

Moving into the information age: From records to Google Earth

David R. R. Smith Psychology, School of Life Sciences, University of Hull
e-mail: davidsmith.butterflies@gmail.com

Introduction

Many of us are avid recorders of Yorkshire' varied flora and fauna. Over a recording career of many years a single recorder can personally amass large data sets or, if involved with networks of similarly-minded people, groups of recorders can rapidly build truly massive data sets numbering many tens of thousands of records. Sometimes we wish to make sense of these large data sets by mapping their geographic location, perhaps to understand distributional trends in space across time. Though data sets can be easily mapped in software such as MapMate or Levana, the resulting maps are basic outline projections of the reference region on which records are plotted (Figure 1). Though useful for understanding distribution in the abstract they lack the directness (and ability to zoom in and out) offered by scaleable photographic representations of location as provided by satellite imagery embedded and rendered in a digital environment. As County Butterfly Recorder (Butterfly Conservation) for Yorkshire I wished to make use of the software Google Earth, which is the pre-eminent example of scaleable mapping software, to map butterfly sightings across a number of years. The object was to motivate recorders in the coming butterfly-recording year to go out and explore those parts of Yorkshire which had not so far provided any butterfly sightings. The problem was how to take almost 75,000 records, representing just two years worth of butterfly sightings, currently residing in the butterfly software Levana and plot them in Google Earth. This turned out to be a relatively painless journey but requiring some tricks that I thought it would be useful to share. I would hope that this article will allow other people with data bases of records that they would like to present in Google Earth to do exactly that.

===== INSERT FIGURE 1 ABOUT HERE =====

Method

Stage 1 We start with whatever system your current records are in. I keep my butterfly records in Levana – others may keep their records in MapMate or in Excel. Most recording software will offer an export function which takes internal data records and outputs them into another generic file format suitable for other software to read. A common output file format is called .csv where data are separated by commas. Levana takes records consisting of dates, grid references, site names, species, number seen, recorder etc and outputs all these as tab-separated values (.txt file). Levana's export function has several variables which can be set to select only those records which you are interested in exporting – I wanted to export all butterfly records in Yorkshire in 2015 and 2016, so I selected those for export. You should now have an output file (.txt format for Levana).

Stage 2 The next step is to import the exported file into Excel. If you keep your records in Excel spreadsheets already then skip the previous paragraph and this one! If not read on. Open Excel up and open the .txt file (your output file might be different say .csv which will depend on the original software you use. You will need to navigate to the correct folder and

make sure that you have told Excel to look for the file format your exported output file is in (setting to All Files (*.*) in Excel is the easiest way). Excel will ask you a series of questions about the file you are reading in related to the way it is structured – I have found that for .txt files accepting all the default answers to those questions gets the data into Excel. The questions ask consecutively whether data is delimited or fixed width (they are delimited), what the delimiter is (it's a tab), and what data format the columns are (they are 'general' being a mix of text, dates and numbers). You should have something like Figure 2.

===== INSERT FIGURE 2 ABOUT HERE =====

Stage 3 At this point the file has more columns than you need. I only want grid references. So delete all columns except the GridRef column. You should have something like Figure 3.

===== INSERT FIGURE 3 ABOUT HERE =====

Highlight the remaining column, go to the Data tab, click on Remove Duplicates and click on OK. What I have done is remove all entries that are duplicated – so the 300 visits to that popular butterfly location are reduced to just one entry. This results in just over 63,000 grid references being discarded from the Excel file and leaves just under 8,000 unique grid references.

Stage 3a (optional) If you only want to plot your sightings in Google Earth (at the resolution they are at) then this stage can be skipped. For my purposes, I wanted to put a pin in every monad (1 x 1km square) that had been visited so that by omission people would know which monads in Yorkshire had not returned a butterfly sighting. So I am faced with the problem that many of the 8,000 or so unique grid references in my Excel file are different locations *within* the same monad – I want to treat all locations *within* the same monad as being at the level of the monad. This means that higher resolution grid references (such as SE567453) need to be converted into lower six-figure resolution (such as SE5645). This is the tricky stuff.

