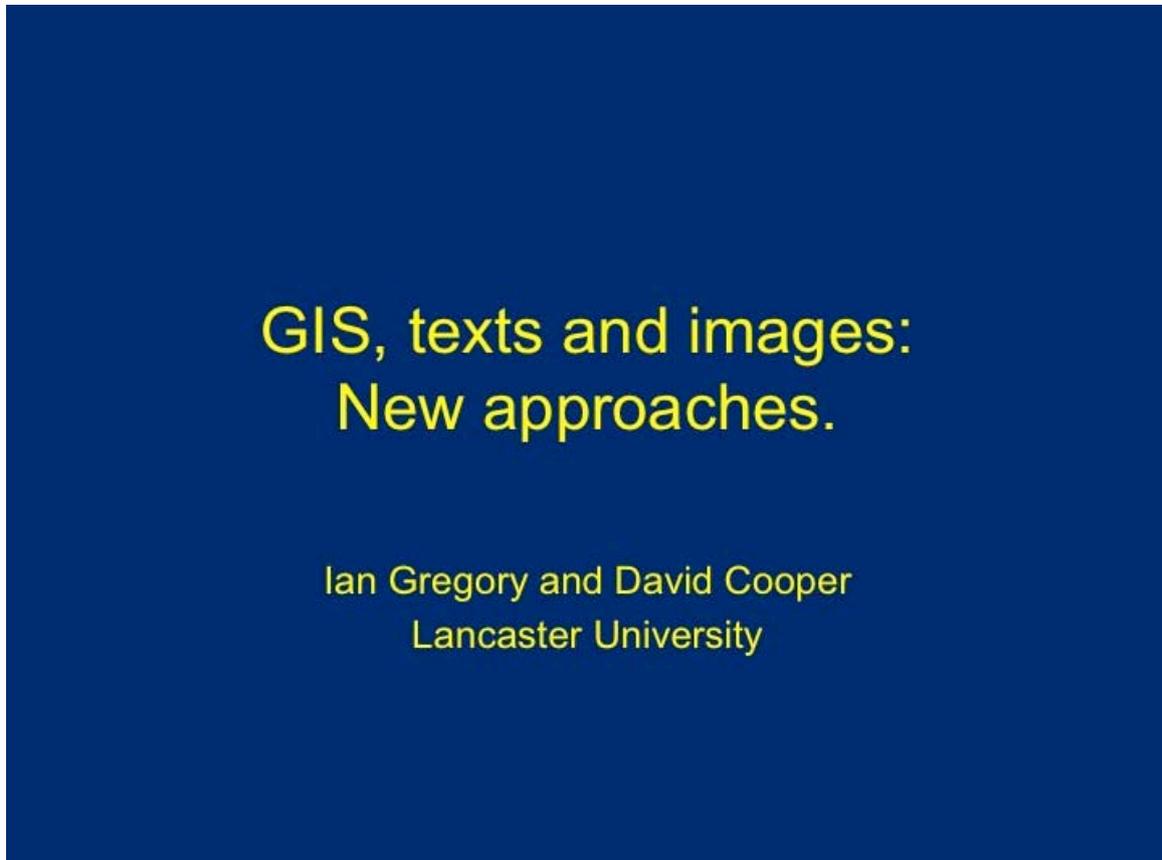


**GIS, Texts, and Images: New approaches.**

© Ian Gregory\* and David Cooper\*\* 2010

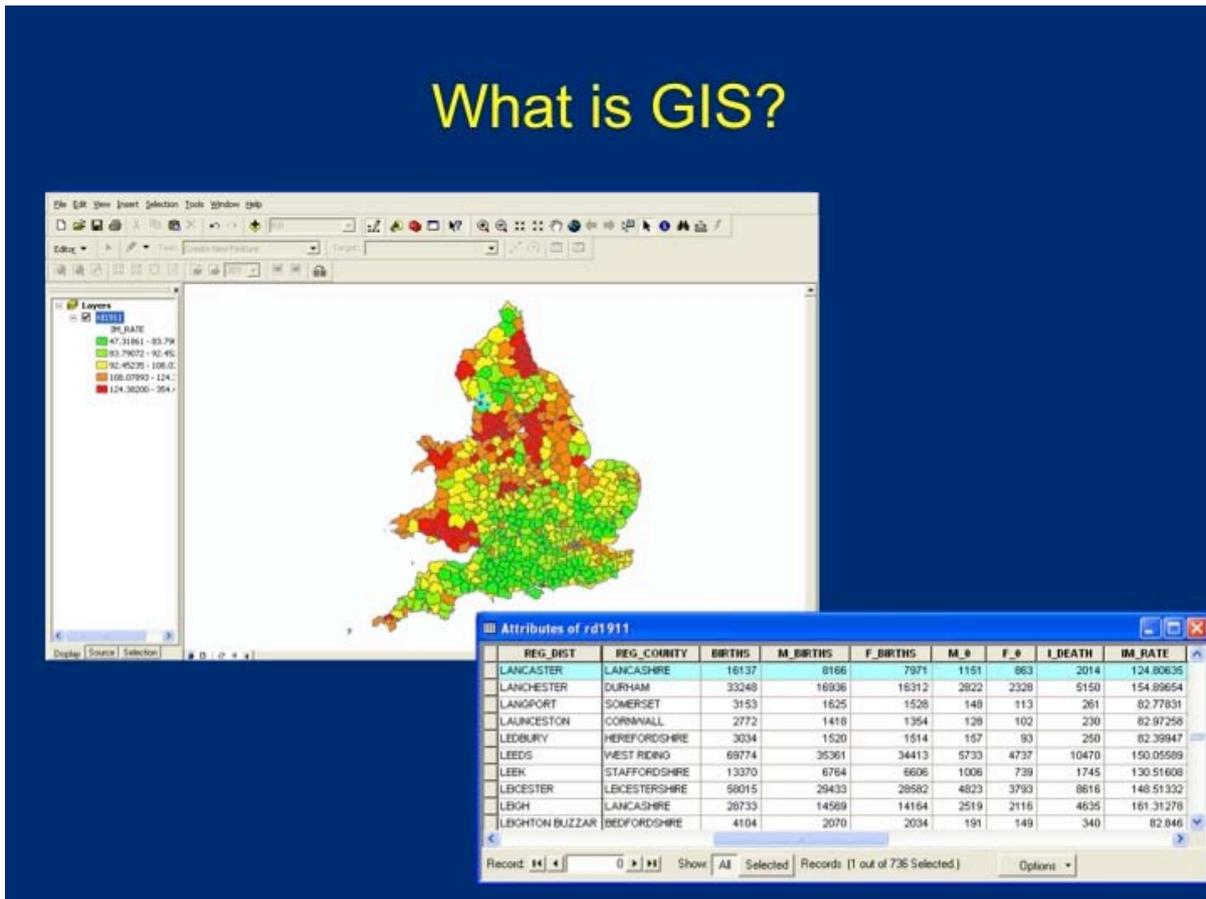
**Slide 1: Title page**



Geographical Information Systems (GIS) have become widely accepted in historical research, and there are increasing calls for them to be used in disciplines further across the humanities. The difficulty is, however, that GIS comes from a quantitative, social science paradigm that is frequently not well suited to the kinds of sources that are widely used in the humanities. The challenge for GIS, if it is to become a widely used tool within the humanities, is thus two-fold. First, approaches need to be developed that allow humanities sources to be used within a data model that is usable by GIS. Second, and more importantly, researchers need to demonstrate that by using GIS they can make significant new and substantive contributions to knowledge within humanities disciplines such as Literary Studies. This paper explores both of these questions focussing primarily on cultural representations of the English Lake District but also looking at seventeenth century newsbooks published in London.

**Slide 2: What is GIS?**

## What is GIS?



Slide 2 shows a ‘traditional’ GIS database as used in historical research. It takes mortality data from the 1900s. These data are in tabular form: each row refers to a particular place for which there are columns of mainly numeric data providing information about that place. What makes a GIS different from a conventional database is that in the GIS each row of data is linked to a map-based representation of where that place is, in this case in the form of ‘polygons’ – areas that represent the districts of the time. This allows the researcher using GIS to ask questions not only about what is happening but also where it is happening. In many ways the map shown on this slide can be thought of as a response to a database query where each polygon is shaded according to the numeric data associated with it.

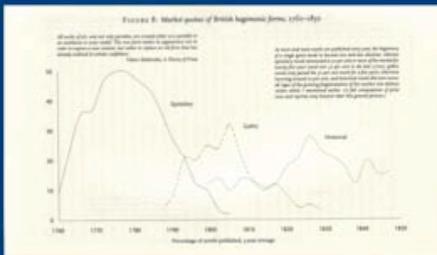
The problem with this, of course, is that quantitative history is only a relatively small part of the discipline of history and many other humanities subjects have no tradition whatever of using quantitative sources. If GIS is going to be used in these subjects, we first need to be able to convert sources such as texts and images into a format that can be used within a GIS.

