Critical review

Placing of photos on the internet: Critical analysis of biases on the depictions of France and Afghanistan on FLICKR

Christoph Lambio*, Tobia Lakes

Humboldt-University, Unter den Linden 6, 10099 Berlin, Germany

ABSTRACT

The Internet and the representation of space therein are almost omnipresent in society and everyday life. Peer-produced geographic data is gaining a particular importance through increasingly available digital tools and techniques that shape the perception of space in the internet, such as flickr, OpenStreetMap or Wikipedia. However, few studies focused on how space is represented, and by whom it is described. We hypothesize that the alleged opening up of geographic information and the assumed benefits for every individual and society through the occurrence of 'easy-to-use-mapping-tools' was premature. To explore these assumptions, a comparative study of the flickr worldmap was undertaken and roughly 6.8 million metadatasets of geocoded photos in France, and roughly 50,000 metadatasets in Afghanistan were downloaded and the metadata was analyzed. Our results indicate that photos geocoded in France show a large diversity of motives, while photos geocoded in Afghanistan are mostly limited to content containing warfare when they are up loaded in English. The content of the photo and therefore the representation of space strongly depend on who uploaded the photo, particularly in Afghanistan. We can show that the representation of space on the internet, for the case of flickr, is strongly dominated by perceptions of Western societies and individuals. We therefore confirm our hypothesis that the supposed opening up of geographic information systems through 'easy-to-use-mapping-tools' and their democratization thereof was premature. Moreover, we highlight the importance of understanding who contributes online content to be able to evaluate peer-produced data, its value, and its possible applications to avoid reproducing biases.

1. Introduction

The rapid development of internet technologies and the popularity of social media resulted in an evolution of the production of decentralized and volunteered data, which also applies to the production of geographic data (Graham, 2010). Thanks to the simplification of the (geographic) data production, everyone with a computer and internet access has the possibility to produce geographic content and become part of the 'geoweb' (Crampton, 2008:2f). Turner's (2006) conceptualization of 'neogeography' describes a transformation in the traditional mapping practice, namely the shift from officials and professionals providing geographic data to non-trained individuals, creating geographic content (i.e., Goodchild, 2007; Elwood, 2008). This change in the production of geographic information led to assumptions of the democratization of GIS and map-making (Haklay, 2013).

Geographers have known for a very long time that maps have a particular power. Shedding a critical cartographic light on peer produced maps will answer how material places are represented online,

which plays an increasingly important role in how we learn about the world (Zook et al., 2011). It is not just the sheer amount of information, but rather what kind of information one finds online that brings a place into being (Graham, 2010). Online information can be accessed 'on site' with a smartphone in real-time; thus, an additional layer of information, a "digital sixth sense" influences our perception of space (Zook and Graham, 2011). It is a blend of online and offline information, a new constructed reality (ibd.) that shapes and reproduces cultural, social, and political factors (Crutcher and Zook, 2009; Graham et al., 2013). As Zook and Graham (2011) formulated, (online) maps are "distorted mirrors" (Zook and Graham, 2011:130), which are "grounded in people's interpretations of place" (Graham, 2010:433). Therefore, how place is represented online is of specific relevance, and it is even more surprising that only a few empirical studies have focused on the analysis of this phenomenon up to now. The aim of this paper is to address this lack in research and discuss initial insights as to how two countries – France and Afghanistan – are perceived, mapped and represented online.

Table 1
The deployed ranking after the process of stemming for English tags in Afghanistan.

The 55 English tags most used for photos in Afghanistan

No	Tag	Amount		No	Tag	Amount		No	Tag	Amount		No	Tag	Amount	
		Absolute	Relative			Absolute	Relative			Absolute	Relative			Absolute	Relative
1.	Afghanistan	11.637	27.30%	15.	SSG [anonymized]	555	1.30%	29.	Tajikistan	399	0.94%	43.	Kunar Province	312	0.73%
2.	Kabul	3.370	7.91%	16.	1158th Infantry	553	1.30%	30.	Wakhan	398	0.93%	44.	Women	297	0.70%
3.	Army	2.157	5.06%	17.	Pamir	547	1.28%	31.	Combat	398	0.93%	45.	Islam	293	0.69%
4.	War	1.482	3.48%	18.	Travel	547	1.28%	32.	People	398	0.93%	46.	Balazs Gardi	293	0.69%
5.	Military	1.223	2.87%	19.	Laghman Province	533	1.25%	33.	Landscape	388	0.91%	47.	2011	291	0.68%
6.	Afghan	1.181	2.77%	20.	Taskforce Diamondback	521	1.22%	34.	Helmand	374	0.88%	48.	Wildkat Photography	270	0.63%
7.	Afg	931	2.18%	21.	TF Diamondback	521	1.22%	35.	Marine	370	0.87%	49.	Develop	269	0.63%
8.	OEF	865	2.03%	22.	Kandahar	505	1.18%	36.	Mehtar Lam	354	0.83%	50.	Geotag	259	0.61%
9.	Mountain	861	2.02%	23.	NATO	474	1.11%	37.	Bamiyan	350	0.82%	51.	Agriculture	253	0.59%
10.	Asia	773	1.81%	24.	Deploy	454	1.06%	38.	AF	349	0.82%	52.	World	249	0.58%
11.	Soldier	727	1.71%	25.	Jalalabad	443	1.04%	39.	Taliban	348	0.82%	53.	School	248	0.58%
12.	ISAF	674	1.58%	26.	Herat	434	1.02%	40.	Afghan National Army	346	0.81%	54.	US Army	248	0.58%
13.	Operation Enduring Freedom	577	1.35%	27.	Children	432	1.01%	41.	Conflict	335	0.79%	55.	Helicopter	245	0.57%
14.	Bushmaster	576	1.35%	28.	ANA	409	0.96%	42.	Central Asia	334	0.78%				