First I need to know the length of all grid references. So SE3096 would be 6, SE309396 would be 8 and so on. Within Excel this is very easy. Just type =LEN(E3) in the cell in the neighbouring column. Hit return. This command means take whatever is in cell E3 and calculate its length. If you do not see a number but perhaps a date in cell F3, then make sure that the column F is selected as being in Number format rather than something else such as being in Date format (see the Number box circled in Figure 4).

===== INSERT FIGURE 4 ABOUT HERE =====

Left-click on the bottom right corner of the black box in cell F3 and drag the box down all the way down the column. The formula will be automatically updated so that what was done in the first row is done for all rows you drag the box down. This is the joy of Excel. You should

now have for every grid reference in column E a neighbouring entry in column F showing the length of the grid reference. In my Excel sheet this ranged from 6 (i.e., SE3034) to 12 (i.e., SE1428856637). This is resolution ranging from 1 x 1km to 1 x 1m.

You need to have all your grid reference lengths in the file sorted going from shortest to longest. Simply highlight column F and click on the Sort and Filter button in Excel, Sort Smallest to Largest and Expand the selection. The last bit – Expand the selection – is necessary if you have more than one column, in this example of just one column it is not critical.

The next step is to extract only the 1 x 1km part of all grid references. So if it is SE3034 you want SE3034 and if it is SE1428856637 you want SE1456. The best way to do this is decompose a grid reference into three components each in its own column: so letter code (i.e., SE), 1km easting number (i.e., 14) and 1km northing (i.e., 56) go in separate columns. In column G I shall extract the letter code, by typing =MID(E3,1,2) (you may need to replace the commas with semi-colons if you are using OpenOffice) in cell G3 and hitting return. This simply means look at whatever is in cell E3 and, starting from the first character, extract the first two characters. This will give SE. To extract the easting simply use the formula =MID(E3,3,2) in column H. This means look at whatever is in cell E3 and starting from the third character extract the first two characters. This would give 14. We need to be very careful when extracting the northing as the location of this varies depending on the length of the grid reference. If length 6 (i.e., SE3024), the northing is found by the formula =MID(E3,5,2). If length 8 (i.e., SE306389) the northing is obtained by =MID(E3,6,2). If length 10 (i.e., SE30673897) the northing is obtained by =MID(E3,7,2). If length 12 (i.e., SE3067238971) the northing is =MID(E3,8,2). This goes in Column I. You must be careful that you drag down the formula in Excel to match the length of the grid reference. Each time that you move from one length of grid reference to a new one you will need to manually alter the formula once before dragging the formula down further.

Finally you need to take the three components you have built (letter code, easting and northing) and build them back together to form a monad grid reference in a separate column J. Use the formula =(G3&H3&I3) to achieve this. You should have something like Figure 5. Regardless of whether you originally had grid references of different resolutions, they are all now converted into monads.

===== INSERT FIGURE 5 ABOUT HERE =====

We now have our grid reference data in a suitable format for the next stage which is to convert them into the format required for plotting in Google Earth.

Stage 4 Copy column J (or if you skipped Stage 3a and you just want to plot the exact location of your records then just copy your column of grid references). Go to the web page <http://gridreferencefinder.com/> which takes you to an excellent resource known as the UK Grid Reference Finder. At the very bottom of the web page there are several links. Click on the one labelled 'Coordinate Batch Conversion Tool'. This opens a page which essentially converts one kind of location code to latitude/longitude coordinates which can be plotted in Google

Earth. Paste the contents of Column J into the empty box space under ‘Step One – Paste your data to convert’. Then select the ‘Grid Reference’ radio button. Click on the ‘Step Five – Convert’ button. Click on ‘Step Seven – Export to Google Earth (KML) File’. This will save a file called GridReferenceFinder.kml on your computer as a Download. KML files are in a file format used to store geographic information in a way suitable for Google Earth. Rename it whatever you like but keep the .kml suffix.