### Slide 3: From History to Literature

## From History to Literature

“...this is an essay on literary history: literature, the old territory (more or less), ... within that old territory, a new object of study: instead of concrete, individual works, a trio of artificial constructs – graphs, maps and trees – in which the reality of the text undergoes a process of deliberate reduction and abstraction. ‘Distant reading’, I have once called this type of approach;” (Moretti, 2005: p. 1)

### Graphs (p. 16)



### Maps (pp. 55)



### Trees (p. 73)



Before we explore this, it is worth pausing to consider why geography and mapping may be important to a discipline such as Literary Studies. In his 2005 book *Graphs, Maps, Trees*, F. Moretti explores the idea of what he calls “distant reading.” This involves the researcher summarising and abstracting a large volume of information about a literary source or sources, rather the more traditional approach to the discipline which involves reading a single book or small number of books in detail. He illustrates the potential of distant reading based on the three devices from the title: time-series graphs are used to explore the rise and decline of sales of books from different genres (p. 16); maps are used to diagram the relations between protagonists and their objects of desire in novels about young men in Paris, (p. 55); and trees are used to explore the evolution of the use of clues in detective novels where some clues branch into new developments while others fail to evolve (p. 73). The use of graphs and trees appears convincing; however, the use of maps is far less so at least from a distant-reading perspective. There are two reasons for this: first, while there may be much that can be learned from maps such as the ones that he uses, the effort required to create such maps, which must involve very careful reading of the texts, goes against the idea of distant reading. Second, the use of manual cartography means that once the map is drawn it cannot easily be interrogated or refined.

To examine the extent to which GIS can help resolve these problems, a pilot project was developed, *Mapping the Lakes*, based on the literature of the English Lake District. The Lake District is England’s most popular national park and is an intricate landscape of lakes and mountains including Sca Fell Pike which, at 978m, is the highest mountain in England. The

area has a rich literary heritage personified in particular by William Wordsworth and the Lake Poets. This combination of geography and literature makes it an ideal area for this study.

**Slide 4: Literary Mapping of the Lakes: A pilot for humanities GIS**

**Literary Mapping of the Lakes:  
A Pilot for a Humanities GIS**

- British Academy funded pilot project
- Two tours of the Lake District
  - Thomas Gray, 1769 (9,000 words)
    - Proto-Picturesque
  - ST Coleridge, 1802 (10,000 words)
    - Romantic
- Aims:
  - Can we create a GIS of text?
  - What can it offer to literary research?
- Method:
  - Texts typed up by hand
  - Places tagged manually
  - Conversion
  - Analysis

For the initial work on the Mapping the Lakes project, two early descriptions of tours of the area were selected: Thomas Gray's proto-Picturesque tour of 1769 and Samuel Taylor Coleridge's 1802 'circumcursion.' There were two reasons for selecting these tours. First, as Malcolm Andrews (1994) notes in *The Picturesque: Literary Sources and Documents*, Gray's tour became well known as a precursor of the classic Picturesque tour. Coleridge, by contrast is closely associated with the Romantic movement. This distinction is important. Generally speaking, the Picturesque movement is closely connected with the early development of landscape tourism in which the observer travels around the landscape with a tendency to observe it from defined spots of beauty. Like William Gilpin whose writings led the movement, the observer records and perhaps even perceives observations in a stylised manner. The Romantic movement, of which Wordsworth is the leading figure, both developed this perceptual mode and reacted against it. While continuing to stress the aesthetic quality of the landscape, the Romantic writer became part of the landscape rather than remaining a detached observer. From an intellectual viewpoint, therefore, the differences we can find between these two accounts is clearly important. From a more practical point of view, both of these accounts are relatively short, at around 10,000 words each, making them relatively easy to handle within the limitations of the project.

There were therefore two questions for the project: can we create a GIS of text; and what can it offer to literary research? To answer these, four stages had to be gone through: the texts were typed up by hand, place names were tagged manually, they were then converted to a GIS, and subsequently analysed.

#### Slide 5: Place names coded in XML

**Place names coded in XML**

```
<p in_text="Y">On Sunday Augt. 1st - half after 12 I had a Shirt, cravat, 2 pair of
Stockings, a little paper & half a dozen Pens, a German Book (Voss's Poems)
& a little Tea & Sugar, with my Night Cap, packed up in my natty green oil-
skin, neatly squared, and put into my <format format_type="I">net</format>
Knapsack / and the Knap-sack on my back & the Besom stick in my hand, which
for want of a better, and in spite of <person>Mrs C.</person> & <person>Mary</
person>, who both raised their voices against it, especially as I left the Besom
scattered on the Kitchen Floor, off I sallied - over the Bridge<my_comment><pl_name
visited="Y">Greta Bridge, Keswick</pl_name></my_comment>, thro' the Hop-Field,
thro' the <pl_name visited="Y">Prospect Bridge</pl_name> at <pl_name
visited="Y">Portinscale</pl_name>, so on by the tall Birch that grows out of the
center of the huge Oak, along into <pl_name visited="Y">Newlands</pl_name>--
<pl_name visited="Y">Newlands</pl_name>is indeed a lovely Place-the houses...
```

Slide 5 shows a snippet of Coleridge's account as it was marked up using XML tags. As can be seen, the tags highlight a variety of things including editorial comments, people and most importantly, place names, highlighted with the *pl\_name* tag. We also used an element to flag whether the author was actually at the place mentioned (*visited* = "Y"), or whether he was talking about it from afar (*visited* = "N"). Thus, if a writer was to say "From Place A I could see Place B," Place A would be tagged as visited while Place B would be tagged as not visited.

#### Slide 6: Convert to a GIS

## Convert to a GIS

	A	B	C	D	E	F	G	H	I	J	K	L
111	Great Gevel	N	Thursday Aug 5th	5/8/1802	Great Gable	510500	321500					
112	Kirk Fell	N	Thursday Aug 5th	5/8/1802	Kirk Fell	510500	319500					
113	Green Crag	N	Thursday Aug 5th	5/8/1802	Green Crag	498500	320500					
114	Piler	N	Thursday Aug 5th	5/8/1802	Piler	512500	317500					
115	Sheepie	N	Thursday Aug 5th	5/8/1802	Sheepie	511500	319500					
116	Hay Cock	N	Thursday Aug 5th	5/8/1802	Haycock	510500	314500					
117	Great End	N	Thursday Aug 5th	5/8/1802	Great End	508500	322500					
118	Esk House	N	Thursday Aug 5th	5/8/1802	Esk House	509500	323500					
119	Bow-fell	N	Thursday Aug 5th	5/8/1802	Bow Fell	506500	324500					
120	Scie Fell	Y	Thursday Aug 5th	5/8/1802	Scie Fell	506500	320500					
121	Scie Fell	Y	Thursday Aug 5th	5/8/1802	Scie Fell	506500	320500					
122	Broad Crag	N	Thursday Aug 5th	5/8/1802	Broad Crag	507500	322500					
123	Dow Crag	N	Thursday Aug 5th	5/8/1802	Dow Crag	506500	322500					
124	Scie Fell	Y	Thursday Aug 5th	5/8/1802	Scie Fell	506500	320500					
125	Hollow Stones	Y	Thursday Aug 5th	5/8/1802	Hollow Stones	507500	320500					
126	Helvellyn	N	Thursday Aug 5th	5/8/1802	Helvellyn	515500	334500					
127	Bonowdale	N	Thursday Aug 5th	5/8/1802	Bonowdale	517500	324500					
128	Castle Crag	N	Thursday Aug 5th	5/8/1802	Castle Crag	515500	324500					
129	Dewent Water	N	Thursday Aug 5th	5/8/1802	Dewent Water	521500	325500	Lake				
130	Scie Fell	Y	Thursday Aug 5th	5/8/1802	Scie Fell	506500	320500					
131	Esisdale	N	Thursday Aug 5th	5/8/1802	Esisdale	500500	317500	Valley				
132	Broad-crag	Y	Thursday Aug 5th	5/8/1802	Broad Crag	507500	322500					
133	Scie Fell Man	N	Thursday Aug 5th	5/8/1802	Scie Fell	506500	320500					
134	Dow-crag	N	Thursday Aug 5th	5/8/1802	Dow Crag	506500	322500					
135	Broad-crag	N	Thursday Aug 5th	5/8/1802	Broad Crag	507500	322500					
136	Haw	Y	Thursday Aug 5th	5/8/1802								
137	Dow-Crag	N	Thursday Aug 5th	5/8/1802	Dow Crag	506500	322500					
138	Dow-crag	N	Thursday Aug 5th	5/8/1802	Dow Crag	506500	322500					
139	Esk House	N	Thursday Aug 5th	5/8/1802	Esk House	509500	323500					
140	Kirk	N	Thursday Aug 5th	5/8/1802	Bow Fell	502500	319100	Blower				

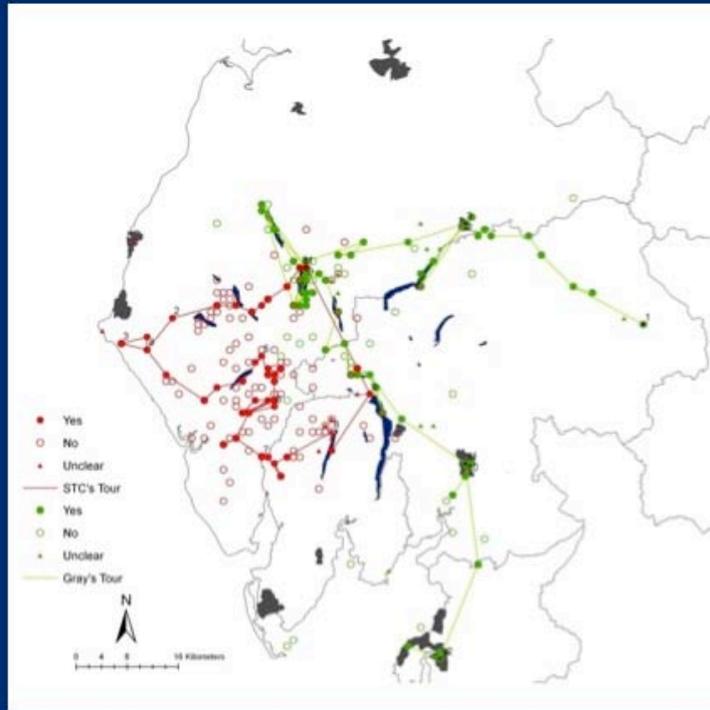
OS 1:50,000 gazetteer – all places on 1:50,000 maps

