WEB 2.0 GEOTAGGED PHOTOS: ASSESSING THE SPATIAL DIMENSION OF THE PHENOMENON

Vyron Antoniou, Jeremy Morley and Mordechai Haklay
Department of Civil, Environmental and Geomatic Engineering, University College, London

Among popular Web 2.0 applications are the social networking, photo-sharing websites like Flickr, Panoramio, Picasa Web, and Geograph. The phenomenon of user-generated content and the increased presence of geographic information in such applications have motivated researchers to consider them as a source of geographic information. In this paper we question whether such web applications can serve as reliable sources of spatial content. We differentiate between spatially explicit and implicit applications, in accordance to their declared aims, and evaluate if they have an impact on the spatial distribution of geotagged photos for Great Britain. We also compare the spatial distribution of the photos submitted with population data, and the patterns of contribution to these sources over a period of 18 months. Finally, at a larger scale, we examine the spatial distribution of photos and their spatial density for 15 test areas and look into issues such as data currency and user behaviour. Our findings show that only web applications that urge users to interact directly with spatial entities can serve as reliable, universal sources of spatial content.

Parmi les applications Web 2.0 populaires, nous avons accès aux sites Web de réseautage social et de partage de photos comme Flickr, Panoramio, Picasa Web et Geograph. Le phénomène du contenu généré par l'utilisateur et la présence accrue de l'information géographique dans ces applications ont motivé les chercheurs à les considérer comme source d'information géographique. Dans le présent article, nous tentons de déterminer si de telles applications Web peuvent servir de sources fiables de contenu spatial. Nous faisons la différence entre les applications spatialement explicites et implicites, conformément à leurs buts déclarés. Nous évaluons aussi si ceci a une incidence sur la distribution spatiale des photos avec balise géographique (géotag) pour la Grande-Bretagne. De plus, nous comparons la distribution spatiale des photos soumises avec les données sur la population et les modèles de contribution à ces sources sur une période de 18 mois. Enfin, à plus grande échelle, nous examinons la distribution spatiale des photos et leur densité spatiale pour 15 régions d'essai et évaluons des questions telles que l'actualité des données et le comportement des utilisateurs. Nos constatations démontrent que seules les applications Web qui incitent les utilisateurs à interagir directement avec les entités spatiales peuvent servir de sources universelles fiables de contenu spatial.

Vyron Antoniou

Jeremy Morley

Mordechai Haklay

1. Introduction

The advances in the technologies associated with the World Wide Web (the Web) over the past few years have been significant, leading to the identification of a step change in web technologies that is termed Web 2.0. While users contributed to the Web in its early days, Web 2.0 users are now central to the process of creating, sharing, consuming and disseminating information. This bi-directional flow of information leads to the growth in user-generated content (UGC). UGC has also affected the delivery of Geographic Information (GI) on the Web. By today's standard, the majority of web maps around 2005 had a very basic and trivial form. Today, common users and experts interact and publish spatial information using sophisticated web mapping services ranging from Google Maps, Microsoft's Bing Maps, or Yahoo! Maps, through to Virtual Globes like Google Earth or NASA's World Wind.

The growth of UGC opens up new opportunities for mapping the world. As *Goodchild* [2007b] states, the arguments made by *Estes and Mooneyhan* [1994], about a mistaken popular notion of a well-mapped world, are still true. In fact, mapping and map updating programs are suffering serious delays in many countries due to budget limitations and rapid development. Thus, in the area of GI, it seems that UGC opens up the possibility of supplying information that can be used to update and maintain existing databases. The core question that is raised by this suggestion is, how effective is this data for updating existing spatial information databases?

Researchers have acknowledged the fact that the evolution of the Web and the maturation of certain enabling technologies, such as Javascript Application Programming Interfaces (APIs), GPS-

enabled everyday devices and the AJAX technology [*Miller* 2006; *Turner* 2006; *Goodchild* 2007a; *Haklay et al*. 2008], have paved the way for collecting, managing and delivering GI over the Web. Moreover, reduced entry costs in the field of mapping and laypersons' familiarization with the subject of geography [*Goodchild* 2008a], leads to a situation where there is no longer a firm need for mapping expertise in order to publish maps on the Web [*Goodchild* 2008b]. Together with the high cost of spatial data [*Goodchild* 2008b], both have played their role in the rise of neogeography and Volunteered Geographical Information (VGI).

Photo-sharing websites emerged around 2004, and the social impact they had has motivated many researchers to examine and analyze further this phenomenon and its possible applications.

As the phenomenon of VGI started to draw attention among scholars, a growing debate about the term and specifically the word 'volunteered' began [*Elwood* 2008a]. Researchers [*Obermeyer* 2007; *Sieber* 2007; *Elwood* 2008b; *Bishr and Mantelas* 2008] suggest that the term 'volunteered' can be misleading with regards to the particularities of the generated data and the intentions of the data providers. In a sense 'volunteered' implies a noble and altruistic gesture, as if the users donate the data, personal or not, to the world for any use. Acknowledging these issues raised, a more general but still precise descriptive term is used in this paper—User Generated Spatial Content (UGSC).

In order to evaluate the usability of UGSC for updating existing data sets, we focused on one class of UGSC source—the social photo-sharing web applications of Flickr, Panoramio, Picasa Web and Geograph. In these web applications users can upload photos, add titles, tags, descriptions, and comments to the photos, form groups, and socialise with other users. In addition, geographic information can be added to the photos. The photos can be geotagged (i.e. associated with the co-ordinates of their capture locations) and might depict spatial entities that can provide useful information such as the development of a new building, or indication that a new road has been opened. The captions and tags (i.e. words that describe the content of the photo) might refer to place names or administrative boundaries. User groups formed might focus on spatially related issues like construction sites or points of interest. These web applications are based on the fundamental principles that powered the evolution of Web 2.0 [*O'Reilly* 2005] with little or no differentiation regarding their structure, functionality, or the need for user participation. From a geographic information retrieval (GIR) point of view, we suggest that such web sources can be categorized as *spatially implicit or spatially explicit. Spatially explicit* applications, like Geograph and Panoramio, urge their contributors to interact directly with spatial features (i.e. to capture spatial entities in their photos) while at the same time encourage that photos, and thus the content, be spatially distributed—for example, Geograph is geared toward capturing photos for each square kilometre in the U.K. In contrast, Flickr and Picasa Web are more socially oriented, as they are aimed at allowing people to share their photo albums, and thus are regarded as *spatially implicit* web applications. The support of geotagged photos is one of the many interesting features that spatially implicit applications have, but spatial information is neither one of the core features nor is it the main motivation of their users, in contrast with what takes place in spatially explicit applications, in which the users are explicitly expected to use geography and location as a motivational and organisational factor. Our research clearly shows that not all applications are suitable to serve as sources of spatial content despite the magnitude of contribution usually associated with such applications. Given the current rapid growth in interest of the GI community for UGSC, there is a need to investigate the spatial dimensions of the phenomenon in order to avoid potential pitfalls, understand the possible contribution of social networking websites, and provide the baseline for further research.

