Assesment of Environmental Stewardship Scheme agreements using the sentiment expressed in trail users' tweets. An exploratory analysis of the Pennine Way National Trail, England.

• Tom Wilson1 • Robin Lovelace1 • Andrew Evans1

**Abstract** The Environmental Stewardship Scheme (ESS) is the most widespread approach to environmental management in England. ESS provides government-financed payments to farmers or land managers in return for an environmentally sensitive approach to farming. Currently ESS has agreements in place on over 70% of agricultural land in England.

England has 15 designated National Trails which pass through diverse landscapes and expanses of agricultural land, a substantial amount of which is managed under ESS. Despite the interactions that trail users have with England’s landscapes, and inadvertently with land managed under ESS, their opinions are not currently considered with regard to the effectiveness of ESS in the maintenance and enhancement of the landscape.

This research presents methods to extract the sentiment contained within trail users' tweets with the aim of reducing this knowlege gap. Here we present an exploratory analysis into the feasibility of using the sentiment expressed within trail users' Twitter messages (tweets) to assess the effectiveness of Environmental Stewardship Scheme agreements in place along the Pennine Way National Trail, England.

**Keywords** big data analysis • sentiment analysis • Environmental Stewardship Scheme • Volunteered Geographic Information • National Trails • social media

# Introduction

The Environmental Stewardship Scheme (ESS) is one of the the most widespread approaches to environmental management of farmland in England (Franks and Emery 2013). ESS provides financial rewards to farmers and land managers in return for reductions in farming intensity and the adoption of measures to protect the surrounding environment. The success of early agri-environmental schemes was measured by levels of participation but more recently, and for the lifetime of ESS which was introduced during 2005 and 2006, the focus has shifted to analyse the environmental benefits provided under the scheme (Franks and Emery 2013), for example with regard to landscape character (Natural England 2014a), the enhancement of grassland, moorland and heath (Natural England 2014b), bird populations (Davey et al. 2010), and the provision of ecocsystem services (Department of Food and Rural Affairs 2009).

England has 15 designated National Trails which pass through diverse landscapes and expanses of agricultural land, a substantial amount of which is managed under ESS. In 2012 approximately 12 million visits were made to England’s National Trails (Ramblers, 2012). Despite the interactions that trail users unwittingly have with land managed under ESS there is currently no method to specifically obtain their opinions concerning the effectiveness of ESS in preserving and protecting the environment. The opinions of trail users are in fact limited to broad, large-scale qualitative surveys of visitors to the countryside in general, such as the Monitor of Engagement with the Natural Environment (MENE) which examines the adult population’s engagement with the natural environment (Natural England, 2013). The National Trail User Surveys (The Countryside Agency, 2005; Natural England/Countryside Council for Wales, 2007) were discontinued in 2007.  
\_\_\_

Tom Wilson: [[gy10tlw@leeds.ac.uk](mailto:gy10tlw@leeds.ac.uk)](mailto:gy10tlw@leeds.ac.uk)

1 School of Geography, University of Leeds, Leeds, United Kingdom

The focus of this research is the 431 km (268 mile) long Pennine Way National Trail (PWNT), the oldest of England’s 15 National Trails which officially opened on 24th April 1965 (Figure 1). Along its route the PWNT crosses agricultural land managed under ESS. For this analysis a 5km spatial buffer was drawn around the PWNT which will hereafter referred to as the PWNT corridor.

This research aims to address this potential discontinuity of knowledge by exploring the feasibility of utilising the sentiment conveyed in trail users’ Twitter messages (tweets) to assess the effectiveness of ESS specifically from the trail user perspective. This exploratory analysis will focus on the Pennine Way National Trail. We propose a methodology to geographically and lexically filter relevant tweets that originate from the proximity of the trail, and then perform sentiment analysis to extract the sentiment expressed. Finally we will determine whether this information can be used to assess the effectiveness of ESS.