- Accuracy
- Spelling problems
- Disambiguation

Having done this, extracting the place names from the text and reading them into a database table is a simple process. To convert this into a GIS the essential next stage is to give a coordinate to every place-name. This can be done by using a relational join to link the raw place-names to a place-name gazetteer, effectively a database table that gives a coordinate for every name. In this project the Ordnance Survey's 1:50,000 gazetteer was used. This provides the British National Grid reference for the 1km grid square in which the place names that appear on 1:50,000 maps are placed. Joining this to the place-names from the text produces a table such as the one shown on the slide. The first four columns are taken from the texts and show the raw place-name, whether the author was visiting the site or not, and the date in two formats. The last three columns come from the gazetteer and give a standardised version of the place-name and, most importantly, the northing and easting of its location. Clearly, a certain amount of work is required to ensure accuracy in this. Spelling problems, such as the differences between "Bow-fell" and "Bow Fell" need to be resolved, and names need to be disambiguated where the same name can refer to more than one location. Given the relatively small size of the texts and the study area neither of these presents a major challenge. There are also issues associated with the accuracy of the grid references, which are at best only to the nearest kilometre but for linear features, such as rivers, or vague features, such as valleys, may be somewhat misleading.

### Slide 7: Coleridge and Gray in a GIS

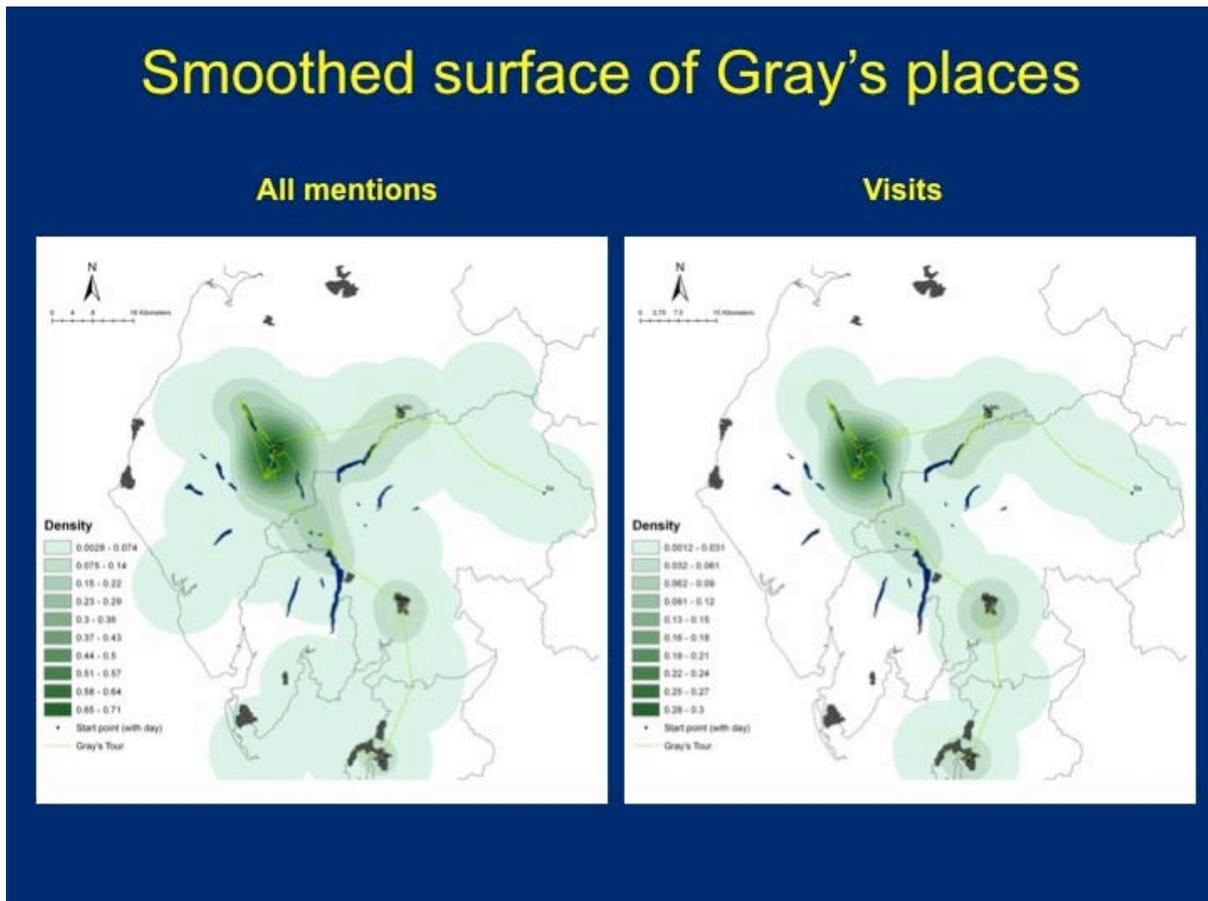
## Coleridge & Gray in a GIS



Once we have coordinates for our place-names, converting these to a GIS is a simple piece of GIS functionality in which the coordinates are converted into point data. The slide shows both tours on a single map with straight lines being used to join together the points mentioned as visited to help illustrate the route taken. Gray started at Brough to the east of the Lake District, travelled to Penrith where he spent two nights, travelling down to Ullswater for the day in between. He then journeyed on to Keswick where he spent six nights travelling out on day trips to the surrounding countryside. He then travelled south, over Dunmail Raise, the main route through the central Lake District, to spend two nights in Kendal, and finally on to Lancaster where the Lake District part of his tour finishes.

Coleridge starts in Keswick, where he lived, and travelled south-west through the Newlands Valley to spend three nights in and around St Bees on the coastal plain, west of what is now the National Park. He then travels back into the Lake District up the valley of Wasdale from where he climbs Sca Fell, the description of whose ascent and descent his account is famous for. From here he travels on through the south-western Lake District and over to Coniston before travelling north up Dunmail Raise to return home.

### Slide 8: Smoothed surfaces of Gray's places

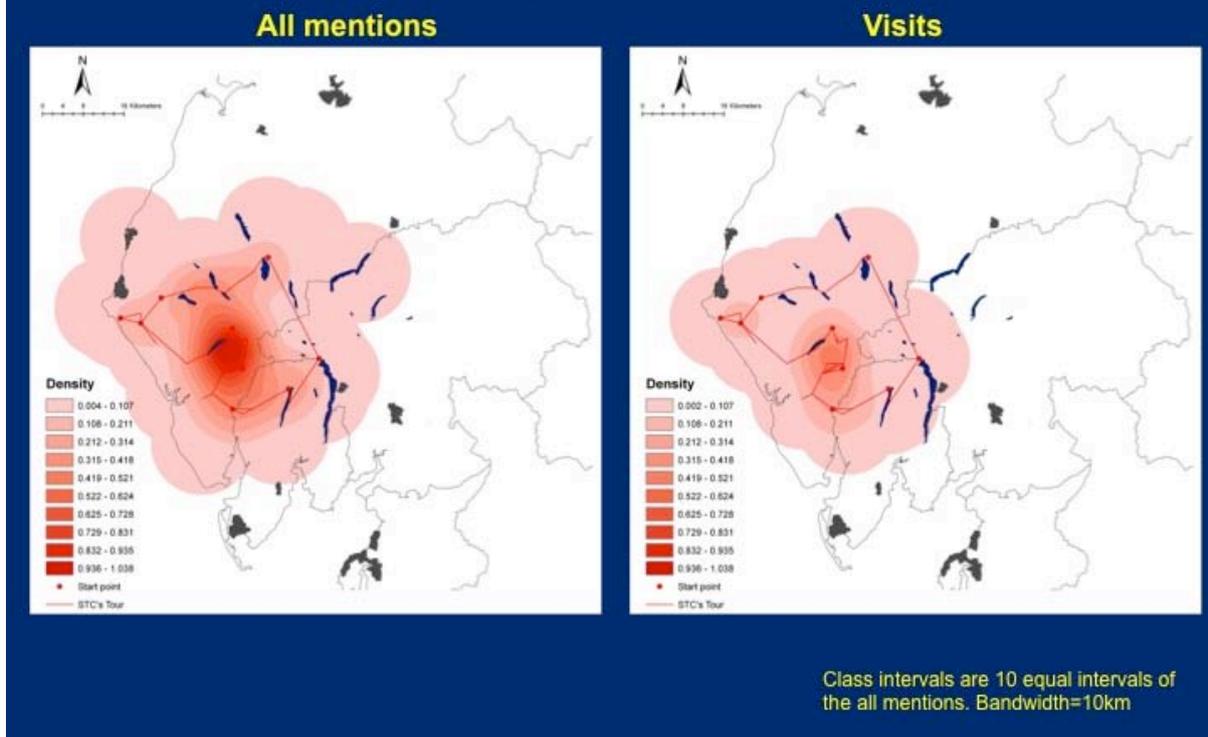


It is well known from a cartographic perspective maps that such as those on the previous slide are difficult to interpret. For this reason techniques have been developed that attempt to simplify them and make them more readily comprehensible. One example of this, pioneered in disciplines such as epidemiology and crime mapping, is density smoothing in which the density of events around each location is mapped with denser locations being shaded in darker colours. The density is calculated using a *distance decay model* such that near events have more impact than those that are further away. In this case an 'event' is a place being named in a tour. As well as simplifying the pattern, this has the second advantage of reducing the accuracy implied by the point map.

