN-GGIM (2013) Future Trends in Geospatial Management. [Online] Available at: http://ggim.un.org/docs/meetings/2ndHighLevelForum/UN-GGIM%20Future%20Trends%20Paper%20-%20Version%202.0.pdf [Accessed: 05/08/14]

Van Zanten, B. T., Verburg, P. H., Espinosa, M., Gomez-y-Paloma, S., Galimberti, G., Kantelhardt, J., ... & Viaggi, D. (2014). European agricultural landscapes, common agricultural policy and ecosystem services: a review. *Agronomy for Sustainable Development, 34* (2), 309-325.

Verburg, P. H., Neumann, K., & Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biology, 17* (2), 974-989.

Vitousek, P. M. (1994). Beyond global warming: ecology and global change. *Ecology, 75* (7), 1861-1876.

Walter, V. (2004). Object-based classification of remote sensing data for change detection. *ISPRS Journal of photogrammetry and remote sensing, 58* (3), 225-238.

Wasson, K., Woolfolk, A., & Fresquez, C. (2013). Ecotones as indicators of changing environmental conditions: rapid migration of salt marsh–upland boundaries. *Estuaries and Coasts, 36* (3), 654-664.

Wiggins, A., & Crowston, K. (2011). *From conservation to crowdsourcing: A typology of citizen science. In System Sciences (HICSS)*, 2011 44th Hawaii international conference on (pp. 1-10). IEEE.

Woodcock, C. & Gopal, S. (2000), Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science, 14*, 153-172.

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., ... & Fay, D. (2011). Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?. *International Journal of Digital Earth, 4* (4), 305-329.

Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., ... & Gong, P. (2014). Meta-discoveries from a synthesis of satellite-based land-cover mapping research*. International Journal of Remote Sensing, 35* (13), 4573-4588.

Zhai, Z., Hachen, D., Kijewski-Correa, T., Shen, F., & Madey, G. (2012). Citizen engineering: *Methods for" crowdsourcing" highly trustworthy results. In System Science (HICSS)*, 2012 45th Hawaii International Conference on (pp. 3406-3415). IEEE.

Zhang, Y., Liu, X., Chang, M. C., Ge, W., & Chen, T. (2012). *Spatio-temporal phrases for activity recognition. In Computer Vision–ECCV* 2012 (pp. 707-721). Springer Berlin Heidelberg.

Appendix 1: Image Classification Experiment Results: Image 1

Image	Participant	Group	Time	Clicks	Broad Class	Actual Broad	Sub Class	Other Class	B Certainty	S Certainty	Extra
1	1	None	128.497	36	22	43	23	A1	95	5	
1	2	Ameteur Enthusiast	184.712	13	22	43	26		63	44	
1	3	Ameteur Enthusiast	372.155	44	43	43	50	A	70	60	
1	4	Ecologist	285.476	23	43	43	44	A2.2	70	9	dry, upland, grassland
1	5	Ecologist	248.816	17	1	43	11	J1.4	40	29	
1	6	Student	166.741	11	1	43	15	b2	61	60	meadow
1	7	Ameteur Enthusiast	252.95	11	1	43	14		40	40	
1	8	Ecologist	137.624	20	81	43	82		50	30	Not sure if I'm familiar enough with UK.
1	9	None	6.476	16	22	43	27	A2 SCrub	81	50	
1	10	None	90.586	8	43	43	45	A2. 1	60	19	
1	11	Student	206.881	18	1	43	11	B	60	40	Green, grassy, quite bare
1	12	Student	314.978	24	22	43	23	C	88	30	

1	13	Ameteur Enthusiast	133.11	19	22	43	24	B1.2	73	38	Calcifuge
1	14	Ecologist	67.451	13	1	43	16		60	59	
1	15	Ecologist	620.076	40	22	43	33	J1.3, J2?	71	50	Agriculturally influenced
1	16	None	235.505	12	22	43	34		30	0	
1	17	Bird	62.965	16	22	43	25	A2.2, A1	91	82	
1	18	Ecologist	191.352	9	22	43	34		69	30	
1	19	Ecologist	179.609	14	43	43	44		41	20	
1	20	Ameteur Enthusiast	24.883	7	22	43	27		58	53	
1	21	Ecologist	135.85	6	43	43	44	B1.1	83	79	
1	22	Ecologist	119.238	8	22	43	27	D1	70	80	Meadow
1	23	Ameteur Enthusiast	32.192	29	1	43	13		62	71	
1	24	Bird	22.769	10	43	43	50	A1.3, D1.1	81	70	
1	25	Ameteur Enthusiast	142.73	33	22	43	26	B2	80	80	Grassland /
1	26	Ameteur Enthusiast	350.85	26	43	43	44		75	15	
1	27	None	24.64	20	22	43	34		81	50	
1	28	None	650.93	32	43	43	50	A2.2	25	23	