# Background

* Environmental Stewardship Scheme
  + what it is (larggest agri-environmental scheme in England). Broad and shallow approach. 5 main objectives:
    - Conservation of wildlife and their habitats
    - Maintenance and enhancement of landscape quality and character
    - Protection of the historic environment
    - Protection of soils and reduction of water pollution
    - Providing opportunities for people to visit and learn about the countryside.
* ESS aims to provide the funding and guidance to enable farmers and land managers to fulfil these objectives
  + how it works (payments to landowners for managing their land in an environmentaly concious way)
  + histroy of the scheme
    - Government’s response to increasing levels of agricultural intensification and its negative impacts on wildlife and landscape character. ESS follows on from Countryside Stewardship Scheme, is more flexible, and a broad and shallow approach (open to all).
  + significance of ESS in the context of the National Trail system
    - ESS agreements within the corridor of a National Trail have a particularly important role to play in providing positive experiences to users of the trail. One of the primary objectives of ESS is the maintenance and enhancement of landscape quality and character. Furthermore, Natural England identifies enhancement of the landscape, natural, and historic features within the trail corridor as one of its quality standards of the National Trail system.
    - Current methods to determine trail users' opinions
      * large scale qualitative surveys such as Monitor of Engagement with the Natural Environment (MENE): annual representative quota-sampled survey of ~50k people. Very broad definition of natural environment.
* Sentiment Analysis
  + what is it?
    - Seeking the sentiment conveyed by others is fundamental to the decision-making of individuals and organisations alike. As a consequence, sentiment is seen as a key influencer of action, and central to human behaviour (Liu 2012)
  + why conduct sentiment analysis (current uses and purpose)
    - In short there is a human desire to know what other people think.
  + approaches to sentiment analysis (lexical, machine learning)
  + sentiment analysis of social web challenges (abbreviations, slang, poor grammar)
  + SentiStrength tool used in this project (advantages and disadvantages, suitability for this research)
    - SentiStrength was developed by Thelwall et al., University of Wolverhampton <http://sentistrength.wlv.ac.uk>
* Social web data (big data) - Do we need to provide background about this? Robin to contribute stuff here

# Data

* 60,434 geocoded tweets collected between 2014-06-03 and 2014-07-25
  + The Twitter data for this research was acquired through a tweet-harvesting project conducted at the University of Leeds - Robin to add a little more about this process.
* A comma separated file of Twitter data with each row of the file representing a single instance of a tweet. As well as the actual text of the tweet (TweetText), the dataset also included additional metadata about each tweet; a unique id (TweetID), the date the tweet was created (DateCreated), the time the tweet was created (TimeCreated), the number of followers of the sender (tweeter) (n\_followers), the number other tweeters the tweeter was following (n\_following), the number of tweets tweeted by the user (n\_tweets), and user’s location (user\_location). The user’s location does not refer to the user’s location at the time of sending the tweet, rather it refers to their self-disclosed location extracted from their twitter profile. The dataset did not include a Twitter username or any other personal information. The location from which the tweet originated was provided by the geotag fields in the dataset; longitude and latitude. Every tweet in the dataset included this geocoded information, which represents approximately 1-3% of all tweets (Morstatter et al., 2013; Broniatowski et al., 2013; Hecht and Stephens, 2014). As received the twitter dataset represented 52 days of data collection between 2014-06-03 and 2014-07-25 inclusive and contained a total of 60,466 geotagged tweets and their associated metadata. Figure 4 illustrates the spatial distribution of the tweets within the Twitter dataset.
* GPX file of the route of the Pennine Way National Trail (PWNT) (Walk Unlimited 2014)
  + A GPX file of the route of the Pennine Way National Trail was downloaded from the official website of England’s National Trails (Walk Unlimited, 2014). The GPX file contained a waypoints layer and a route layer. It was the route layer which was used as the basis of this analysis. Using the open-source statistical package R (R Development Core Team, 2008), the route layer of the PWNT was plotted and a 5km buffer drawn around it. This 5km buffer presents the geographical scope of the project and is hereafter referred to as the PWNT corridor. A 25km buffer was also drawn for use in later viewshed analyses. Figure 3 illustrates the geographic scope of these buffers.
* Shapefile of ESS agreement boundaries in England (Natural England 2014c)
  + A shapefile of ESS agreements was downloaded from the Environment Agency Geostore (Natural England, 2014e). This file was a vector dataset that contained the boundaries of all the ESS agreements in England; a total of 48,285. Each agreement also an associated entry in the geodatabase which included the level of the agreement (e.g. ELS, HLS), and other information such as details about the farm, the duration of the agreement, etc. The ESS shapefile was spatially clipped to only include the ESS agreements that were completely or partly within 5km and 25km PWNT corridors. Table 1 provides a breakdown of the types of ESS agreements in place, and the prevalence of ESS in both the 5km and 25km PWNT corridors. Figure 3 illustrates the spatial distribution of ESS agreements within the PWNT corridors.
* Land Cover Map 2007
  + clipped to 5km and 25km PWNT buffers
* Digital Elevation Model
  + clipped to 5km and 25km PWNT buffers
  + low resolution (90m) SRTM
* PWNT trail counter data (provided by Natural England)