Stage 5 Double-click on the .kml file. It will launch Google Earth and go to the location of your sighting placing a yellow pin at the location. You can zoom in and out around the area. That is it! If nothing happens then you probably don not have Google Earth installed on your computer. Go to http://www.google.co.uk/intl/en_uk/earth/ and download the appropriate version of Google Earth for your operating system and computer. Now double click on your .kml file.

Stage 6 Each yellow pin comes with a label Point 1, Point 2 and so on. This can become a problem if you have many pins because it clutters the map and seriously hampers the smoothness of movement within Google Earth. Go to the left side-bar in Google Earth and locate your My Places folder. You might have to expand the folder to see all the folders within My Places. You should be able to see lots of yellow pins each with a tick mark against it. This means that the pin is showing on the map – you want to keep the pin showing but make the label associated with the pin invisible. Right click the label for your yellow pin (Point 1) and select Properties. In the popup window you need to left click on the tab called ‘Style, Color’. Then change the Label scale to 0. You are still plotting the label but with a size of 0 so that you cannot see it (see Figure 6). Click on OK. As long as all the yellow pins are in the same folder (and if they are not then make sure they are by copying them into one folder) then, to make all the pins in the folder have invisible labels, right click on the folder containing all your pins and select Properties. In the ‘Style, Color’ tab make sure that you click ‘Share Style’, and then click OK.

===== INSERT FIGURE 6 ABOUT HERE =====

Extras You can do lots of wonderful further things with Google Earth but I do not want to over-burden this article. The missing squares for butterflies which I developed can be found at <http://www.yorkshirebutterflies.org.uk/missingsquares.html> This is a slightly more sophisticated version with the five VCs outlined in red, thousands of sightings located at different monads on a regular grid, with OS grid references coordinate lines, all parcelled up into one big file (called a .kmz file). KMZ files are many kml files that are compressed into one convenient package file. Figure 7 shows an example screen shot showing the whole of Yorkshire.

===== INSERT FIGURE 7 ABOUT HERE =====

Acknowledgements

My thanks to Paul Millard for assisting in my first attempts at mapping and directing me towards the gridreferencefinder website and its handy way to convert data into .kml format.

