



## Geographies of Information in Africa Wikipedia and Uneven User-Generated Content

By Dr. Mark Graham, Oxford Internet Institute

**W**e all know that Wikipedia is an immense project. It is an incredibly impressive coming-together of human labour on a scale that the world rarely sees. Over the last few years, I have been working on mapping the encyclopaedia. My work has shown that the project is far from complete.

In the map below, I offer a visualisation of all English-language Wikipedia articles on the planet shaded according to the number of words in each article. Yellow dots represent the location of relatively short articles (such as Nyanza), while red dots indicate the location of relatively long articles (for instance Nairobi). A high-resolution version is also available from my website (

Geotagged Articles in English Wikipedia

geography.net). Each one of these yellow dots represents the human effort that has gone into describing some aspect of a place.

The density of this layer of information over some parts of the world is astounding. Much of my work looks at inequality in user-generated content, but it is still hard not to be awed by this cloud

of information about hundreds of thousands of events and places around the globe.

There are a staggering number of articles in the United States (over 180,000 of them) and tens of thousands in many European countries, Japan, Australia and India. It is also important to point out that there are far fewer in much of the rest of the world. In fact, there are only a few countries in Africa that contain more than 1000 articles!

Interesting patterns emerge: the average word length of articles in the US is 750, while many European countries have lower means: e.g. Italy (550), Germany (439), Spain (397), France (260), and Poland (233). Articles in the UK and Ireland average 687 and 749 words respectively. The immediate conclusion here should be that it is easier for editors in English speaking countries (all of which tend to have high averages) to expand articles than editors in countries in which English isn't the native language.

But the native language of a country clearly isn't the only factor at play. The countries with the highest average word counts to their articles are Iraq, with an average of 1091 words in its 538 articles, the Philippines, with an average of 1085 words in its 2736 articles, and North Korea, with an average of 947 words in its 292 articles.

On the bottom end of the scale we have Azerbaijan (159), Estonia

(209), and Kenya (223). The results tell us that there are apparently a lot of stub articles written about Azerbaijan, Estonia and Kenya. Whereas there are very few stubs in places like Iraq and North Korea: a finding that makes a lot of sense.

It must be very hard for English speaking editors to create articles (even stub articles) about things like small stadiums in provincial towns in North Korea and Iraq. But uploading this sort of information about the equivalent type of place in Estonia or Kenya is far less of a problem.

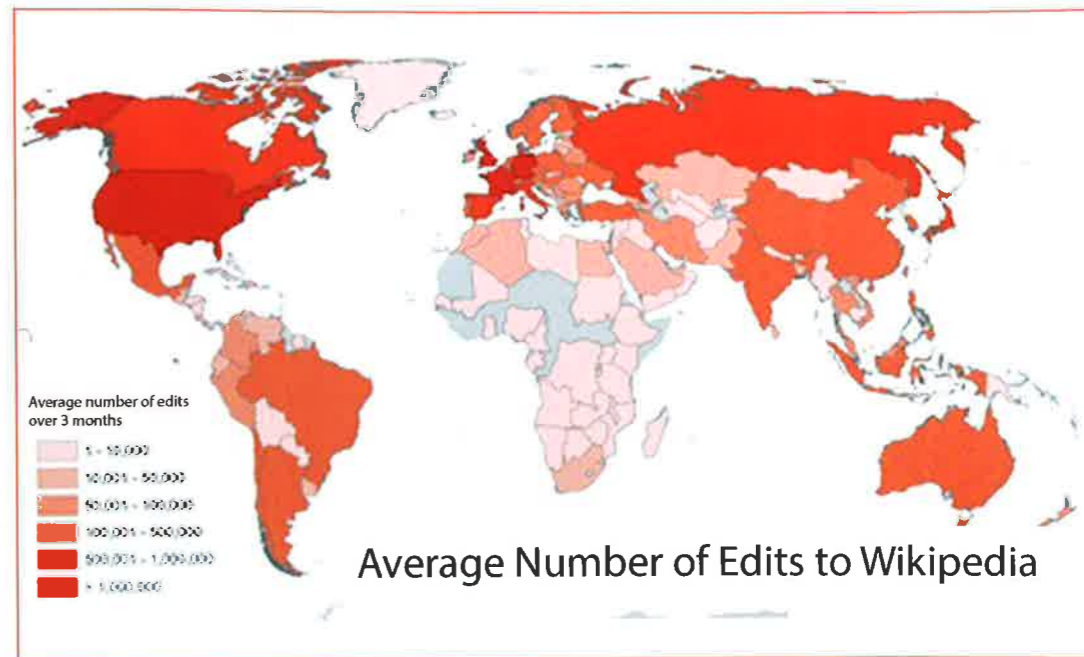
It is maybe even more important to look at where all of this information is coming from. The map below shows you the average quarterly number of edits over a two-year period (2010-2011). The inequalities in the amount of content produced are stark: the US, Germany, the UK and France all have an average of over a million edits each quarter.

However, when you look at most of Africa and the Middle East, the average number of edits per quarter is only a few thousand. Interestingly, there are more edits that originate in Hong Kong each quarter than the entire continent of Africa.

Much of this variation can actually be explained by internet population (i.e. the total number of internet users in a country). However, even accounting for their generally low internet populations, most countries in Sub-Saharan Africa still fall below their expected number of edits (my team and I are currently working on some statistical models and writing a paper about this topic).

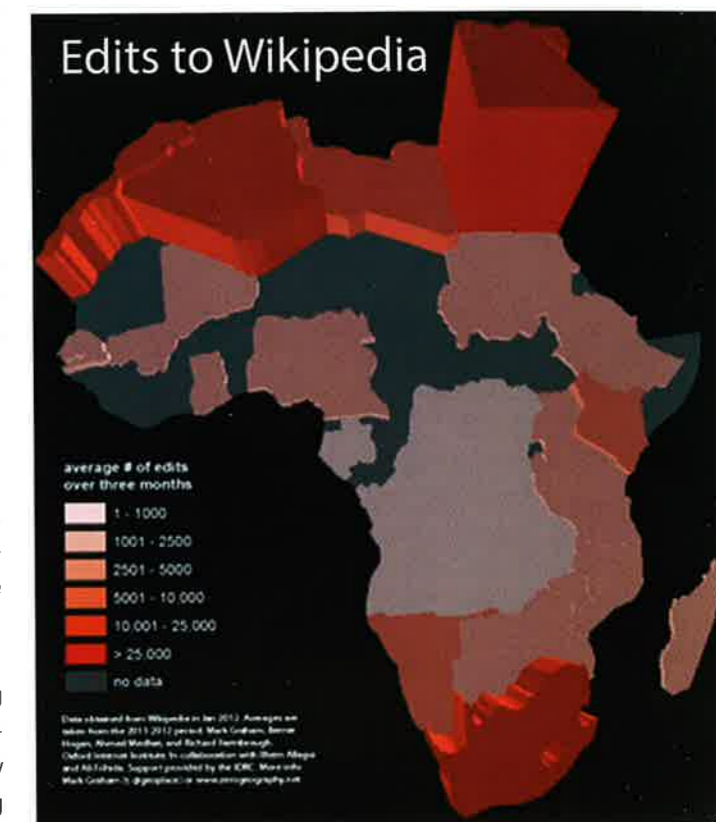
Finally, if we zoom into the number of edits to Wikipedia emerging from Africa, we get a much better sense of the scale of the differences in contributions. In particular, the map really highlights how much participation we're seeing from Egypt. It is also worth noting that Rwanda and a few other African countries don't even show up on the map due to having so few contributions to even register in the data collector.

I think what this work shows us is that even though Africa now has almost 150 million internet users, it remains largely left out of one of the world's most important and visible sources of information. Wikipedia is a platform that prides itself in lack of barriers to the access and creation of knowledge, so we need to then ask why Africa is both literally and figuratively left off the map. We unfortunately don't yet have definitive answers, but it is this question that will guide the work of my research team over the next few years.



Average Number of Edits to Wikipedia

World map showing number of edits to wikipedia



Map of Africa showing edits to wikipedia

If you have any opinions or insights into these preliminary results, please do get in touch through either my twitter account (@geospace) or website ([www.geospace.co.uk](http://www.geospace.co.uk)).

Dr. Mark Graham is a Research Fellow at the Oxford University's Oxford Internet Institute