The slide uses density smoothing on Gray's tour showing one map of all places mentioned and one of those actually visited. The maps stress the central importance of the area around Keswick to Gray's account although other clusters such as Penrith and Ullswater, Kendal, and Lancaster are all apparent. What becomes clear is the importance of urban centres and valleys to Gray's text.

### Slide 9: Smoothed surfaces of Coleridge's places

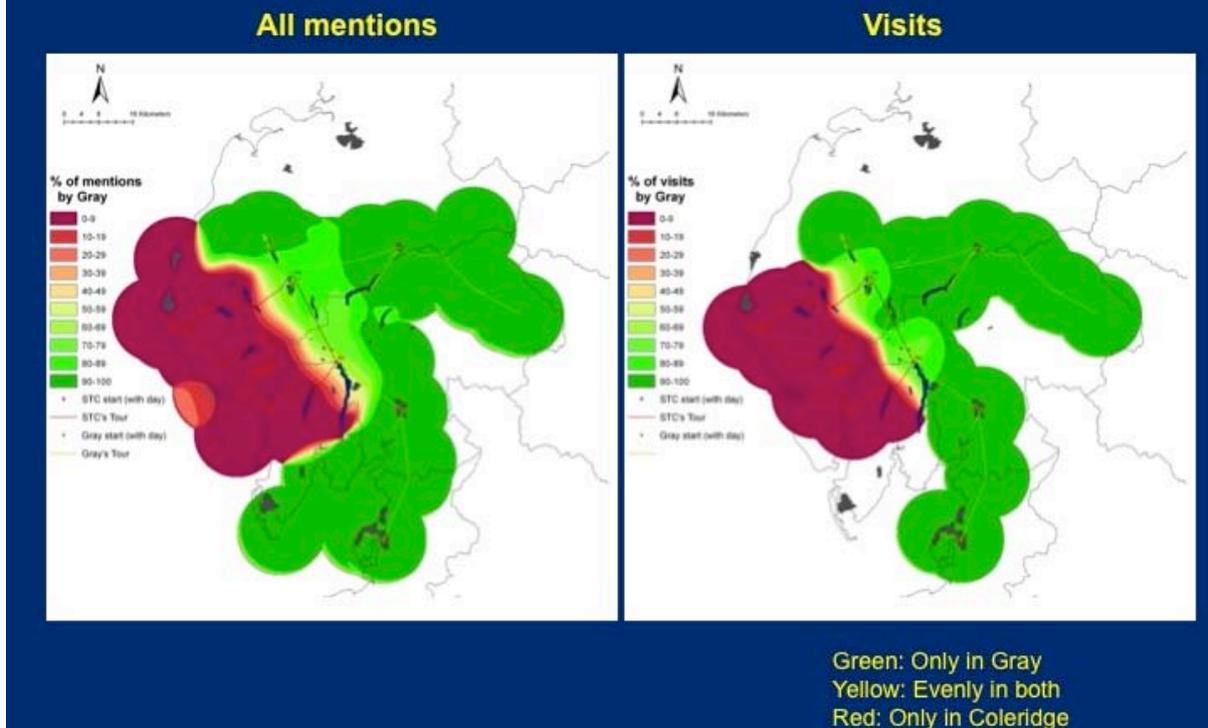
## Smoothed surface of Coleridge's places



This slide shows the same two maps for Coleridge's place-naming. His tour shows a very different pattern with the account being clustered on the area around Sca Fell.

### Slide 10: Comparing Gray and Coleridge

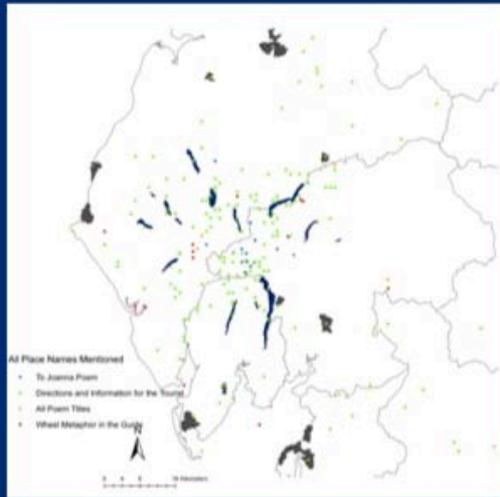
## Comparing Coleridge and Gray



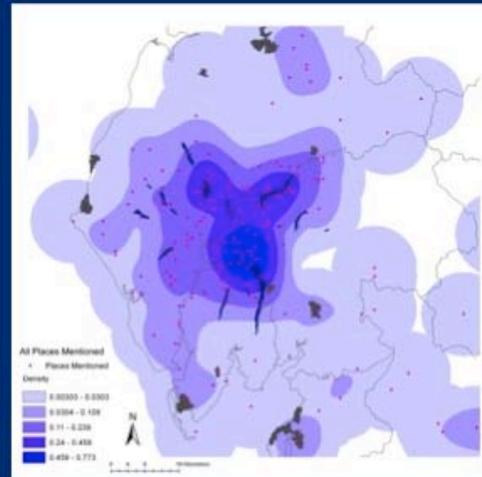
This slide moves further into analytic mapping. The density smoothed surfaces for the two authors have been brought together to allow them to be compared. Where a location is only mentioned by Gray, it is shaded green; where it is only mentioned by Coleridge, it is shaded red. Areas that are mentioned by both are shaded in yellows and pale greens. The salient element of these maps is the lack of overlap. Coleridge's account concentrates on western and south-west parts of the Lake District; Gray's follows a linear trajectory west from Brough to Keswick and then south-east to Lancaster. The two barely overlap except for around Keswick, where Coleridge only started and finished, and over Dunmail Raise, which neither author spends much of his account describing. Thus while both authors are known for their writing on the Lake District, the places they talk about are largely mutually exclusive.

### Slide 11: Wordsworth and the Lake District

## Wordsworth and the Lake District



Places Mentioned

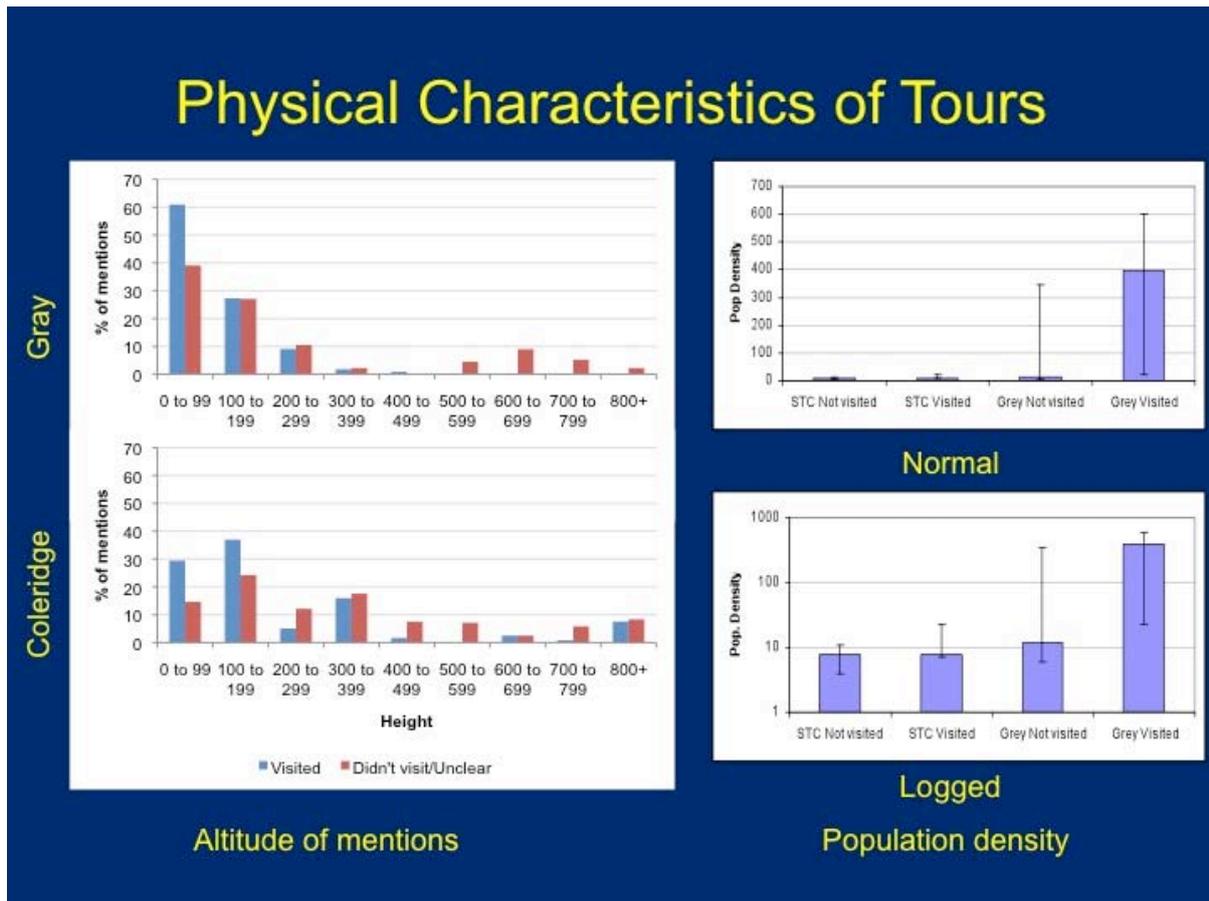


Density Smoothed

Thanks to Kirsten Hansen,  
Mt. Holyoke College, USA

As a brief aside, work has also been done looking at Wordsworth's writing. These maps bring together places named in the titles of his poems, the poem *To Joanna*, one of the few that contains many place names, and sections of his *Directions and Information for the Tourist*. Here again, there is a significantly different pattern. Wordsworth concentrates on the area around Grasmere, where he lived, and from here the pattern spreads north-east around Ullswater and north-west towards Keswick and Borrowdale.