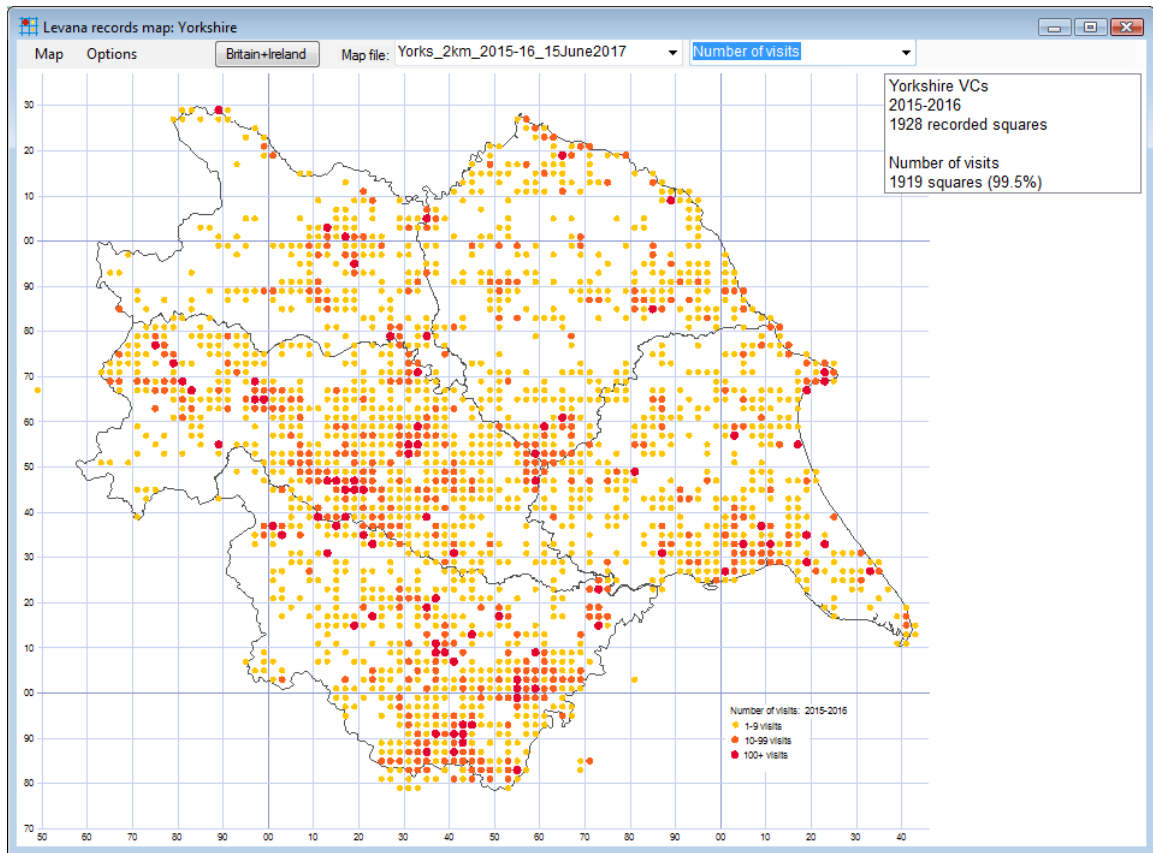


Figure 1. Butterfly sightings displayed on a map in Levana.

	A	B	C	D	E	F	G	H	I
1	Export from Levana files on 15 June 2017								
2	Recorder	Site	County	Habitat	GridRef	Date	Species	Adults	Males
3	confidential	Leeds, Moor Allerton			SE309396	06/08/2015	Small White	1	
4	confidential	Leeds, Moor Allerton			SE309396	06/08/2015	Red Admiral	1	
5	confidential	Leeds, Moor Allerton			SE309396	06/08/2015	Small Tortoisel	4	
6	confidential	Leeds, Moor Allerton			SE309396	06/08/2015	Meadow Brow	1	
7	confidential	Leeds, Wellington Hill			SE354383	08/08/2015	Small White	2	
8	confidential	Leeds, Wellington Hill			SE354383	08/08/2015	Red Admiral	1	

Figure 2. Results of reading in the exported file into Excel. Entries in the recorder name column have been relabelled for confidentiality.

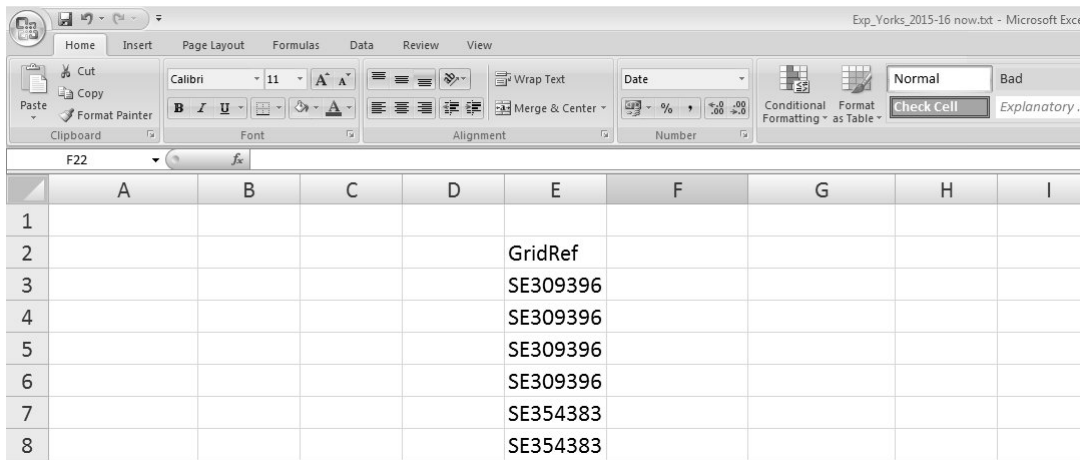


Figure 3. Only the GridRef column has been left.