1	29	Student	450.21	31	1	43	11	?	20	21	
1	30	None	328.86	20	43	43	50		86	70	
1	31	None	85.85	3	43	43	50		20	20	
1	32	None				43					
1	33	Ameteur Enthusiast				43					
1	34	Ecologist	222.96	16	43	43	50	A2.2, C1.1	91	81	Heavily grazed
1	35	Ecologist	244	7	22	43	28		100	47	It's certainly a grassland community.
1	36	Ecologist				43					
1	37	Student	250.54	11	43	43	44	B2	70	30	Mixed shrubland and grassland, bunch grasses,
1	38	None	173.34	9	22	43	28	B2	100	80	
1	39	Ecologist	94.706	23	22	43	26	?	80	10	
1	40	None				43					
1	41	Student				43					
1	42	Student				43					
1	43	Ecologist			43	43	50	A2	71	30	
1	44	Ameteur Enthusiast	40.22	7	22	43	27		70	29	

1	45	Bird	179.92	10	43	43	45	J1.4	20	0	
1	46	Ecologist	191.03	19	22	43	26	?	90	75	Grassy plain grazed by wild animals.
1	47	Student	150.31	13	22	43	26		100	55	Field
1	48	Student	15.67	11	22	43	27		49	20	Possible pasture
1	49	Student				43					
1	50	Student	160.63	14	22	43	26		71	8	
1	51	Student	22.65	7	43	43	44	A2.1, A2.2	60	50	
1	52	Ecologist	93.38	6	22	43	24		92	71	
1	53	Ecologist	311.21	28	1	43	13		71	83	
1	54	Ecologist	137.78	9	22	43	28		81	30	
1	55	Ecologist	256.03	31	22	43	30	A1.3	52	23	Looks like overgrazed grassland
1	56	None				43					
1	57	Ameteur Enthusiast				43					
1	58	Student	3097.31	37	43	43	50	B6?, B1.2?, D5?	60	29	well-drained soil. grasses,
1	59	Bird				43					

1	60	Ameteur Enthusiast	119.637	11	43	[43]	45	C2	64	30	
1	61	Bird	627.113	21	43	[43]	45	?	100	80	sparsely wooded
1	62	Bird	108.526	9	43	[43]	50	A	100	91	Fern invasion
1	63	Ameteur Enthusiast				[43]					
1	64	Student	196.933	13	22	[43]	25		100	30	managed grassland
1	65	Bird	255.23	14	43	[43]	50		70	30	
1	66	Ameteur Enthusiast				[43]					
1	67	None	452.679	12	22	[43]	26		72	19	
1	68	Ecologist				[43]					
1	69	None	368.748	30	1	[43]	13	A2.2	64	32	
1	70	None	281.274	15	1	[43]	12		9	9	
1	71	Bird	623.38	9	22	[43]	28		50	39	
1	72	None				[43]					
1	73	Ecologist	163.907	15	43	[43]	50	D1.1	80	80	
1	74	Ecologist				[43]					
1	75	Ameteur Enthusiast	281.802	15	22	[43]	34	A?	91	60	Open grazing landscape with
1	76	Ecologist				[43]					
1	77	Student	390.137	25	22	[43]	30	A2.2	90	65	The habitat is dominated by two

											different types of grass, which are grouped together in patches
1	78	Ecologist				43					
1	79	Ecologist	344	21	43	43	50	?D1.2, ?B1.2, ?B2.2	80	60	Heath / grassland mosaic
1	80	Student				43					
1	81	None				43					
1	82	Bird	376	6	22	43	25		100	60	
1	83	Bird				43					

.

N.B. the coding used to represent the land cover classes is explained in Tables 5.9 and 5.10