# Methods

* Spatial selection
  + in R the Twitter dataset was reduced to only the tweets which originated from within the PWNT 5km corridor
* Lexical selection
  + in R the tweets relevant to hiking the PWNT were selected using natural language processing
* The text of each tweet was processed and ‘cleaned’ of spurious characters
* Each tweet and its spatial information was imported into ArcMap 10 (ESRI, 2011)
* The TweetText and TweetID were input into SentiStrength for sentiment analysis
* Sentiment analysis outputs were combined with tweet spatial information in ArcMap 10/QGIS
* Viewshed analyses were conducted for each overall positive and overall negative tweet. Viewshed analyses were conducted in ArcMap 10 (ESRI, 2011) and included:
  + Calculation of the viewshed
  + Determination of majority land cover class within the viewshed
  + Determination of the ruggedness within the viewshed
  + Determination of the presence of ESS agreements within the viewshed.

# Results

* The sentiment expressed in the twitter messages along the PWNT trail corridor:
  + 47 tweets expressed positive sentiment,
  + 22 negative sentiment
  + 102 expressed no sentiment (i.e. neutral).
  + 10 tweets expressed both positive and negative sentiment.
  + Overall tweet sentiment consisted 40 positive, 16 negative and 105 neutral.
  + 105 tweets did not convey any sentiment and were classed as neutral. It was discovered that of these 105 tweets, 94 contained a URL within the TweetText.
    - Missing sentiment in images?
  + [Link to interactive map of trail user tweets](http://tom-wilson.info.s3-website-us-west-2.amazonaws.com/PennineTweets.html)

# Discussion

There are several issues with social web data that deserve attention. First is the accessibility to the data. To use Twitter as an example, none of the public APIs provide direct, unfettered access to Twitter data. The Twitter API is believed to be subject to a ‘streaming cap’ of about 1% of all tweets at any point in time (Driscoll and Walker, 2014). In reality it is only the social web companies themselves that have full access to the data (Manovich, 2011), and at the same time have full control as to who can access the data (boyd & Carwford, 2012). So of any data collected using the public Twitter API there also exists an additional ~99% of data in not accessible. This data cannot be accounted for because of a lack of transparency regarding the exact streaming cap, and the process of selecting which tweets are available via the API (boyd & Crawford, 2012). Nevertheless is important that this is recognised within the research.

Second is the representation of social web data. Twitter usage requires access to the internet via a desktop computer or a mobile device. As such, Twitter usage is limited to internet users. Moreover Twitter usage is not evenly distributed among internet users (Driscoll and Walker, 2014). Specific to this research, not all trail users are necessarily internet users, and even those that are internet users may not wish to tweet when out hiking. Furthermore, a person needs both a smartphone and a data plan in order to send a tweet whilst out hiking on the PWNT. Consideration of this is needed to avoid the creation of a ‘digital divide’ whereby only trail users with smart phones and data plans are heard.

Third is the question of ethics. As boyd and Marwick (2011) succinctly put it; “there is a considerable difference between being in public and being public” (boyd & Crawford, 2012 p673). Although Twitter data may be classed as public data, consideration should be given to the subjects of the study. Twitter users should understand that their data is public (unless they specify otherwise in their preferences), but there is the chance that they may not. Even if they do know their data is public they may not intend for their data or tweet to become public (Eckert et al., 2013). The Twitter dataset in this study did not contain Twitter usernames or personal information, and no tweets have been published in this report. It may be necessary to disclose the purpose of data collection in the pursuit of a social web opinion-mining campaign.