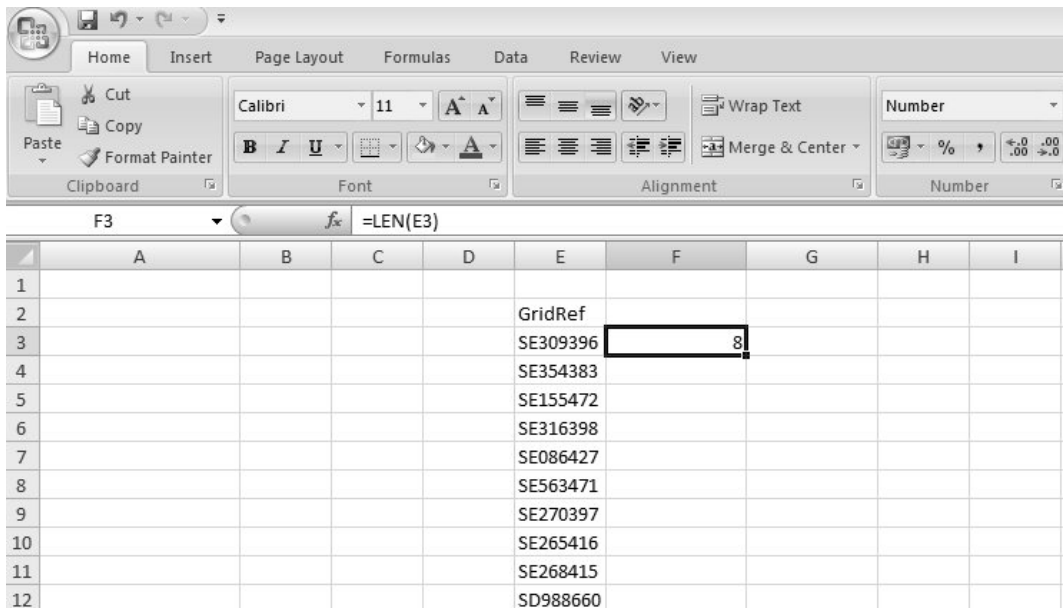


Figure 4. Result after typing in the first formula to calculate the length of the grid reference.

	A	B	C	D	E	F	G	H	I	J	K
5415					SE06634989	10 SE	06	49		SE0649	
5416					SE06604991	10 SE	06	49		SE0649	
5417					SE3449539835	12 SE	34	39		SE3439	
5418					SE3376892024	12 SE	33	92		SE3392	
5419					NY8500626806	12 NY	85	26		NY8526	
5420					SE3264291859	12 SE	32	91		SE3291	
5421					SE3350379798	12 SE	33	79		SE3379	
5422					NZ1762601960	12 NZ	17	01		NZ1701	
5423					NZ1844900332	12 NZ	18	00		NZ1800	
5424					SE2705987939	12 SE	27	87		SE2787	
5425					SE2163497674	12 SE	21	97		SE2197	
5426					SE1826197922	12 SE	18	97		SE1897	
5427					SE1229687875	12 SE	12	87		SE1287	

Figure 5. Here we have decomposed the original grid reference (Column E) into three separate components (Columns G to I) and then rebuilt the grid reference at monad level (column J).

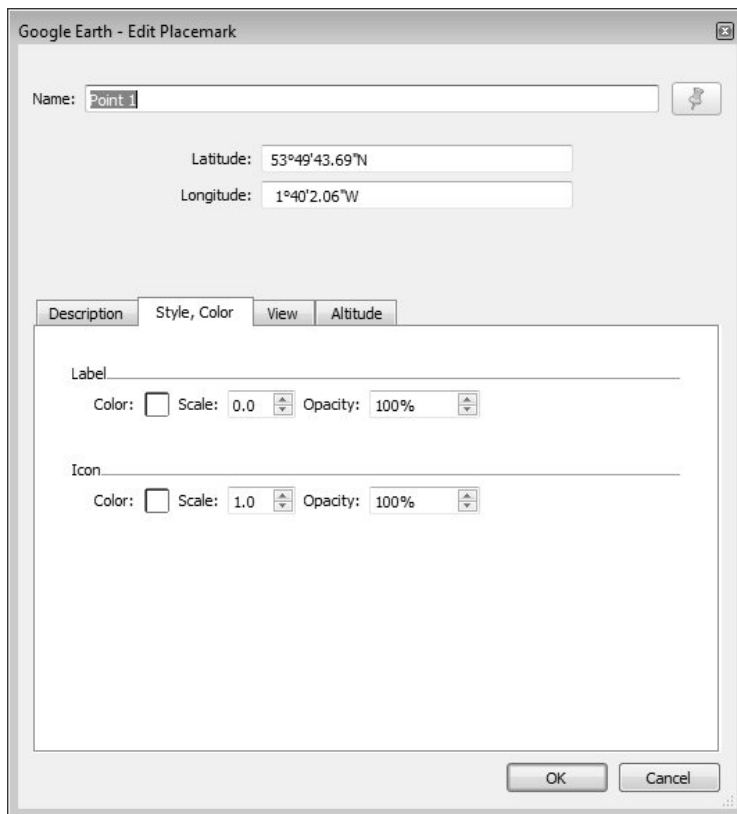


Figure 6. Plotting invisible labels requires you to set the Label Scale value to 0.

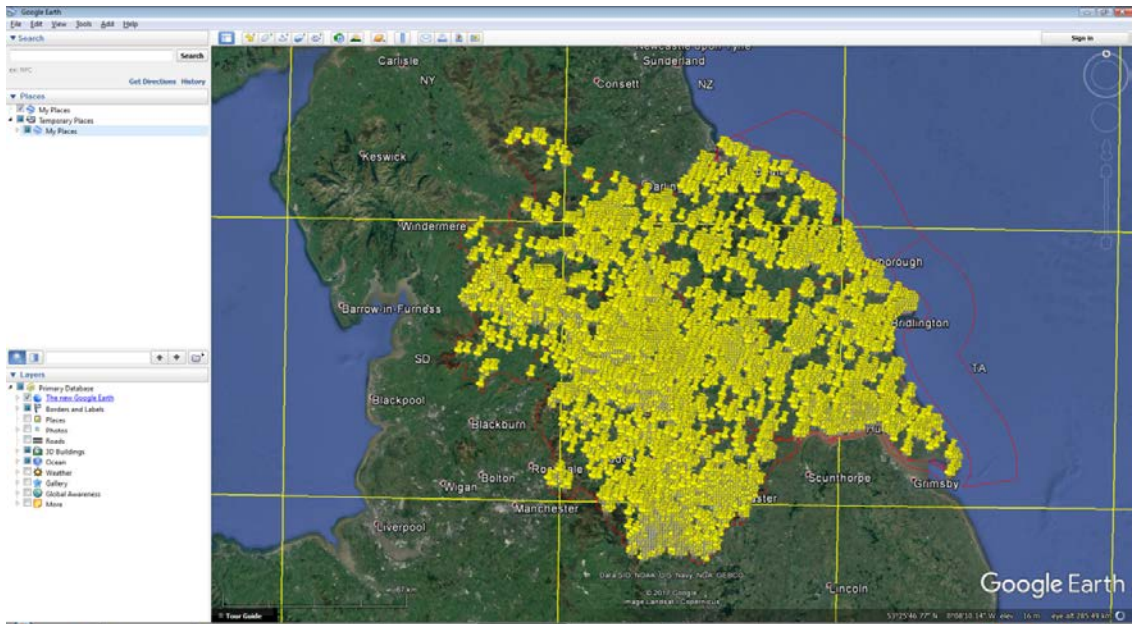


Figure 7. Screen shot of Google Earth from about 250 km above Yorkshire, with the 2015-2016 butterfly sightings at monad level. The lower version shows little advantage in a printed page, but on screen the ability to zoom in to varying scales offers much greater versatility.