**Slide 12: Physical characteristics of the tours**



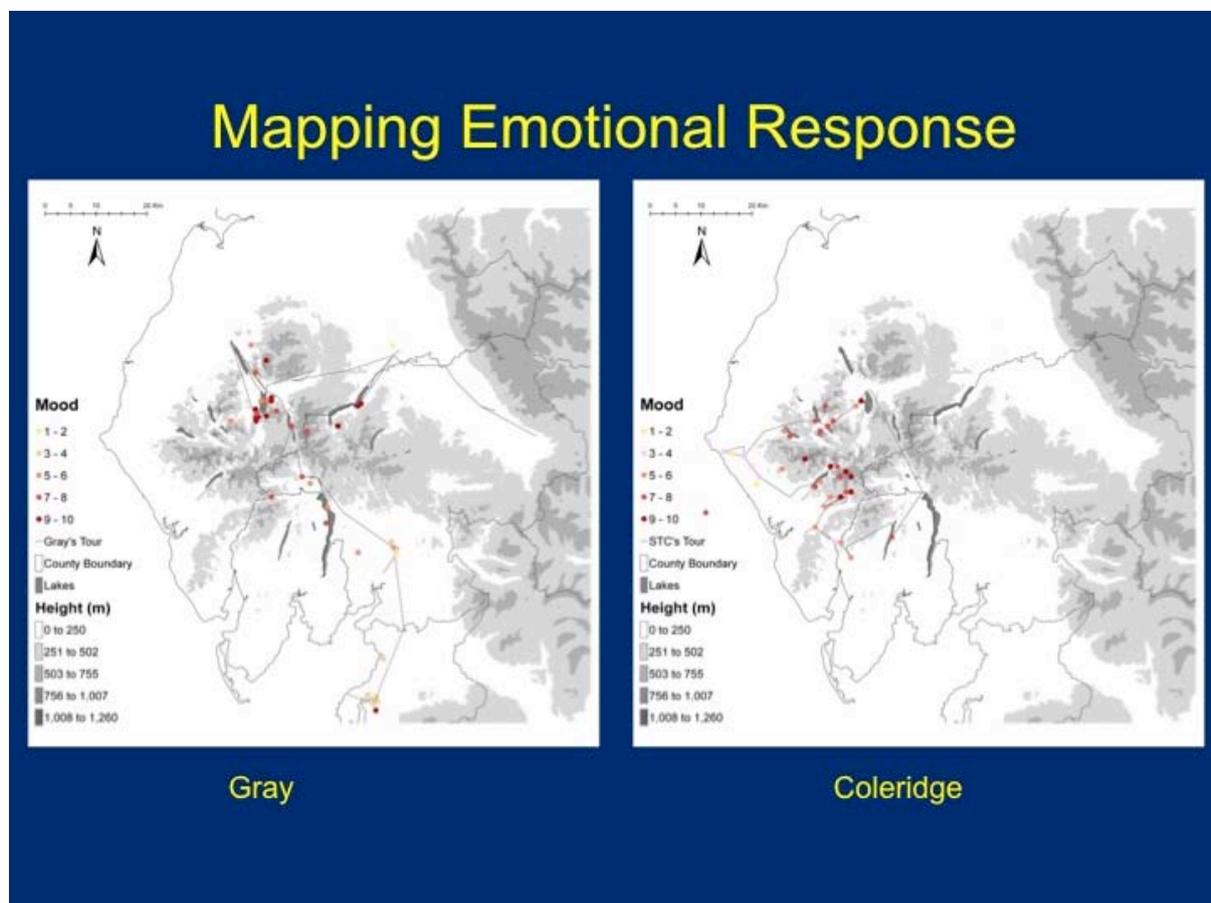
One of the big advantages of GIS is its ability to integrate data from apparently disparate sources. The previous maps have implied that Gray concentrated on the more urban areas and the valleys, while Coleridge consciously sought out the more remote and upland parts of the Lake District. Using location to integrate data from other sources can help us explore this further. A useful GIS-based source of information about height is a *Digital Elevation Model* (DEM), a representation of the Earth's surface that gives heights for every location. Integrating a DEM with the point data on place-names allows us to allocate a height to every mention. Rather than mapping them, these can then be graphed. The graph in the top-left of the slide shows heights of places mentioned by Gray distinguishing those places that he visits from those that he talks about from a distance. A clear pattern is apparent. He spends all of his time at low altitudes, with over 60% of visited places being under 100m and all being under 1000ft. Most of the places he mentions without visiting are similarly low although there are also mentions of high places, particularly those over 600m, the higher Lake District peaks. He almost completely ignores places in mid-altitudes. This pattern seems to fit well with the concept of Gray as a Picturesque tourist: he spends his time in the valleys and passes, describing the areas around him and looking up to the high peaks.

Coleridge's pattern is striking for both its differences and its similarities. Like Gray he spends much of his time at lower altitudes but not to quite the same extent. Coleridge also visits places across the height range including a cluster of mentions in the very highest intervals, over 800m when he climbs Sca Fell. While his account is famous for this ascent, it only

occupies a relatively small proportion of the heights of the places that he visits. It is also noticeable that Coleridge does not ignore mid-height places.

A similar exercise can be carried out using population density. This was available for parishes from the 1851 census using parishes, a relatively small type of administrative unit represented in a GIS as polygons. Again by integrating these with the point data of place-names from the tours we can allocate a population density to each mention. The 1851 census is noticeably later than the two tours, of course, but as the urban structure is unlikely to have changed much the relative densities are probably good enough to provide a general pattern. The pattern basically confirms what would be expected. Gray spends much of his time in relatively dense places while the average density of places that Coleridge (STC) visits is much lower. Perhaps more interesting is the fact that the average densities of places that they mention without visiting are fairly similar although Gray's are more varied in range.

### Slide 13: Mapping emotional response



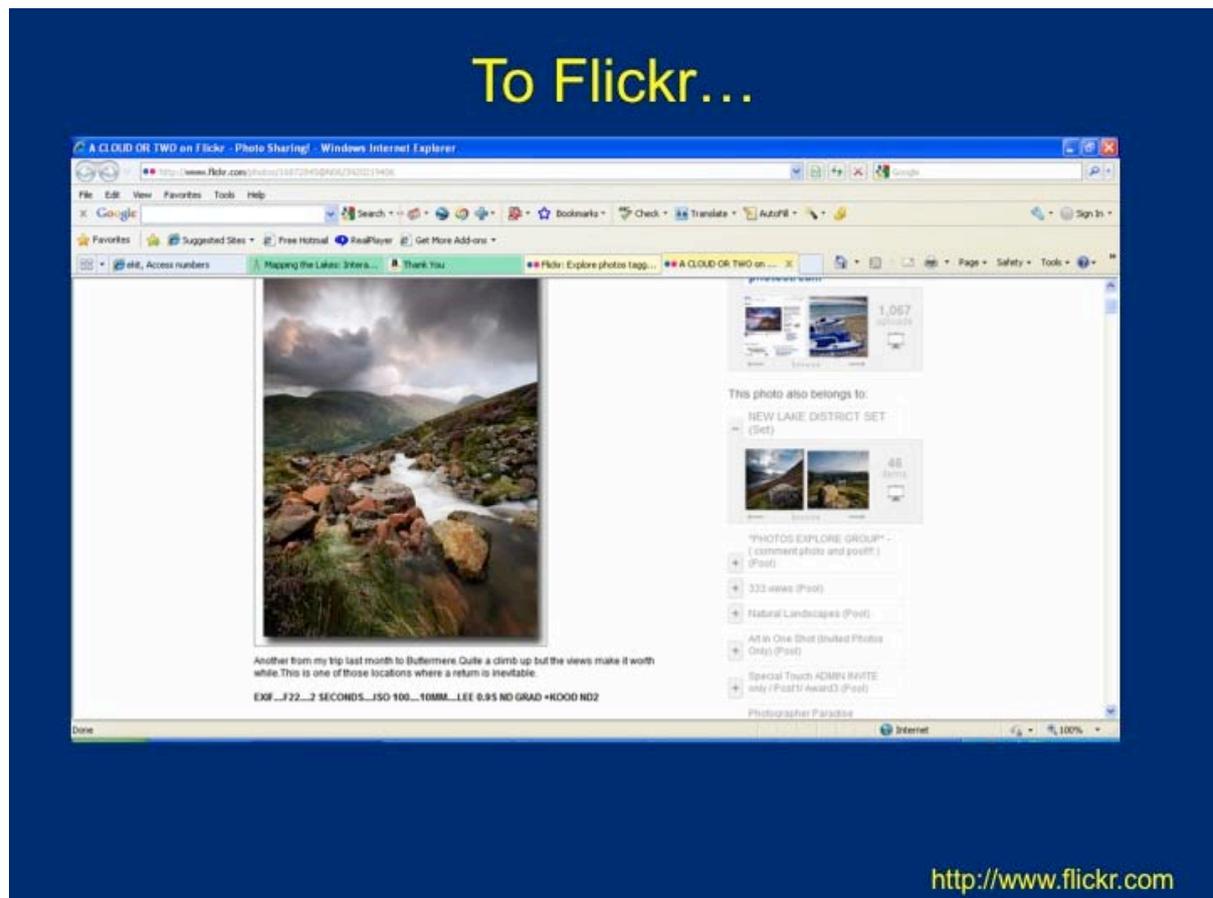
As well as mapping where the writers were talking about, we were also interested in mapping their response to the landscape. To do this a ten point scale was devised that associated the emotional response that the writers had to the places that they were talking about. At the bottom of the scale were words such as “dull” and “tedious” while at the opposite end, attracting a score of 10, were words such as “sublime” and “terrifying.” As shown in the slide, mapping these for the two authors gives a somewhat different pattern than the simple maps of where they were talking about. For Gray, rather than Keswick, the emotional centre is