# Reccomnedations

A process to select trail users’ tweets from a larger dataset of Twitter and extract the sentiment conveyed has been developed. The exploratory analysis of the data used in this research did not provide conclusive results with regard to the effectiveness of ESS, but it did uncover interesting insights which deserve further attention.

As previously mentioned, an interesting avenue of future research is to determine the extent of image-sharing in trail users’ tweets. Furthermore, are tweeting trail users attempting to convey sentiment through these images? Borth et al. (2013) present findings of visual sentiment ontology which could provide a foundation of future research in this area.

Based upon the findings of this research it is recommended that Natural England proactively initiate a social media strategy as a method of eliciting the sentiment of its trail users from their social web data. Natural England should select a hashtag with which it would like users to tag their tweets. This research has uncovered that trail users already utilise hashtags within their tweets. Assigning a hashtag specific to this campaign will facilitate with grouping and selection of tweets during data analysis. Furthermore, the hashtag can form the basis of a promotional and educational campaign designed to inform users of the purpose and mechanism of the campaign.

Promotion of the campaign would also provide the opportunity to increase awareness of the use of social web data for this purpose. This is important from both an ethical and representative perspective: Trail users should be alerted to the fact that the sentiment they convey and comments they make are public. In terms of representation, a greater number of people need to be encouraged to participate in this scheme for it to be anywhere close to representative, and ensure that trail user opinions are not subject to a digital divide whereby only those with a smart phone and a data plan are able to offer their opinion.

Initiation of a social media campaign is also likely to increase the amount of data available for analyses such as those presented in this report, and allow for the process to be refined further.

# Conclusions

This research has demonstrated a methodology to filter relevant tweets and perform a sentiment analysis on short informal texts sent by trail users along the Pennine Way National Trail. With regard to the feasibility of assessing the effectiveness of ESS agreements the initial findings are limited, both in terms of the number of relevant tweets and the sentiment conveyed. Additional work is needed to collect additional data, perhaps over a longer period of time. A key finding is that a high percentage of neutral tweets contained a URL to an image and it could be that these images are intended to convey sentiment. Further research should therefore focus on developing methods to extract the sentiment, if any, conveyed in these images.

# ADDITIONS:

* statistics in results
* broader potential impact of study
* section regarding social media analysis in general
* policy reccomendations

# References

Davey, Catherine M, Juliet A Vickery, Nigel D Boatman, Dan E Chamberlain, Hazel R Parry, and Gavin M Siriwardena. 2010. “Assessing the Impact of Entry Level Stewardship on Lowland Farmland Birds in England.” *Ibis* 152 (3). Wiley Online Library: 459–74.

Department of Food and Rural Affairs. 2009. “Provision of Ecosystem Services Through the Environmental Stewardship Scheme.” Defra project code NR0121.

Franks, J R, and S B Emery. 2013. “Incentivising Collaborative Conservation: Lessons from Existing Environmental Stewardship Scheme Options.” *Land Use Policy* 30 (1): 847–62.

Liu, B. 2012. “Sentiment Analysis and Opinion Mining.” *Synthesis Lectures on Human Language Technologies* 5 (1): 1–167.

Natural England. 2014a. “Assessment of the Effects of Environmental Stewardship on Landscape Character. Correlative Analysis of Datasets to Assess the Degree of Success in Delivery of Environmental Stewardship Objectives.” *Natural England Commissioned Reports Number 158*.

———. 2014b. “Assessment of the Effect of Environmental Stewardship on Improving the Ecological Status of Grassland, Moorland and Heath.” *Natural England Commissioned Reports Number 156*.

———. 2014c. “Environment Stewardship Scheme Agreements.” *[Shapefile] Created by Natural England, August 2014*. <http://www.geostore.com/environment-agency/WebStore?xml=environment-agency/xml/ogcDataDownload.xml>.

Walk Unlimited. 2014. “The Pennine Way.” <[online] Available at: http://nationaltrail.co.uk/pennine-way. Accessed 7 January 2015>.