In the following, we begin with an examination of related research in this field. Next, we discuss the data sources and the technologies used in this research (Section 3), followed by the presentation and analysis of the data collected (Section 4). In Section 5, we elaborate upon our findings. Finally, we close with some conclusions and indications of our future work.

2. Related Research

Photo-sharing websites emerged around 2004, and the social impact they had has motivated many researchers to examine and analyze further this phenomenon and its possible applications. For example, researchers have started to examine: the novelties offered by photo-sharing applications like Flickr that stimulate users' participation [*Cox* 2008], the role that such applications can play in times of disaster and crisis [*Liu et al*. 2008], the motivations that make users add tags to photographs [*Ames and Naaman* 2007], and why and how groups form and interact in the context of photo-sharing websites such as Flickr [*Negoescu and Perez* 2008].

In terms of GIR, as *Jones and Purves* [2008] note, only in recent years have we seen efforts to retrieve GI from the Web. Much of the ongoing research in this direction has used text documents as input to: detect topological relationships among

places [*Schockaert et al*. 2008], automatically create gazetteers by combining official data with informal data harvested from the Web [*Goldberg et al.* 2009], create gazetteers of informal place names [*Twaroch et al*. 2008; *Jones et al*. 2008], or create region boundaries by delineating imprecise regions [*Arampatzis et al*. 2006]. Another class of applications uses photos and associated text from photo-sharing websites as input. *Purves and Edwards* [2008] used UGSC to define descriptive terms for place, *Dykes et al* [2008] used visualization tools to explore the spatial distribution of descriptive terms for place, and *Popescu et al*. [2008] presented a method for automatically creating gazetteers by using both photo-sharing and text-based websites.

In summary, while the nature of photo-sharing websites received attention, there is no systematic analysis of their use as a potential source of GI.

3. Data Collection

In this study, the sources of data used here include Flickr, Picasa Web, Panoramio, and Geograph. The reasons for choosing these websites are twofold. Firstly, these photo-sharing applications are among the most popular websites and secondly, Application Programming Interfaces (APIs) are provided for them which allow access to their data. Apart from the commonalities among the chosen sources, there are also some differences that make each source unique compared with the other sources. Flickr is a social networking website that allows users to publish and share photos of any content. Users can also comment and add tags and descriptions to photos. In contrast with Flickr where this functionality is only available online, Picasa Web is a similar Web application which is also supported through a desktop application (i.e. Picasa) that allows users to perform those actions for their photos off-line and upload the content when they wish to do so. Panoramio is a web application that urges users to freely upload photos to describe places that they like or want to annotate. Finally, Geograph encourages its users to submit photos for every square kilometre of the U.K. and Ireland and thus declares a more precise aim to attract user participation and geographic coverage. From these four sources, we classify Geograph and Panoramio as spatially explicit sources since they urge their users to directly capture geographical information in their photographs and submit it to the web application. Flickr and Picasa are spatially implicit, as their main aim is to allow users to manage their photos, and the geographical functionality is an add-on.

For assistance in data visualization and gathering, we also used two datasets provided by Ordnance Survey (OS): The Great Britain boundary and the 1 km^2 National Grid. We used the polygons of the Great Britain boundary to select the National Grid tiles that would be used for our study, 238 920 tiles in total. In order to examine the spatial distribution of the geo-tagged photos, we examined how many photos have been submitted to each selected source for each one of these tiles. Thus, our results have a spatial resolution of 1 km^2. In order to compute the population density surfaces (see Section 4.4) we used population data for Great Britain provided by CASWEB service.

We collected data in three different stages. In the first, we collected the number of photos in each of the 238 920 tiles from all four sources to understand the general pattern of UGSC distribution. The second data collection was performed only for the most popular tiles of Flickr and Geograph (see Section 4.4 for more details). Finally, to examine the phenomenon in more detail, below the coarse level of 1 km, we chose 15 areas in Great Britain (Table 1) and collected detailed data for 1/25 (141 km^2) of the common popular tiles for both Flickr and Geograph, from January 2005 until April 2009. This dataset consisted of the actual photographs submitted (50 504 from Flickr and 11 937 from Geograph), details associated with each photo (e.g. title, tags, comments, dates taken/posted), the users (e.g. user name) and the location (e.g. geometry).

Study Area	Torquay	North London	Chester	Leeds	Dundee	Swindon	Oxford	Chatham	Glasgow	Edinburgh	East London	West London	Cambridge	Portsmouth	Nottingham		
Num. of Tiles	10	15	9	16	5	6	14	12	6	9	9	9	6	6	9	141	Total Num. of Tiles
Num. of Flickr Photos	930	7993	3753	3642	1156	674	11861	2266	625	6837	2337	3444	673	2982	1331	50504	Total Num. of Photos
Num. of Geograph Photos	887	1114	1781	1888	215	465	704	874	224	472	341	347	1047	1187	391	11937	

Table 1: The selected study areas in Great Britain.

4. Results

4.1 Descriptive Statistics

The analysis began by calculating the descriptive statistics of the different datasets (Table 2). The source with the most geo-tagged photos for Great Britain was Flickr with 1 654 277 photos, followed by Picasa Web with 1 255 515, Geograph with 1 078 150 and lastly Panoramio with 410 236 (this might be affected by the Panoramio's API as it seem to provide inconsistent results). The same ordering of sources applies to the maximum numbers of photos per tile. An indication of distribution can be seen in Geograph's maximum value (1 317 photos) which was significantly lower compared to the maximum values of other sources (Flickr 38 506, Picasa Web 31 947 and Panoramio 10 191 photos).

4.2 Spatial Distribution

In the next step we calculated the frequency of the number of photos submitted per tile (Figure 1).

	Num. of Tiles	Min	Max	Sum	Mean	Std. Deviation
Geograph	238920	0	1317	1078150	4.51	12.95
Flickr	238920	0	38506	1654277	6.92	196.56
Panoramio	238920	0	10191	410236	1.72	34.97
Picasa Web Albums	238920	0	31947	1255515	5.25	136.46

Table 2: Statistics of the photos submitted to the selected photo-sharing websites (Geograph, Flickr, Panoramio, Picasa Web Albums).

It is evident that, for all the sources except Geograph, the majority of tiles contain no geotagged photos from those that have been submitted since the websites were launched. More specifically, for Picasa Web there are no photos for 72.7% of the tiles, for Flickr the percentage is 84.6% and for Panoramio 86.3% (Panoramio's API did not report back tiles that had a single photo). In contrast, for Geograph only 9.2% of the tiles do not contain a geotagged photo.

While the sheer numbers of geotagged photos available from the selected social-networking websites can give us an understanding about the magnitude of the phenomenon, they do not provide adequate spatial distributions. In order to investigate this specific issue we associated the 1 km^2 National Grid with the data collected from the web sources and visualised the results. Visualization is a powerful method for data exploration and hypotheses testing, especially for large and unknown datasets [*MacEarchren and Kraak* 2001]. Figure 2 shows the spatial distribution of each source's photos.