allows the user to turn this old-map-view on or off. This window also maps the locations of the places named in the text, distinguishing between those that are visited, which are emphasised, and those that were not. In the current display only places mentioned on the first day of Coleridge's tour are mapped. Other days, or places from Gray can be added using the table of contents. Clicking on one of these points brings up a speech bubble with some basic information about the place including the author and the date of the mention. The bubble also includes a hyperlink that will take the reader to the appropriate location in the text for where this mention occurs. As will be described in more detail below, a second hyperlink takes the reader to the Flickr website where they can explore user generated photos of the surrounding landscape.

The idea behind this was that a user can read the text in the normal linear manner in the lower window while following the places named using the map above. If required he or she can follow hyperlinks to see what the landscape around this place looks like today as shown on Flickr. Alternatively, the user can look at the map, select a place of interest and read what the authors said about this place and places nearby in other parts of the text. In this way the GIS offers a way of exploring the texts in detail thus helping with more traditional scholarship by emphasising the geographies inherent in the texts.

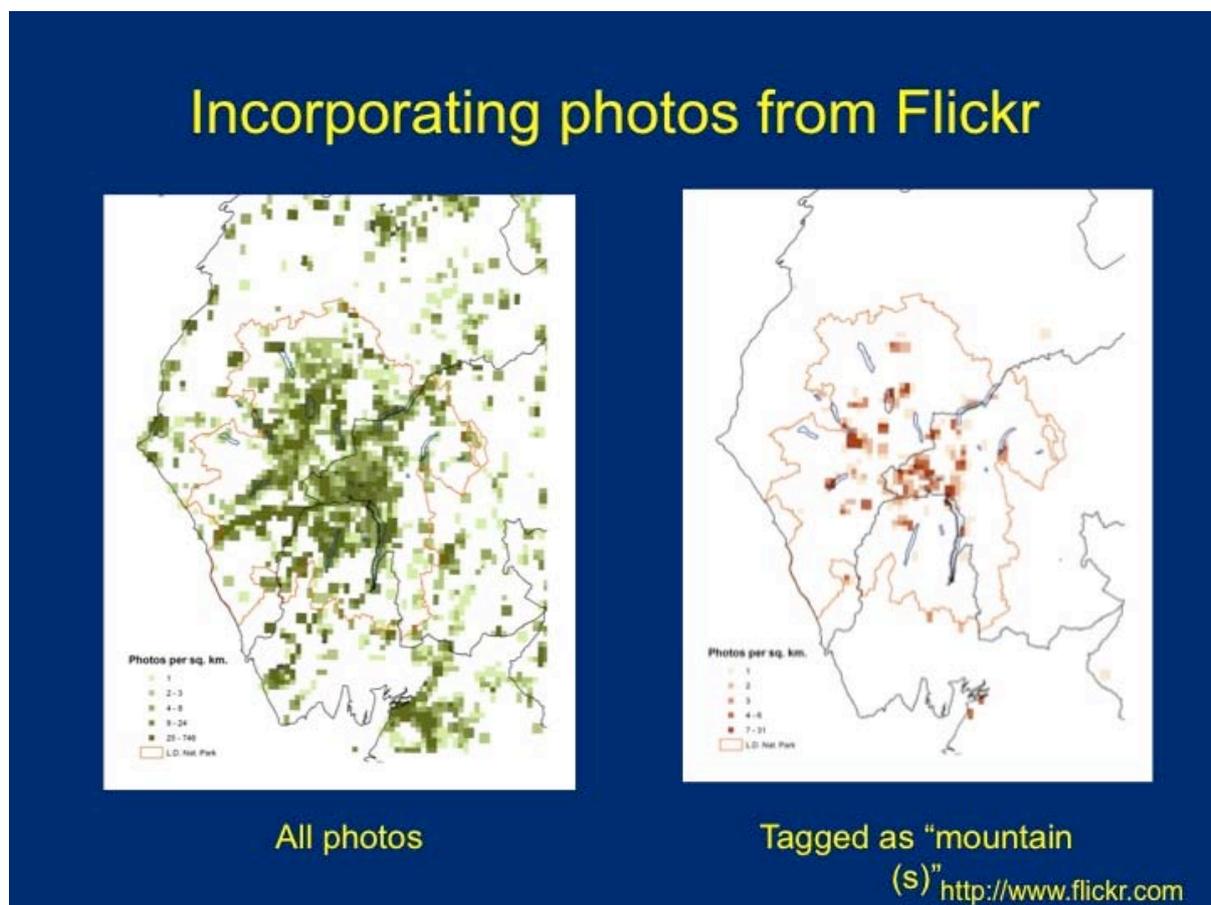
### Slide 15: To Flickr...



Flickr is a photographic website that allows users to upload their own photographs. Users add tags, or metadata, to the photographs to allow other users to find them. They can also add

geo-tags, latitude and longitudes that provide a location for the photograph. This is how we integrate these images with the texts. A latitude and longitude is passed from the Mapping the Lakes system to Flickr along with a zoom level, which specifies how near to our mention the photographed place has to be in order for the photo to be included. The tag “landscape” is also passed to Flickr in order to narrow down the number of photos included. Flickr presents the reader with a map that includes the locations of some of these images and thumbnails of them. The reader can click on the locations or the thumbnail to see a complete image, or move onto a new set. The initial idea behind this was to assist the reader to visualise what the writers were writing about by including photos of the landscape, thus enhancing the texts and the reader’s ability to close read them.