Sum of all 55 tags: 42,630
Total photos in English: 15,894
Total photos tagged in English: 13,882
Photos in English but not tagged: 2012
Number of different tagss: 12,194

2. Representation of France and Afghanistan on flickr: A comparative approach

We choose two countries that differ in geography, culture and history; namely, France and Afghanistan because we want to highlight the possible differences that occur in the creation of cyberspaces and the depiction of space. The data was downloaded from flickr through the flickr API using the flickr.photo.search method.[1] Here, we used a geo query request, which returns 4000 photos per bounding box.[2] Since the number of available photos was significantly higher for both countries, the bounding boxes were split into smaller bounding boxes (France: n = 162,947, mean size = 3.38 km^2; Afghanistan: n = 158,042, mean size = 4.07 km^2). We downloaded all geotagged photos for the two countries with the upload date equal or before the 12th March 2015 and the respective latitude and longitude, tags, the number of views, the description, the date when the photo was uploaded and when it was taken.

By counting the most frequently used tags, we can establish the most widely photographed subjects in the respective country. To first identify the language(s) we used two language libraries (LID) with a Python binding: the LID used in the Google browser Chromium[3] (cld), and the langid[4] library. To increase the accuracy of the LIDs and to reduce the required data to a sample for which the language is most likely faultlessly identified, a strict logic for language detection was developed. To count the most used tags and establish a ranking of what

is photographed the most, we then reduce the words to their linguistic stem by using the nltk-[5] and the hazm[6] library.

3. A biased representation of place in the flickr map?

Our findings highlight that the flickr cyberspace and map around France and Afghanistan significantly differs. Of the 12 million available datasets on flickr for France, 6.8 million were downloaded and of those 1.7 million had at least one tag and could be clearly identified by language. In contrast, only 53,000 datasets were online for Afghanistan and of those 51,700 were downloaded and of those 17,700 had at least one tag and could be clearly identified with language. In comparison to France, Afghanistan obviously resides in a "black hole" on the flickr world map (Graham, 2010: 429).

By comparing the English tags one can easily see that the flickr cyberspace around Afghanistan is rather hostile, while the flickr cyberspace created around France is friendly and welcoming. For Afghanistan, terms like 'Army', 'War', 'Military', 'OEF', 'Soldier', 'ISAF', 'Operation Enduring Freedom', 'Bushmaster', 'SSG [anonymized]', '11558th Infantry', etc. dominate in the analyzed tags (see Table 1). From the 13,822 photos that were English tagged, about 18% contain content of warfare, with the war connoted words resulting in 2540 photos. In contrast for France we see a positive and colorful image of the country picture emerging with combinations of words, such as 'color', 'blue', 'light', 'beach', 'snow', and 'holidays' (see Table 2). Most of the tags refer to traveling, nature, and landscape, or describe particular scenery in detail or they include words like 'car', 'architecture', 'art', 'festival' or 'music' connected to cultural events or achievements in construction and engineering.

How and by whom is this flickr cyberspace of war, fear and terror

---

[1] For a more detailed description of the script and the processing schema, the readers may contact the authors.

[2] See https://www.flickr.com/services/api/flickr.photos.search.html for the limitations of the API.

[3] See https://pypi.python.org/pypi/chromium_compact_language_detector/0.2 for more information.

[4] See https://pypi.python.org/pypi/langid/1.0dev for more information. A detailed explanation of the langid library can be read here: Lui and Baldwin (2011) and/or Lui and Baldwin (2012).

[5] http://www.nltk.org/api/nltk.stem.html; – In this study the 'SnowballStemmer' is used.

[6] https://github.com/sobhe/hazm.