Although the distributions differ between the sources, it can be suggested that there are two different patterns of distribution. The first (to which only Geograph belongs) covers almost all of the area with only a small number of clusters. Geotagged photos submitted to Geograph leave very few blank spots which are mostly in the sparsely populated, barren areas of the Highlands in Scotland. The stated aim of Geograph that was noted above has proven to be a strong motivation for the users to provide consistent

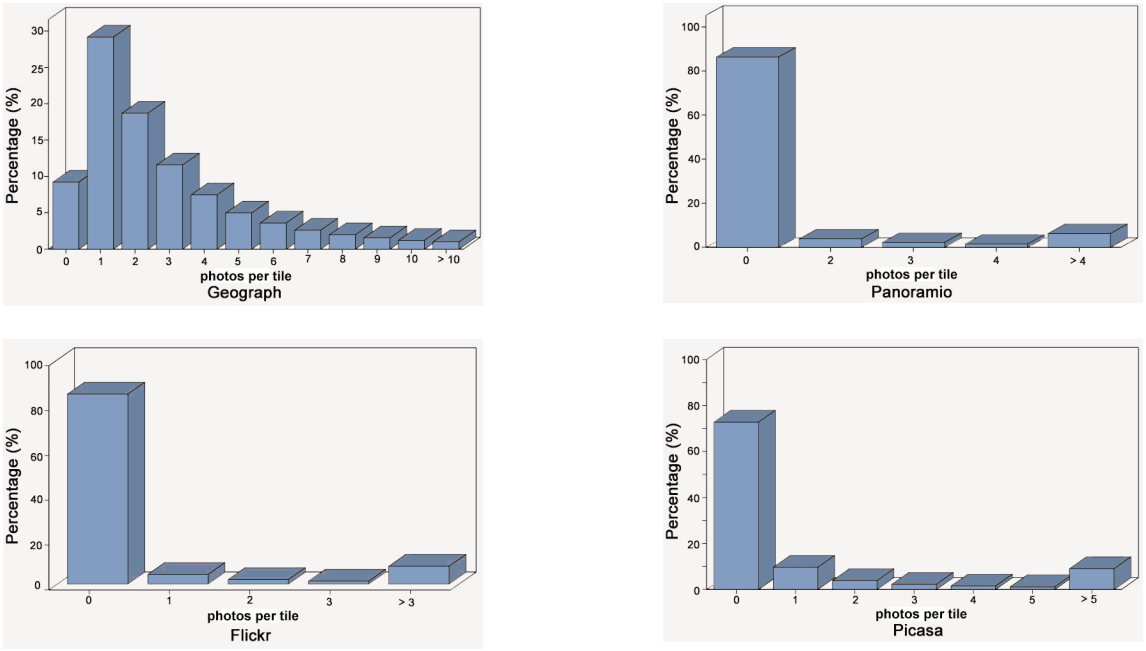


Figure 1: Frequency of photos per tile for each of the selected sources.

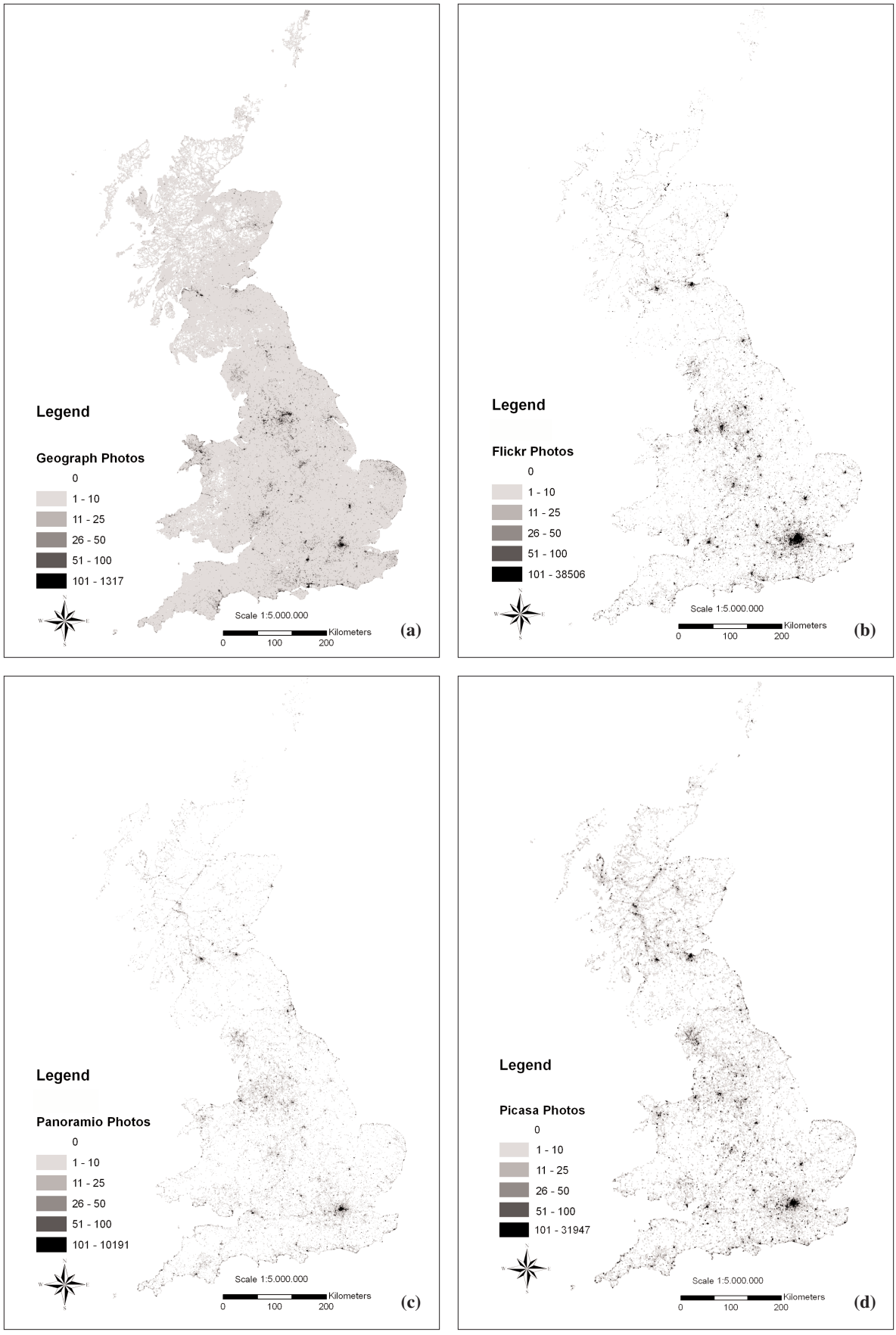


Figure 2: The spatial distribution of photos in (a) Geograph, (b) Flickr, (c) Panoramio, and (d) Picasa Web Albums.

coverage. Nonetheless, for the majority of the tiles, only a few photos have been submitted since the launch of the website in 2005; for example, almost 60% of the tiles have three or fewer geotagged photos. This does not apply to urban areas and tourist attractions as users' participation in these areas is more intense and thus clusters appear. For example, for Greater London area, a total of 35 275 photos have been submitted which approximately corresponds to 22 photos per km^2. This number is almost five times higher than the overall average of 4.5 photos per km^2.