### Slide 16: Incorporating photos from Flickr

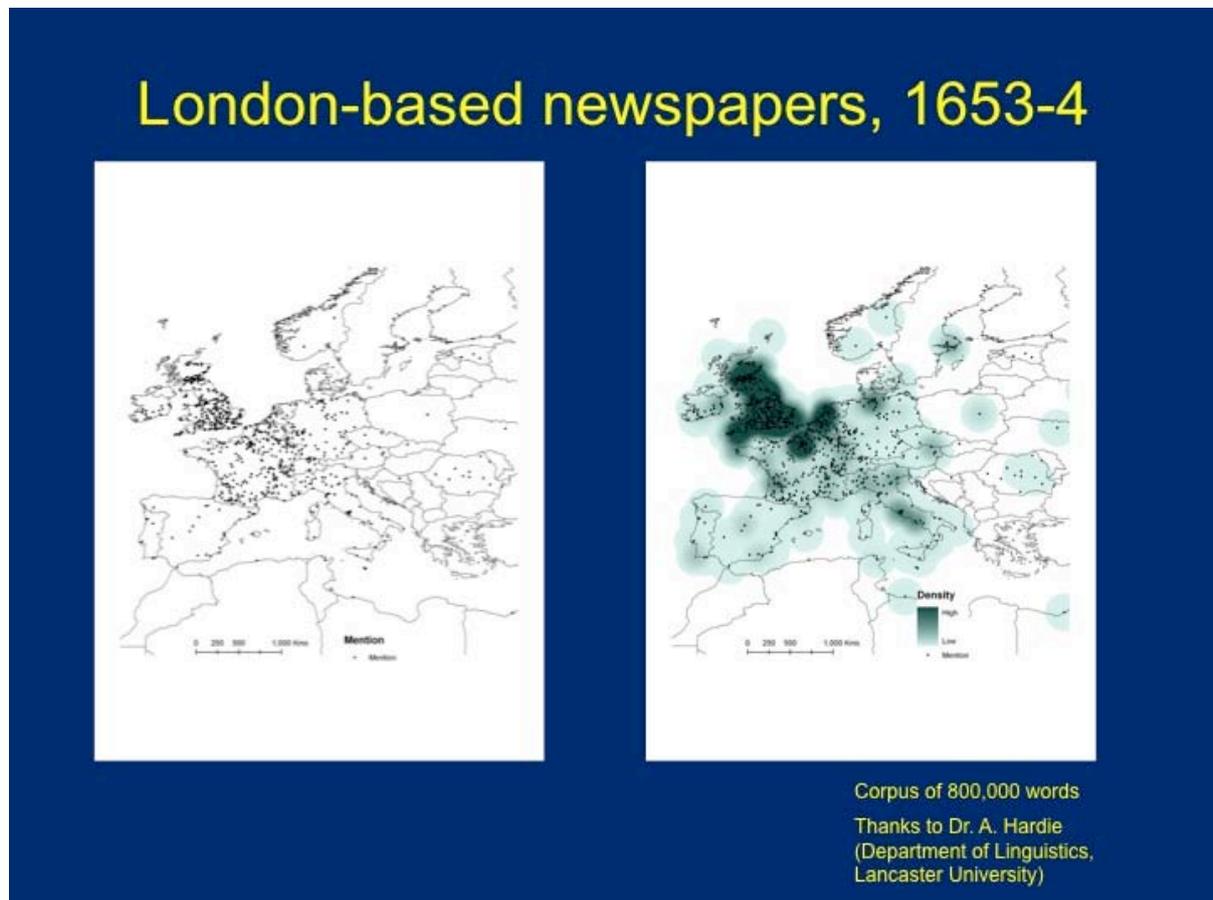


Here we move to a new idea: if photos can be used for close reading, can they also be used as part of an exercise in distant reading? Using the Flickr API, a program was written that extracted all of the photographs taken in and around the Lake District. To make the process smoother and reduce data volumes, the numbers in each 1km square were selected. This is what is shown on the slide which maps the number of photographs in each square. A number of things become apparent. First, although the Lake District is one of the most visited parts of Britain, a remarkably large proportion of it has no geo-tagged photos on Flickr. This applies both inside and outside the modern National Park boundary. Second there are clear clusters of places that are photographed repeatedly including: the main valleys such as Langdale,

Ullswater and Borrowdale; the main settlements such as Keswick, Windermere and Kendal; and some of the smaller roads such as Hardknott and Wrynose passes. This would seem to suggest that, despite the Romantic movement, the way in which the Lake District is represented in photographs today seems to follow the characteristics of Gray's proto-Picturesque tour rather than that of Coleridge, although more work is required on this.

The second map on this slide just maps images tagged as "mountain" or "mountains." Here again there is an interesting pattern as most of these are taken from valleys or from the seafront near Morecambe. While this may seem counter-intuitive, it does in fact make sense as people photographing a mountain tend to do so from the valleys or across Morecambe Bay. This again suggests that these photographs have more in common with the characteristics of Gray's way of viewing the landscape than they do of Coleridge's. There is clearly much more potential for this approach to help us understand how user-generated representations of landscapes relates to earlier representations.

#### Slide 17: London-based newspaper, 1653-4

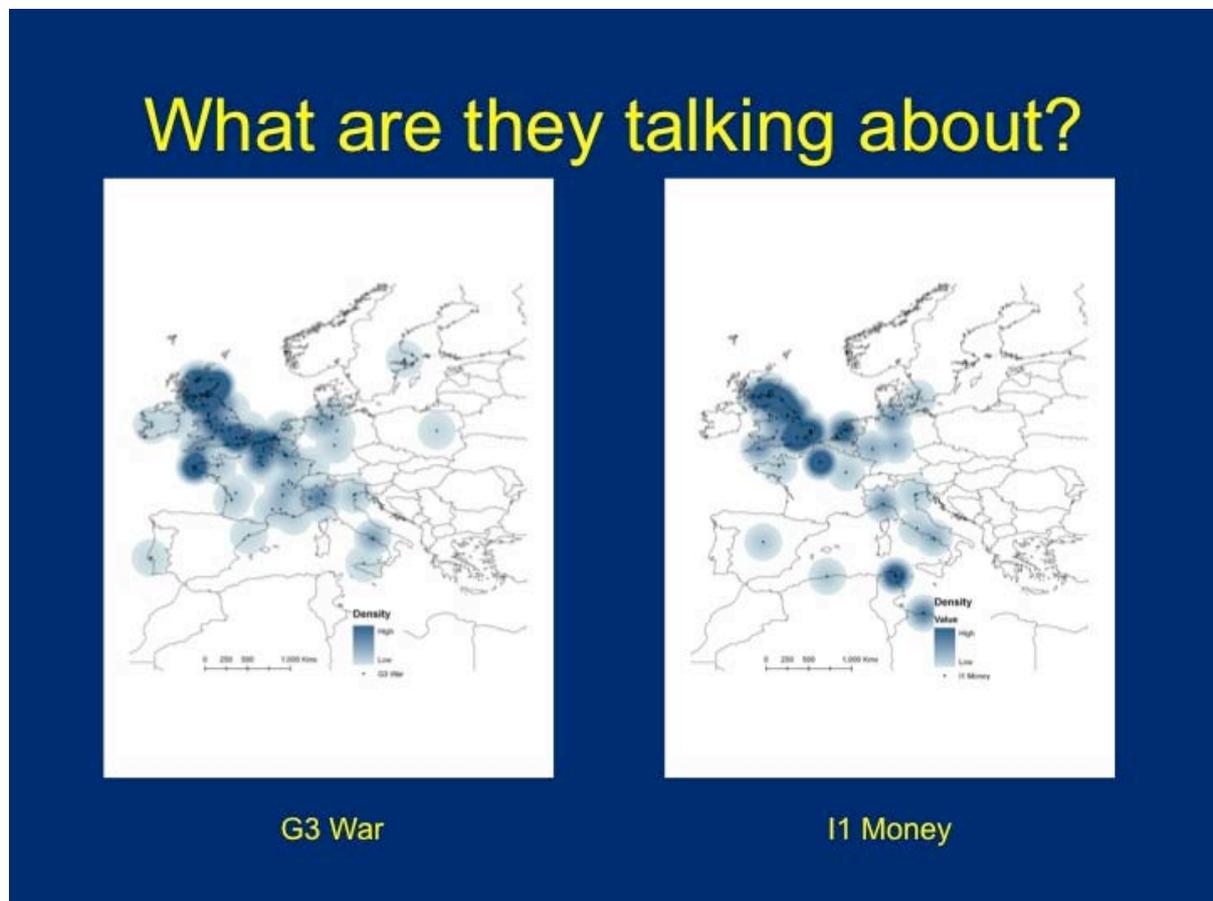


We believe that the Mapping the Lakes project was highly successful in illustrating that we can produce a GIS of texts and use it to help in Literary Studies scholarship. The major limitation with the approach taken was that place-names had to be tagged manually, thus limiting the volume of data that could be included. This poses the question of whether automated or semi-automated techniques could have been used to speed up this process. Techniques from corpus linguistics can help. Place-names are proper nouns, and identifying

proper nouns within a text can be done automatically. These can be matched to a gazetteer to add grid references; in this case World Gazetteer was used. Proper nouns that are not place names can be removed by developing filters such that, if a proper noun is preceded by, for example, “Mister” or “Duke of,” it clearly refers to a person not a place. Some manual input is however inevitable both in terms developing these filters and handling spelling inaccuracy and ambiguity.

This approach was tested on the Lancaster Newsbooks Corpus, 800,000 words of surviving text published in London between 1653-4. The maps show the success of doing this. It shows that London-based newsbooks were primarily interested in England, but there are also clusters of mentions in the Low Countries, Paris, Hamburg, Stockholm, Rome, and a number of other places. These clusters can be explained in part by the fact that when news came to London by ship the port of dispatch was often included in the resulting article.

### Slide 18: What are they talking about?



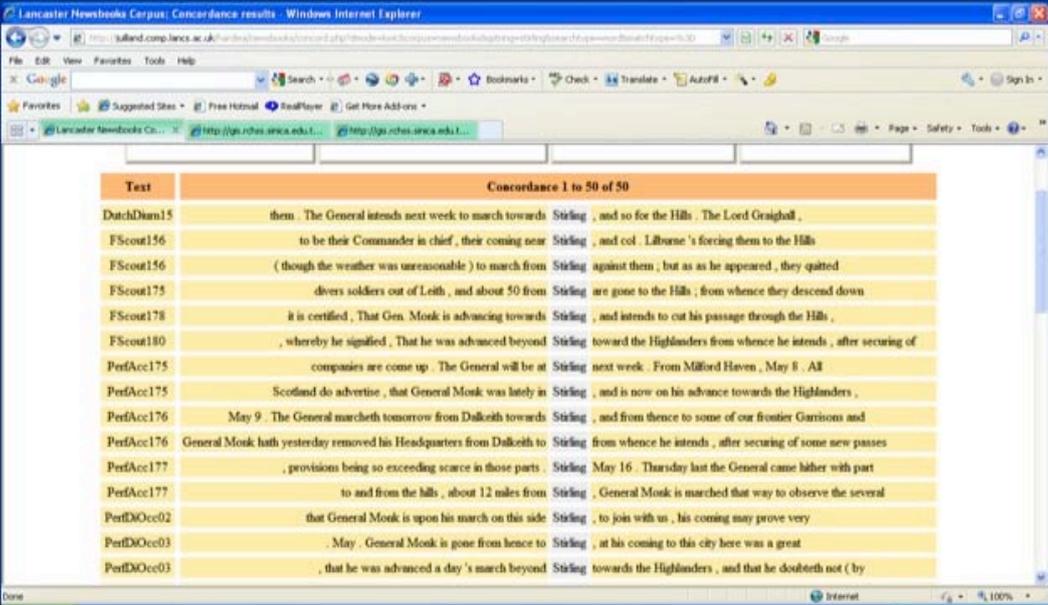
While the preceding maps are interesting in that they tell us locations in which these newsbooks were interested, the question of what they were saying about these different places is likely to be far more important. Again, corpus linguistics can help through the combination of two techniques: semantic tagging and collocation. Semantic tagging involves attaching each word to a broader class of meaning, while collocation basically asks, what other words occur within a set number of words of the word we are interested in. Bringing these two techniques together means that we can identify either “what place-names occur

near words associated with this theme?” or “what types of words occur near mentions of this place name?” In both instances “near” means near in the text rather than in geographical space.