In contrast, the second pattern of distribution (to which Picasa Web, Flickr and Panoramio belong) covers a small percentage of the study area and their pattern is considerably more clustered in urban areas and tourist attractions. For example, 338 198 photos have been submitted to Picasa Web for the Greater London area, which corresponds to an average of 211 photos per km^2, which is 40 times more than the overall average of only 5.25 photos per km^2. Interestingly, while Picasa Web has fewer geotagged photos than Flickr, overall, it provides better coverage of the study area.

Finally, using 3D visualization (Figure 3) we can understand the magnitude of the phenomenon in the clustered areas.

We suggest that this diversity in the spatial patterns of web sources is due to the difference in the nature of the web applications. The spatially explicit source of Geograph provides a more distributed coverage than the spatially implicit sources of Flickr and Picasa Web. Interestingly, Panoramio's spatial distribution is closer to the distribution of the spatially implicit sources, in contrast with our initial hypothesis. This can be partially caused by its API inconsistencies, but taking also into account the photo frequencies (Figure 1), it is reasonable to assume that Panoramio behaves in a socially-oriented manner, and thus like a spatially implicit application. An explanation of this might be that

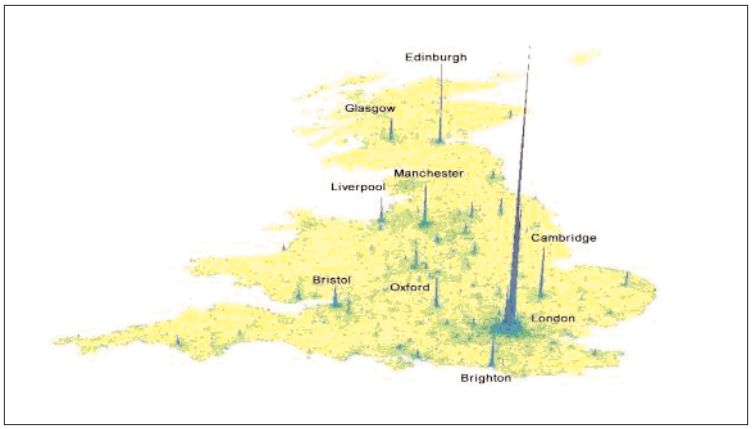


Figure 3: 3D visualization of the geotagged photos collected from Fickr.

while Panoramio is spatially explicit, it does not encourage complete spatial coverage, and therefore its users are tagging places that they like or visit, which are similar to those of the other sites.

4.3 Expectation Surfaces

As *Dykes and Wood* [2008] explain, building a surface that both relates to the intensity of a phenomenon and to the location of people, allows the statistical analysis to be carried out. This can be accomplished by calculating expectation surfaces using the chi-statistic:

$$chi = \frac{(ObservedValue - ExpectedValue)}{\sqrt{ExpectedValue}}$$

In our case, the observed values are the number of photos submitted for a tile and the expected values are from a population density surface based on 2001 population data from CASWEB. This comparison will allow us to understand the correlation of UGSC phenomenon with population density.

The chi index will be negative for the tiles where the observed value (the number of geotagged photos) is lower than expected (according to population data) and positive when it is greater than expected. Figure 4 shows the expectation surfaces for the four different sources; the purple (dark) shade indicates areas where there are more photos than expected, compared to the underlying population, in contrast with the cyan (light) areas. In the majority of the study areas, the number of photos from Geograph is higher than expected versus the underlying population. In spatial terms, this characteristic can prove valuable for a crowd-sourced application because it demonstrates that the population is not the only predictor of data collection, and a linear distribution assumption, that where there are people data will be collected, is invalid. In contrast, the other sources cover only the tourism areas, where few people have permanent residence, better than expected.

Figure 5 shows a magnified view of the chi expectation surface for the Greater London area. As can be seen, even in the urban areas where users' contribution is high, the concentration of photos is confined to a core at the centre of the city, ranging from a very small area, for Geograph, up to a broader one for Flickr. In all four cases though, the outer suburbs of London are underrepresented.

4.4 Data Flow

Another aspect of our study was to examine the data flow (i.e. the number of photos submitted)

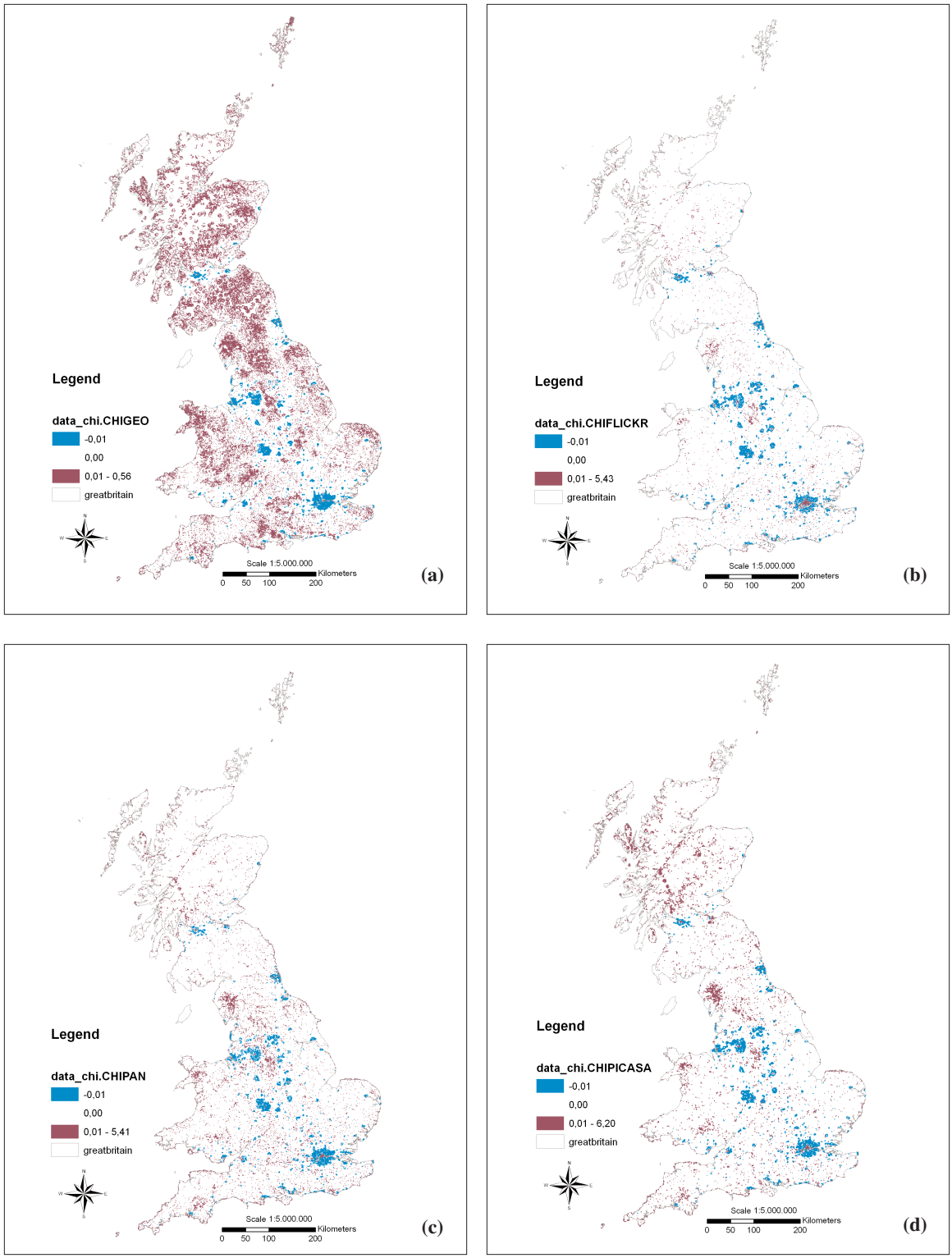


Figure 4: Expectation surfaces from (a) Geograph, (b) Flickr, (c) Panaromio, and (d) Picasa Web Albums.

for the most popular tiles of Geograph and Flickr. A threshold of 15 photos was set to characterise a tile as 'popular,' as this represents an average of one submitted photo per tile per quarter over the four year period. Only tiles which have a total of 15 or more photos since the launch of the web application were included (Figure 6). In this category there are 12 081 tiles for Geograph and 8 889 tiles for Flickr, covering just 5.06% and 3.72% of Great Britain respectively. Significantly, although Flickr has 65% more photos than Geograph, its popular tiles cover 26% less area than Geograph.

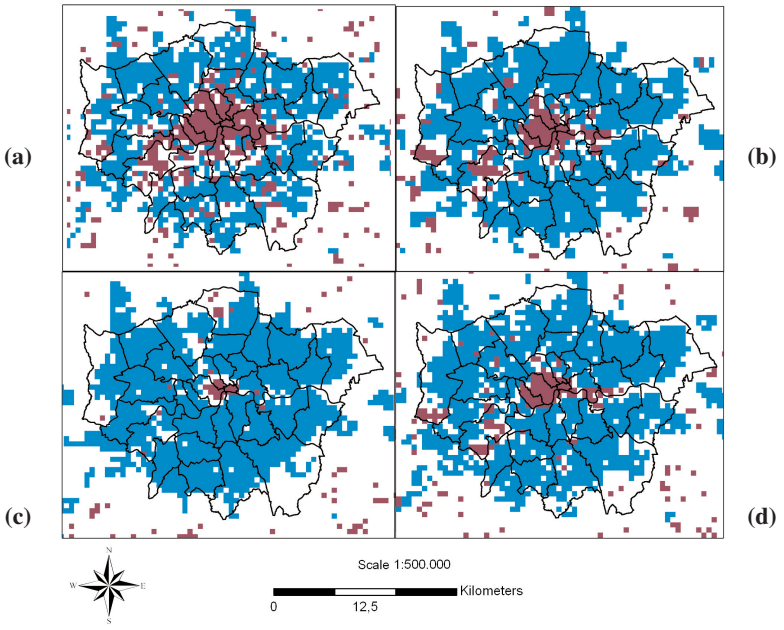


Figure 5: **The chi expectation surfaces for the area of Greater London: (a) Flickr, (b) Picasa (c) Geograph (d) Panoramio**

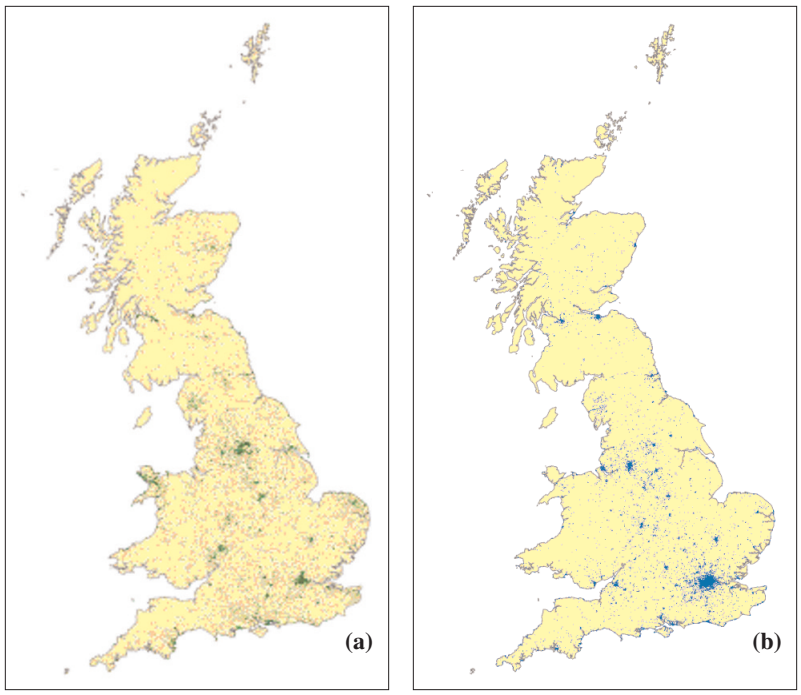


Figure 6: The most popular tiles (with 15 or more photos) in (a) Geograph and (b) Flickr.

	09/2007 - 02/2008	03/2008 - 08/2008	09/2008 – 02/2009
Geograph	7072 (2.96%)	8927 (3.74%)	7879(3.30%)
Flickr	5884 (2.46%)	6474 (2.717%)	6249 (2.62%)

Table 3: The number of popular tiles for which at least one geotagged photo has been submitted over a period of 6 months. In the parenthesis is the percentage area coverage of Great Britain.

Additional to the data flow, the currency of the data available for the popular tiles was examined. When it comes to the use of spatial information (for example, in updating mapping products), currency of data is paramount. Table 3 shows the number of popular tiles, and the overall percentage of area coverage that they represent for Great Britain, for which at least one geo-tagged photo has been submitted over a period of three consecutive six-month periods.

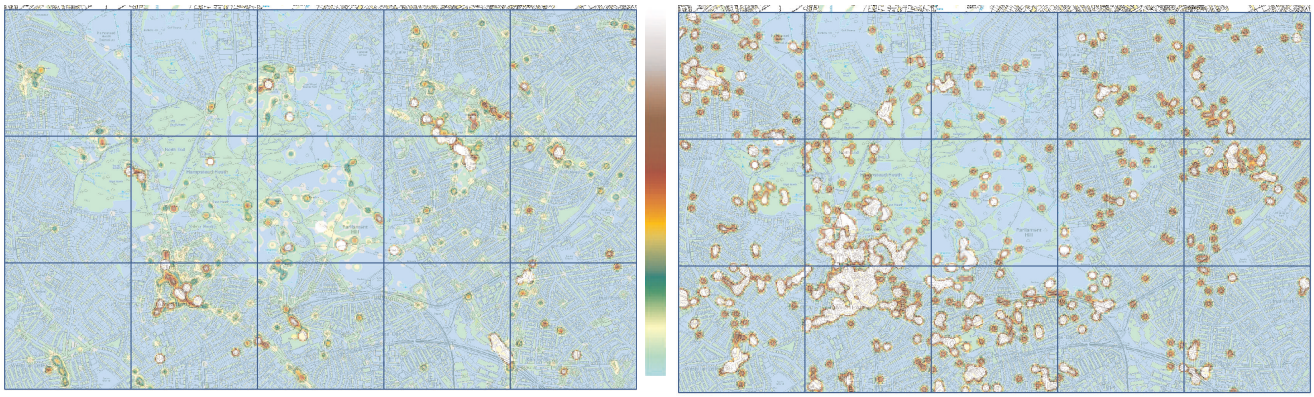
From the sub-group of the popular tiles, 3 782 tiles of Geograph and 3 912 tiles of Flickr had photos submitted for every six-month period.

4.5 Density Analysis

In the next step, the 15 selected areas shown in Table 1 are examined. The detailed datasets collected for these areas and especially the co-ordinates of where each photo was taken, allowed a kernel density analysis for these areas which examines the spatial distribution of the phenomenon at a large scale and compares the behaviour between explicit and implicit sources. This analysis shows that spatially explicit sources provide better coverage of the study areas, even with fewer photos. For example, Figure 7 shows the density analysis for an area of 15 km^2 located in North London. The spatial resolution of the density surface is 10 m, which is approximately the accuracy of a geo-tagged photo and the kernel radius is 50 m. Figure 7a shows the density surface created from photo capture points from Flickr since 2005 (7 993 in total) and Figure 7b shows the density surface for Geograph computed from 1 109 points for the same period. It is clear that, although Flickr has 6.2 times more photos than Geograph in this area, the spatial distribution of the photo-capture points from Flickr is concentrated in a few relatively popular spots (e.g. Parliament Hill and Dartmouth Path at Hampstead in North London). In contrast, for Geograph, the distribution of the photo capture points is more dispersed covering a considerably larger portion of the area.

In an effort to quantify this observation in more practical terms, an event analysis for each study area was carried out. This allowed the quantification of the repetition observed in the photo-capture locations (i.e. camera locations) for each source. Figure 8 shows the percentage of different photo-capture locations for each study area.

The average percentage of unique locations for Flickr is 30.1%, in contrast with 85.6% for Geograph. This means that, on average, 100 photos in Flickr would be taken from only 30 different camera locations in contrast with 85 different locations for 100 photos in Geograph.



(a) (b)

Figure 7: Density surfaces for (a) Flickr, and (b) Geograph.

4.6 User Behaviour

The final analysis focused on user behaviour wirh regard to user productivity, as measured by the time difference between capturing and uploading a photo, and also with regard to the time that users remain active in an area (i.e. time difference between first and last photo submission for each user). This will reveal the currency of the photos submitted to such applications and provide an insight into how users' participation is evolving through time.

Geograph users appear to be more productive than those from Flickr. There are 3 236 Flickr users that have uploaded photos for these 15 study areas in contrast to 538 Geograph users, which mean that individual contribution for Flickr users is 15.6 photos per user in contrast to 22.2 photos per Geograph user. Apart from this, user behaviour is quite similar for both sources. Figure 9 shows the percentage of photos submitted in various time spans (the negative values are due to falsely recorded timestamps by either the users or the web sources). Users usually upload their photos in the first couple of weeks after the capture date. Very few photos (8.4% for Flickr and 9.2% for Geograph) are more than a year old which indicates that such sources can be used in applications that need contemporary photos. Finally, Figure 10 shows how long users remained active in these areas. For both web applications, more than 50% of their users captured data over a period of less than a day and thus it can be suggested that these users were just visiting the area. On the other hand, a rough estimation can be made about the users that are permanent residents in an area (for example, by calculating the users that have stayed active for more than 14 days). This is important because local knowledge has been widely acknowledged as a very important factor in retrieving geographic information [see for example: *Haklay and Tobon* 2003; *Dunn* 2007; *Budhathoki et. al* 2008; *Elwood* 2008c].

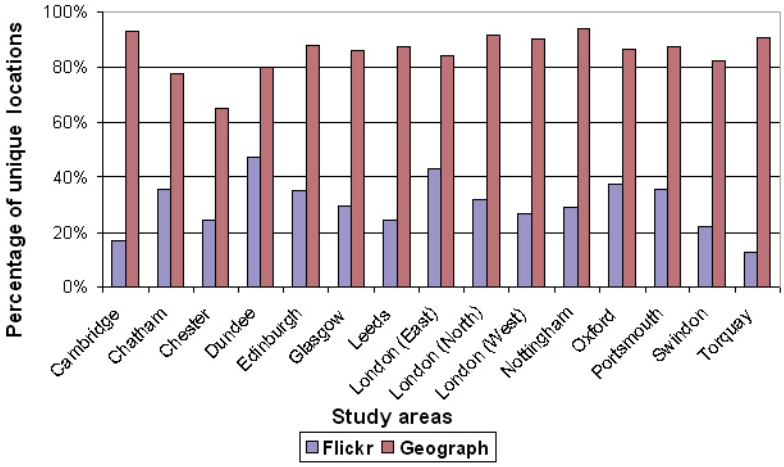


Figure 8: Percentage of unique camera location for each study area.

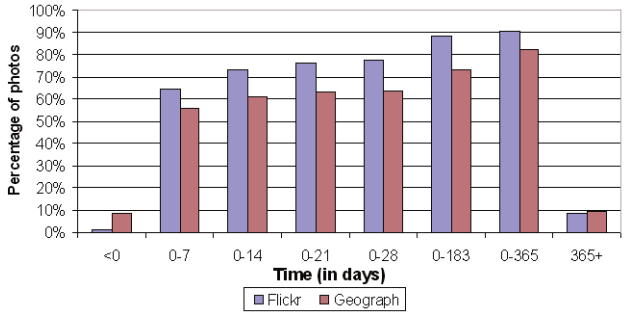


Figure 9: Time difference between capturing and submitting a photo.

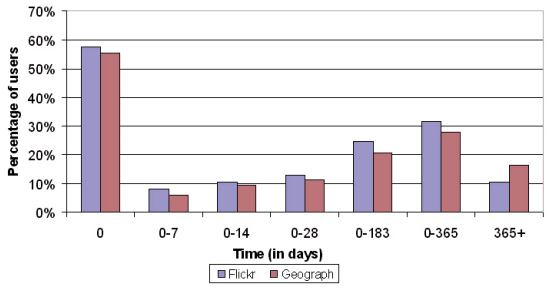


Figure 10: Periods of user activity.

5. Discussion

The magnitude of the activity in photo-sharing websites is arguably impressive. Recording user participation from just four web sources provided over 4.4 million geo-tagged images for Great Britain alone over a period of 18 months. This number indicates that people upload massive quantities of geographic information to the Web. Although such information is loosely structured, given the proper tools and methodologies, GI retrieval from the Web can provide valuable information. However, the study described here shows that a core question is, 'what kind of GI can each web source provide?' Our findings clearly show that photo-sharing web applications can be grouped into spatially explicit and spatially implicit ones. In the first group, GI is the main characteristic of the application. Stated directly or not, explicit web applications urge their users to collect spatial information that complies with some sort of loosely defined specification(s) (e.g. "…submit a photo of a place…," "…cover every square kilometre…"). This results in the creation of a dataset that is richer in GI and better positioned in time and space than the implicit sources. In implicit sources, GI is not prioritized over the rest of the features offered by the web application and thus GI is firmly connected with social behaviour patterns that lead to a clustered distribution of data in popular locations (both highly populated and attractive for social gathering, such as tourist attractions).

Following the rise of Web 2.0, scholars, researchers, and entrepreneurs in the GI domain have high hopes about the importance of user participation and UGSC.

In terms of spatial distribution, though the spatially explicit source of Geograph had the third largest data pool among the four sources, it provides considerably better coverage of the study area. In contrast, the spatially implicit sources of Flickr and Picasa Web fail to adequately cover the study area, except for urban areas and tourist attractions. Even in these areas, the expectation surfaces showed that suburbs are not well represented as compared with the underlying population. Interestingly though, at the detailed level of the distribution of photographs in popular tiles, the differences between implicit and explicit sources diminish. Finally, in contrast with our initial classification, Panoramio behaves as a spatially implicit source even if the aim is to submit photos of places. Apparently, the motivation to publish photos of users' favourite places is not adequate enough to provide extensive spatial coverage for Great Britain.

In terms of data flow, it is evident that the popularity of sources such as Flickr and Picasa Web have grabbed the users' interest, resulting in huge amounts of information being submitted. Taking into account distribution and data flow, we can suggest that explicit sources have the means to distribute data collection at a national level but they lack the power generated by the participation in implicit ones. This is an important differentiation—implicit sources are popular because the geographical functionality is additional to a popular functionality (photo storage and sharing) unlike explicit sources, where the user is primarily interested in the geographical functionality.

The difference between spatially implicit and explicit sources is intensified at the larger scale of the 15 test areas. Our findings show that, even in areas where Flickr appears to have numerous photos, the distribution is still confined to a very few popular spots. By contrast, Geograph's photo distribution is more scattered, covering the area more adequately, even with considerably fewer photos. Interestingly, apart from these differences, it is shown that there is also a common aspect. As Figures 9 and 10 show, by examining non-GI related issues, users of both web applications behave in a similar way. Thus, the crucial issue that makes the web applications differ is, how related to space are the incentives provided by the web application to the users to guide user participation in a spatially-oriented direction? In other words, the GI community needs spatially explicit applications with UGSC that can provide data at a national level that is useful for mapping applications.

Finally, examining the subject from a broader point of view, one of the pillars of the Web 2.0 evolution is the Long Tail [*O'Reilly* 2005; *Anderson* 2004; 2006]. The architecture of web applications like Flickr is heavily based on that principle. The ability of any user to add any content about any subject, popular or not, is the cornerstone that enables the support of users' interest for small niches. That cumulative interest transforms these niches into important elements of the application. Our research shows that in spatial terms, the Long Tail principle is not realised in the spatial distribution of the photos. Spatially speaking, the users are not interested in the small, relatively unpopular, niches of space but focus on the mainstream places. This is in contrast with early observations about UGSC regarding the ability of the phenomenon to record local places and activities that would be otherwise uncovered [*Goodchild* 2007a].

6. Conclusions and Future Work

Following the rise of Web 2.0, scholars, researchers, and entrepreneurs in the GI domain have high hopes about the importance of user participation and UGSC. This is mainly because the evolution of Web 2.0 revealed a new form of collaboration through numerous users who individually upload freely and without constraits, any kind of

explicit or implicit spatial content on the Web, trying to describe, in detail, their neighbourhoods, home towns or vacation places. *Hanke* [2007], the co-founder of Keyhole, suggested during the O'Reilly Where 2.0 conference in 2007, that this is an opportunity for all of us to build "*a map of the world that I think will be more detailed, more comprehensive, more inclusive than any map of the world that has ever been created*." Hanke did not refer simply to satellite imagery, but to "*a map of user annotations, of descriptions, of images, of movies, of sound*." This approach is indicative of the enthusiasm that stems from UGSC.

In our research we challenged the validity of such approaches and examined whether this enthusiasm is justifiable. Although we are aware that there are strong examples of crowd-sourced mapping efforts like OpenStreetMap (based on vector editing), by examining the spatial aspects of popular photo-sharing websites, our findings both justify a moderate optimism and raise some concerns about the overall direction of the phenomenon. We conclude that, in order for a web application to be able to serve as a universal source of spatial content, it must be a spatially explicit one. User participation and contribution must directly interact with spatial entities in a conscious effort to better describe our world. Our findings show that common social networking applications cannot provide the breadth of information needed. It also has been shown clearly that not all web sources, even if they hold some spatial information, can provide GI suitable to support mapping efforts or a broad GIR. Socially oriented web applications can be proved to be an immense pool of spatial information but we have shown that (at least up to now) their scope of usage is limited to urban areas and tourist attractions and consequently, GIR methods that solely use such sources will suffer from that limitation. Nevertheless, there are already early efforts to introduce geo-tagged images into mainstream and commercial GISs [*Woolford* 2008]. Moreover, although still at a research level, there is growing interest from National mapping agencies regarding the potential of the phenomenon (see, for example, the EuroSDR workshops on crowd sourcing for updating national databases).

An interesting question emerges from this work, given the fact that mapping agencies are struggling to find the changes that constantly take place on the ground in order to keep their databases up-to-date, 'what is the spatial pattern of those changes?' If that pattern is similar to the spatial distribution we recorded for the data from spatially implicit sources, then obviously their role and importance will be considerably enhanced. Otherwise, our efforts will be focused solely on spatially explicit ones. Finally, a similar approach can be used to evaluate the contribution of vector-sharing web sources such as OpenStreetMap.

References

Ames, M. and M. Naaman. 2007. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. San Jose, California, USA: ACM, 971-980. Available at: *http://portal.acm.org/citation.cfm?id=1240624.1240772#* [Accessed April 29, 2009].

Anderson, C. 2004. *The Long Tail*. Wired. Issue 12.10, October 12, 2004.

Anderson, C. 2006. *The Long Tail.* Random House Business Books, Croydon, 244 pages.

Arampatzis, A., M. Kreveld, I. Reinbacher, B.C. Jones, S. Vaid, P. Clough, H. Joho, M. Sanderson, M. Benkert, and A. Wolff. 2006. Web-based delineation of imprecise regions. *Computers, Environment and Urban Systems*, 30(4), 436-459.

Bishr, M. and L. Mantelas. 2008. A trust and reputation model for filtering and classification of knowledge about urban growth. *GeoJournal,* 72:133–135.

Budhathoki, N., B. Bruce, and Z. Nedović-Budić. 2008. Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*, 72(3), 149-160.

Cox, A.M., 2008. Flickr: a case study of Web 2.0. *Aslib Proceedings*, 60(5), 493-516.

Dunn, C.E. 2007. Participatory GIS a people's GIS? *Progress in Human Geography*, 31(5), 616-637.

Dykes J., S.R. Purves, J.A. Edwardes and J. Wood. 2008. Exploring volunteered geographic information to describe place: visualization of the 'Geograph British Isles' collection. In *Proceedings of GIS Research UK 16th Annual Conference*. Manchester, Great Britain: 256-267.

Dykes J. and J. Wood. 2008. Mashup Visualization with Google Earth and GIS. Available at: *http://www.gicentre.org/infovis/* [Accessed March 15, 2009].

Elwood, S. 2008a. Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal,* 72:133–135.

Elwood, S. 2008b. Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal,* 72:173–183.

Elwood, S. 2008c. Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*, 22(1), 71.

Estes, J.E. and W. Mooneyhan. 1994. Of maps and myths. *Photogrammetric Engineering and Remote Sensing,* 60 (5), p. 517-524.

Goldberg, D.W., J.P. Wilson, and C.A. Knoblock. 2009. Extracting geographic features from the Internet to automatically build detailed regional gazetteers.

International Journal of Geographical Information Science, 23(1), 93.

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

Goodchild, M.F. 2007b. Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial data Infrastructure Research*, Vol. 2, 23-32.

Goodchild, M.F. 2008a. Assertion and authority: the science of user-generated geographic content. Available at: *http://www.geog.ucsb.edu/~good/papers/454.pdf* [Accessed November 17, 2008].

Goodchild, M.F. 2008b. Commentary: Wither VGI? *GeoJournal* (2008) 72:239–244.

Haklay, M. and C. Tobon. 2003. Usability evaluation and PPGIS: towards a user-centred design approach. *International Journal of Geographical Information Science*, 592, 577.

Haklay, M., A. Singleton, C. Parker. 2008. Web mapping 2.0: the neogeography of the GeoWeb. *Geography Compass,* Vol 2.

Hanke, J. 2007. The evolution of the GeoWeb. [Online] Where 2.0 conference video clip. Available at: *http://conferences.oreillynet.com/cs/where2007/view/e_sess/12583*. [Accessed May 05, 2008].

Jones, C.B., R.S. Purves, D.P. Clough, and H. Joho. 2008. Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science,* 22(10), 1045.

Jones, C.B. and R.S. Purves. 2008. Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3), 219-228.

Liu S.B., L. Palen, J. Sutton, L.A. Hughes, and S. Vieweg. 2008. In search of the bigger picture: the emergent role of on-line photo sharing in times of disaster. In *Proceedings of the 5th International ISCRAM Conference.* Washington, DC, USA, May 2008.

MacEachren A.M. and J. Kraak. 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science,* 28 (1): 3-12.

Miller, C.M. 2006. A beast in the field: the Google Maps mashup as GIS/2. *Cartographica*, 41 (3), 187–199.

Negoescu, R. and D. Perez. 2008. Analyzing Flickr groups. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*. Niagara Falls, Canada: ACM, 426, 417. Available at: *http://dx.doi.org/10.1145/1386352.1386406* [Accessed April 29, 2009].

O'Reilly, T. 2005. *What is Web 2.0: design patterns and business models for the next generation of software*. Available at: *http://www.oreillynet.com/lpt/a/6228* [Accessed May 2, 2008].

Obermeyer, N. 2007. Thoughts on volunteered (geo) slavery. [Online] Workshop on Volunteered Geographic Information. Available at: *http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Obermeyer_Paper.pdf* [Accessed January 3, 2010].

Popescu, A., G. Grefenstette and P.A. Moëllic. 2008. Gazetiki: automatic creation of a geographical gazetteer. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*. Pittsburgh PA, PA, USA: ACM, 85-93. Available at: *http://portal.acm.org/citation.cfm?id=1378889.1378906* [Accessed April 20, 2009].

Purves S.R. and J.A. Edwardes. 2008. Exploiting volunteered geographic information to describe place. In *Proceedings of GIS Research UK 16th Annual Conference*. Manchester, Great Britain: 252-255.

Schockaert, S., D.P. Smart, I.A. Abdelmoty, B.C. Jones. 2008. Mining topological relations from the Web. In *Database and Expert Systems Application, 2008. DEXA '08. 19th International Conference on*. 652-656.

Sieber, R. 2007. Geoweb for social change. [Online] Workshop on Volunteered Geographic Information. Available at: *http://www.ncgia.ucsb.edu/projects/vgi/docs/supp_docs/Sieber_paper.pdf* [Accessed January 3, 2010].

Turner, A.J. 2006. *Introduction to Neogeography.* Sebastopol, CA: O'Reilly Media Inc.

Twaroch, F.A., C.B. Jones, and A.I. Abdelmoty. 2008. Acquisition of a vernacular gazetteer from web sources. In *Proceedings of the First International Workshop on Location and the Web*. Beijing, China: ACM, 61-64. Available at: *http://portal.acm.org/citation.cfm?doid=1367798.1367808* [Accessed April 30, 2009].

Woolford, T. 2008. The Use of Geotagged Images in GIS. [Online] AGI Geocommunity 2008. Available at: *http://www.agi.org.uk/SITE/UPLOAD/DOCUMENT/Events/AGI2008/Papers/TimWoolford.pdf*. [Accessed 03 January 2010].

Authors

Vyron Antoniou is a Captain in the Greek Army and since 1998 has served at Hellenic Military Geographical Service. Currently he is a Ph.D. student at UCL in the department of Civil Environmental and Geomatic Engineering. His research interests are in user-generated content and National Mapping Agencies, spatial databases, vector data transmission over the Web, and web mapping.

Mordechai (Muki) Haklay is a senior lecturer in Geographical Information Science in the department of Civil, Environmental and Geomatic Engineering at UCL, where he is also the Director of the Chorley Institute, which focuses on facilitating interdisciplinary geospatial research at UCL. His research interests are in public access to environmental information, Human-Computer Interaction (HCI) and Usability Engineering for GIS, and Societal aspects of GIS use. He received his Ph.D. in Geography from UCL.

Jeremy Morley is Deputy Director of the Centre for Geospatial Science at the University of Nottingham and was previously a lecturer in Geographic Information Systems at University College, London. His interests lie in GIS interoperability, interfaces between GIS web services and mashup technology, sensor webs, and planetary mapping. ❑