The two maps shown on the slide give examples of doing this for semantic tag G3, words associated with war, and tag I1, those associated with money and finance. In both cases clear but very different patterns emerge. For war there is a strong cluster in eastern Scotland, a cluster that stretches from London into the Low Countries, and a large cluster centred specifically on the French port of Brest. For money, the east coast of Britain stands out as important with a cluster running from Edinburgh, through Newcastle and Scarborough, to London. Clusters are also apparent around Amsterdam, Paris and Tunis.

### Slide 19: Back to the texts...

Back to the texts...



Text	Concordance 1 to 50 of 50
DutchDiam15	them . The General intends next week to march towards Stirling , and so for the Hills . The Lord Graighall ,
FScout156	to be their Commander in chief , their coming near Stirling , and col . Lilborne 's forcing them to the Hills
FScout156	( though the weather was unreasonable ) to march from Stirling against them , but as he appeared , they quitted
FScout175	divers soldiers out of Leith , and about 50 from Stirling are gone to the Hills , from whence they descend down
FScout178	it is certified , That Gen . Monk is advancing towards Stirling , and intends to cut his passage through the Hills ,
FScout180	, whereby he signalled , That he was advanced beyond Stirling toward the Highlanders from whence he intends , after securing of
PerfAcc175	companies are come up . The General will be at Stirling next week . From Milford Haven , May 8 . All
PerfAcc175	Scotland do advertise , that General Monk was lately in Stirling , and is now on his advance towards the Highlanders ,
PerfAcc176	May 9 . The General marcheth tomorrow from Dalkeith towards Stirling , and from thence to some of our frontier Garrisons and
PerfAcc176	General Monk hath yesterday removed his Headquarters from Dalkeith to Stirling from whence he intends , after securing of some new passes
PerfAcc177	, provisions being so exceeding scarce in those parts . Stirling May 16 . Thursday last the General came hither with part
PerfAcc177	to and from the hills , about 12 miles from Stirling . General Monk is marched that way to observe the several
PerfDOcc02	that General Monk is upon his march on this side Stirling , to join with us , his coming may prove very
PerfDOcc03	. May . General Monk is gone from hence to Stirling , at his coming to this city here was a great
PerfDOcc03	, that he was advanced a day 's march beyond Stirling towards the Highlanders , and that he doubteth not ( by

So far, this is clearly an exercise in distant reading in that we have been able to summarise a large corpus of newsbook text without actually having to read any of it. This is clearly useful in summarising the text but is also superficial and does not tell us why the texts are talking about these places. We can, however, refer back to the raw texts to take this further. The slide starts with the example of “Tunis” because it seems surprising that this stands out as a place associated with finance. What we find is that this association arises from a tagging error: there are various statements that include phrases such as “call the Turks... to account for the wrongs they have done unto the Christians...” The issue here is that the word “account” has been wrongly tagged as referring to money. It is not surprising that issues such as this occur

when large volumes of data are analysed automatically; however, refining the map to deal with such errors is not a problem. The occurrence of these errors stresses the importance of human intervention in this type of work: it requires accepting that automated techniques will always have their limitations.

The second example is concerned with Scarborough, again perhaps not somewhere that might be thought of as a major financial centre. Here the pattern seems valid as there are various mentions concerned with Scarborough and “prize ships,” enemy ship captured for profit. The third example picks up on the cluster associated with war in eastern Scotland focussing on one of the places that makes the cluster, Stirling. From this it is clear that a rebellion was going on in Scotland at this time, and this rebellion was attracting much attention in the London newsbooks of the time. A fourth example, not mapped, is the cluster on Brest that is associated with war. This association is caused by there being many examples of “men of war” in the port, illustrating how important Brest was as in terms of naval conflicts of the time.

This approach therefore offers two ways of helping us understanding large volumes of text such as newspapers. First we can use it effectively as a metadata tool to assist us in finding the right articles to read if we are interested in the geographies of the past. Second, it also has an analytic function in allowing us to understand how different places were perceived differently from each other.

#### **Slide 20: Conclusions**

## Conclusions

- **Pilot work with more to do but:**
  - We can use texts within a GIS
  - We can geo-reference them in a semi-automated way to ask “what were they saying about where?”
  - Allows us to integrate a vast amount of disparate data sources:
    - Historical texts, modern images, historical maps, modern topographical data, etc.
  - Has many potential applications in the digital humanities

So far this work is in its early stages but it clearly has much potential. Firstly, we have illustrated that at a technical level it is possible to create GIS databases from large volumes of text. Secondly, and more importantly, it has illustrated that mapping place-names provides a useful scholarly tool to help the researcher to understand these texts. How this mapping is implemented is up to the researcher who can follow a distant reading approach, in which the texts are abstracted and summarised, or, as the work with Google Earth illustrates, follow more traditional forms of literary scholarship by helping the reader understand the complex geographies within the texts under study. One key advantage of a GIS that is perhaps understated is its ability to integrate apparently disparate data sources. In the *Mapping the Lakes* project two different texts were integrated, and these were then further integrated with a wide variety of other sources including: historical maps (added to Google Earth), modern user generate content (from Flickr), modern areal photographs (from Google Earth), modern topographical data, historical census data, and do on. All of these can be used to contribute to our understanding of either or both of the texts. The main conclusion from this work is, however, a simple one: GIS has much to offer to scholarship within the Digital Humanities.

\*Department of History, Lancaster University, Lancaster, LA1 4YG, UK. Email [I.Gregory@lancaster.ac.uk](mailto:I.Gregory@lancaster.ac.uk)

\*\*Department of English and Creative Writing, Lancaster University, Lancaster, LA1 4YD, UK

### Further reading:

- Andrews M. (1994, ed.) *The Picturesque: Literary sources and documents. Vol. I* Helm: Mountsfield
- Bodenhamer D.J. (2007) "Creating a landscape of memory: The potential of humanities GIS" *International Journal of Humanities and Arts Computing*, 1, pp. 85-96.
- Cooper D. and Gregory I.N. (Jan 2011) "Mapping the English Lake District: A literary GIS" *Transactions of the Institute of British Geographers*, 36, pp. 89-108.
- Dunning A., Gregory I. and Hardie A. (2009) "Freeing up digital content with text mining: New research means new licences" *Serials*, 22, pp. 166-173.
- Gregory I.N. and Cooper D. (2009) "Thomas Gray, Samuel Taylor Coleridge and Geographical Information Systems: A Literary GIS of Two Lake District Tours" *International Journal of Humanities and Arts Computing*, 3, pp. 61-84.
- Gregory I.N. and Ell P.S. (2007) *Historical GIS: Techniques, methodologies and scholarship*. Cambridge University Press: Cambridge.
- Gregory I.N. and Hardie A. (in press) "Visual GISing: Bringing together corpus linguistics and Geographical Information Systems" *Literary and Linguistic Computing*
- Moretti F. (2005) *Graphs, Maps, Trees*. Verso: London
- Flickr website: <http://www.flickr.com>

The Lancaster Newsbooks Corpus:

<http://juilland.comp.lancs.ac.uk/hardiea/newsbooks/index.php>

Mapping the Lakes: <http://www.lancs.ac.uk/mappingthelakes>

World Gazetteer: <http://www.world-gazetteer.com>

### **Acknowledgments:**

The work on the Literary Mapping of the Lakes project was made possible by an award from the British Academy (SG46004). The authors would also like to acknowledge their gratitude to Sally Bushell and Zoe Bolton (Department of English & Creative Writing, Lancaster University) for their assistance with this project. Additional work on Wordsworth was done by Kirsten Hansen while she was an intern at Lancaster University in the summer of 2009. As well as thanking her we would like to thank Prof R. Schwartz (Mt Holyoke College) and the Department of History, Lancaster University for their help in arranging this internship. The work on the 17<sup>th</sup> century newsbooks was done in collaboration with Andrew Hardie (Department of Linguistics and English Language, Lancaster University).

We would also like to thank Oxford University Press for their kind permission to quote from: Gray (1971) *Correspondence of Thomas Gray*, and Coleridge (1956) *Collected Letters of Samuel Taylor Coleridge* (<http://www.oup.co.uk>). Extracts from Coleridge (1957) *The Notebooks of Samuel Taylor Coleridge* are reproduced by permission of the Taylor & Francis Group (<http://www.taylorandfrancisgroup.com>). This work uses data provided through EDINA UKBORDERS with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown.