Revista Brasileira de Cartografia (2017), Nº 69/8, Edição Especial "Geovisualização, mídias sociais e participação cidadã: apoio à representação, análise e gestão da paisagem": 1586-1597.
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

BIG-DATA TO SUPPORT THE STUDY OF THE ENVIRONMENTAL DISASTER IN MARIANA, MG: FROM VGI TO SOCIAL MEDIA GEOGRAPHIC INFORMATION

Big Data como Suporte ao Estudo do Desastre Ambiental em Mariana, MG: Do VGI ao Social Media Geographic Information

Júnia Lúcio de Castro Borges[1], Ana Clara Mourão Moura[1,2], Priscila Lisboa de Paula[2] & Pedro Benedito Casagrande[3]

[1] Universidade Federal de Minas Gerais - UFMG
Escola de Arquitetura, Laboratório de Geoprocessamento / Programa de Pós-Graduação em Arquitetura e Urbanismo
Rua Paraiba 697, sala 410A, Belo Horizonte – MG, cep 30130-140
juniaborges@yahoo.com.br

[2] Universidade Federal de Minas Gerais - UFMG
Escola de Arquitetura, Graduação em Arquitetura e Urbanismo
Rua Paraiba 697, sala 410A, Belo Horizonte – MG, cep 30130-140
priscilalpaula@gmail.com

[3] Universidade Federal de Minas Gerais - UFMG
Instituto de Geociências / Programa de Pós-Graduação em Geografia
Av. Antonio Carlos 6627, Belo Horizonte – MG, cep 31270-901
pedrobcasagrande@gmail.com

Received on January 26, 2017/ Accepted on July 14, 2017
Recebido em 26 de Janeiro, 2017/ Aceito em 14 de Julho, 2017

ABSTRACT

The use of Social Media Geographic Information in a planning process would improve our knowledge about local values, urban and landscape development and could support decision-making. This paper presents a case study of an environmental disaster that happened recently in one of the most important social economic areas in Brazil to understand how this type of information could be used as a systematized planning input. The authors seek to understand if it is possible to use VGI and SMGI (Volunteered Geographic Information and Social Media Geographic Information) to capture social values, such as genius loci(the essence of the place)and expectations, the values that should be considered in a disaster recovery plan. To do that, we tested an active VGI platform and passive SMGI posts' analysis. Our best results so far relied on a proposal for image analysis from Instagram posts to separate them in the themes of "everyday life", "landscape" and "memes". Presenting a logic and a tool to do this classification automatically, we provide a first step to be used in big-data, as they are characterized by a big amount of data, and we can develop a first analysis in the main pictures posted by people from that area. It is a contribution to use social media to measure values and collective expectations from a place.

Keywords: Big-Data, Volunteered Geographic Information, Social Media Geographic Information.

RESUMO

O uso da Informação Geográfica por Mídia Social em um processo de planejamento apresenta potencial para ampliação de conhecimentos sobre valores locais, desenvolvimento urbano e paisagístico, e poderia apoiar a tomada de decisões. Este artigo apresenta o caso de estudo de um desastre ambiental que aconteceu recentemente em uma das mais importantes áreas econômicas do Brasil para entender como esse tipo de informação poderia ser usada como base um planejamento sistematizado. Os autores procuram entender se é possível usar VGI e SMGI para capturar valores sociais, tais como genius loci e expectativas, os valores que devem ser considerados em um plano de recuperação de desastres. Para fazer isso, foi testada uma plataforma VGI ativa e realizada a análise de mídia passiva, das postagens em um SMGI. Os melhores resultados até agora se basearam em uma proposta de análise de imagem das postagens em Instagram, separados segundo os temas de “vida cotidiana”, “paisagem” e “memes”. Apresentando uma lógica e uma ferramenta para fazer esta classificação automaticamente, expõe-se um primeiro passo para ser usado em big-data, pois são caracterizados por uma grande quantidade de dados, e desenvolvimento da primeira análise nas principais imagens postadas por pessoas do lugar. É uma contribuição para uso das mídias sociais com foco na verificação de valores e expectativas coletivas de um lugar.

Palavras-chave: Big-Data, Informação Geográfica Voluntariada, Informação Geográfica por Mídia Social.

1. INTRODUCTION

The use of Social Media Geographic Information in a planning process would improve our knowledge about local values, urban and landscape development and could support decision-making. This paper presents a case study of an environmental disaster that happened recently in one of the most important social economic areas in Brazil to understand how this type of information could be used as a systematized planning input. The authors seek to understand if it is possible to use VGI and SMGI to capture social values, such as *genius loci* (the essence of the place) and expectations, the values that should be considered in a disaster recovery plan. To do that, we tested an active VGI platform and passive SMGI posts’ analysis. Our best results so far relied on a proposal for image analysis from Instagram posts to separate them in the themes of “everyday life”, “landscape” and “memes”. Presenting a logic and a tool to do this classification automatically, we provide a first step to be used in big-data, as they are characterized by a big amount of data, and we can develop first analysis in the main pictures posted by people from that area. It is a contribution to use social media to measure values and collective expectations from a place.

Analytics gathered from Social Media Geographic Information (SMGI) improve comprehension of community values (BORGES *et al*., 2015). There are several information attributes made available by ordinary citizens within social network environments such as location, texts, videos, preferences, images and audio. The development of a systematic methodology to analyse information posted by ordinary people would contribute to landscape and urban planning processes. This paper presents a new methodological direction to investigate image collection from Social Media.

The term Volunteered Geographic Information (VGI) was proposed by Goodchild (2007) in the sense that citizens act as sensors. Its development leads to innovative models using social networks and is being denominated Social Media Geographic Information. Campagna (2016) understands it as a deviation from a traditional vision of VGI and states that it may be used for both leisure and professional reasons. Ideally, it would allow integration and sharing of the resulting information flow.

Craglia (2007) understands the production of geographic information from social media as a more comprehensive phenomenon from user-generated content, born with web2.0, based on the logics of wikis and blogs. Bao *et al.* (2015) call the system *location-based social networks* (LBSN).

Davis Jr. *et al.* (2016) presents a classification according to the way data is captured and the manner of user collaboration: active or passive. If data is captured by sensors embedded in their mobile devices(for example), it is classified as crowdsensing. This is the case of doing check-in in a place or using the GPS to control the traffic (*Waze*). If the data is registered without a sensor, for example by a computer based VGI or a LBSN platform, the process is crowdsourcing. In both cases the capture can be passive (the users make posts that are captured by Apps and are used in a spatial analysis to support strategies from industries, commerce, or citizens behaviors' analysis, like Tweets or Instagram) or active (the users consciously produce information they know is going to be used, as Open Street Map or a *Ushahidi* Project, in which the users must go to the web to register a comment that is associated to a location). Beyond crowdsensing and crowdsourcing, a third classification defended by Davis Jr. *et al*. (2016) is called by authors Mission Oriented and it is a process where users are selected by similarity of their profile to the task and therefore asked to contribute.

Most of the studies in social media were based on the interest of capturing data from mobile phones to understanding urban movements over time (AHAS & MARK 2005). This capture was classified as a passive and anonymous capture of citizens' data, that results in "sensors of a network" (PUCCI *et al*., 2015). The main contribution is the capture and mapping of temporary practices in the territory, which opens conditions to understand behaviors and values (PUCCI *et al*., 2015 and AHAS *et al*., 2010). This possibility of social-positioning provides data about who is moving, where and how (AHAS & MARK, 2005).

As a diffused media in the territory, Cheng and Wicks (2014) states it has become a rich source for detecting, monitoring, analyzing stories, especially in disaster events. The authors present case studies, in which social media was used to register information about disasters, demonstrating how the methodology of spatial models are used to construct clusters. According to them, tweet clusters emerge during space-time relevant events, because people tend to use more social media in those situations. The authors map and study the movements of clusters and call these actions citizen journalism, as the users describe and provide information about the world around them. They mention case studies using Twitter to register earthquakes, fires and floods (EARLE *et al*., 2011; STARBIRD & PALEN, 2010; BRUNS *et al*., 2012; apud CHENG & WICKS, 2014). In all these situations, the methodology was based on the use of hashtags, of the words related to the disaster, like "earthquake" and so one. They observed that more than 75% of the disasters were detected by Twitter, what means it can be used as social alerts in UK, US and Australia.

Twitter is not very used in Brazil. But even in countries where it is more used, we must observe that data produced by social media is not distributed demographically by age and socioeconomic conditions. Cheng and Wicks (2014) present numbers for Twitter, but we can think that it is the same conditions in all others social media: 6% of the users are under 22 years old, while 60% are under 35 years old. In case studies about social values and behaviors, this unequal distribution of social media users in society could result in difficulties in the use of the data. However, as the authors stated, a disastrous event creates a significant emotional impact on all people, and not only in some social groups. It means that the data provided is reliable, regardless of the demographic conditions.

The analysis of photos posted by users on social media is also a field of interest to produce social data. Jankowski *et al*. (2010) analyzed movement patterns from photos posted on the Flickr website. Their interest was to capture the photographers' movements, to understand preferences for urban landmarks and pertinent travel itineraries. They used data for location and key-words, but the images themselves were not analyzed.

The case study presented here was applied in the environmental disaster in Mariana, Minas Gerais, Brazil. On November 5, 2015, two dams collapsed in the area, sending a torrent of mining sludges through the village of Bento Rodrigues. The muddy floodwaters from an iron ore mining operation destroyed hundreds of homes, killed some residents, and left others missing. 62 million cubic meters of wastewater were unleashed. Multiple rivers were polluted with wastewater and mud. In Barra Longa—a village about 80 kilometers (50 miles) from the dams—the river surged as much as 15 meters and flooded homes. As health officials conducted tests, cities as far as 300 kilometers (200 miles) downstream lost access to drinking water. The mud and polluting materials made their way to the Atlantic Ocean killing fish along the river, Rio Doce (REUTERS, 2015). It is considered the largest environmental disaster in Brazil.(Figure 1 and Figure 2).

Fig 1. Mariana's disaster – The rupture of the dams. Source: Screenshot from PigMine 7, Youtube

Fig 2 – The area before and after the disaster. Source: NASA Earth Observatory images by Joshua Stevens, using Landsat data from the U.S. Geological Survey. Caption by Kathryn Hansen. Instrument(s):Landsat 8 – OLI.

After the environmental disaster happened on November 5, 2015, a case study using VGI and SMGI, from active and passive perspectives, was developed. In active studies, a project was created in the application, *Crowdmap*, to receive comments and posts. In the strategy of investigating passive SMGI the studies were based on the analysis of Instagram images.

The purpose of the research was to contribute towards the identification of collective values, aiding local image reconstruction as well as subsidizing a landscape recovery plan in a new hazardous area. In spatial planning, it is highly recommended that citizens be heard. Acknowledgment and comprehension are the most important value variables in landscape and urban planning.

The hypothesis is to determine whether it is possible or not to establish a connection between the SMGI posts and the level of affection shared. We tested a VGI system, creating a project in the *Crowdmap* platform, with the expectations of receiving posts and comments from society, classified by the main values in the specific situation, but the results could only be observed with the use of a SMGI analysis. As Borges e*t al*. (2016) had already stated, "the potential of SMGI is amplified by the possibility of reaching a larger share of the population, whereas from a VGI project, the type of analysis should attain a higher complexity and focus on the question asked but with a more focused participation".

2. METHODOLOGY

The methodology was based on the production of cartographic data, construction of a VGI application and the use of SMGI data.

2.1 Cartographic data

The first step in the studies was the creation of maps from the territory, organizing topographic information and data about rivers, main localities and spatial references. The goal was to locate the starting point of the disaster and to create information about the buffer zone and to understand administrative data (the limits about Minas Gerais and Espírito Santo states) (Figure 3).

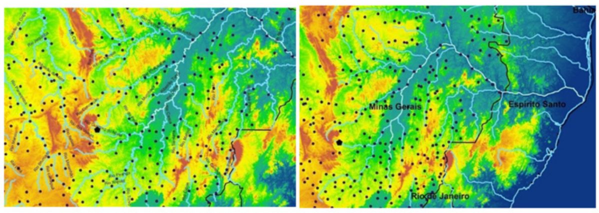


Fig. 3 – The image of the territory. Source: Produced by the authors, using SRTM and IBGE data.

2.2 VGI application

The second step was the creation of a VGI platform based on the *Crowdmap application*. We created a logo with the goal to associate the *genius loci* of the place (the mountains from Minas Gerais and the river) for the project, expecting to use visual values and a play on words to arouse interest among users. (Figure 4).

Fig. 4 – Doce Rio Doce (Sweet river sweet) - Logo created for the project in *Crowdmap*.

The platform (Figure 5) was created to receive posts and comments. The objective was to map the *genius loci* of the place, "sweet memories”. We asked people to register: Special Landscapes, Daily life, Local culture and history, Moments from personal history. They could mark a point and associate comments, pictures, videos (Figure 6). They could also put a buffer zone on the image of the river. The application could be used on a desktop or even on a mobile. The user could register to receive alerts if any other posts were registered in a place of his interest, defining the distance from a point to capture these alerts and send him the information by email (Figure 7).

Fig. 5 – Home interface - https://docerio.crowdmap.com Source: The authors.

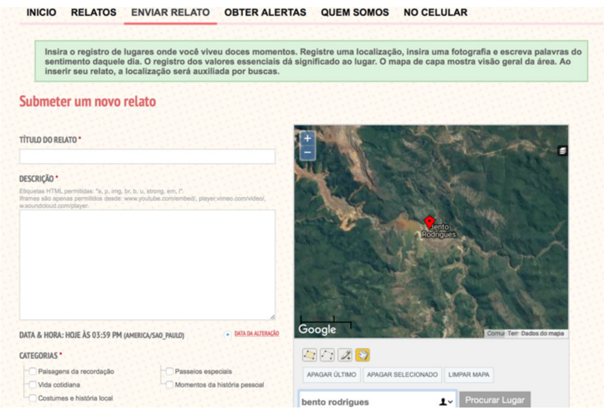


Fig. 6 – To send a post – https://docerio.crowdmap.com Source: The authors.

Fig. 7 – Alerts interface - https://docerio.crowdmap.com Source: The authors.

All users in a VGI application could not only post their registers, but also see the other users' posts, per theme or all the posts. It was also possible to register “like” or unlike”, in a very sharing system. We expected to receive many posts, and we were very surprised with the lack of interest from society. The use of the system is very easy, but we must recognize that people are not very used to see the territory from above and to register information on a map. Moreover, we read from a bibliography (CAMPAGNA *et al*., 2013; 2014) that the VGI is a very difficult application to be maintained. The ones that were successful were those associated to big projects. And even in those, people tend to go once to the website, but if they don't get any return or result they don't use it a second time.

To face general problems of irregularity in data production by social media, some authors are proposing investments in “recommendations” (BAO, 2015; CHEN *et al*., 2014; MATEVELLI *et al*., 2015, DAVIS JR. *et al*., 2016). Davis Jr *et al*. (2016) believe that there are several problems on data collection and their use in crowdsourcing initiatives. They highlight four main problems: (i) gradual reduction users' interest; (ii) irregular coverage on theme or interest space; (iii) doubts about data reliability and strategies validation; and (iv) how to provide feedback of contributions. (Figure 8).

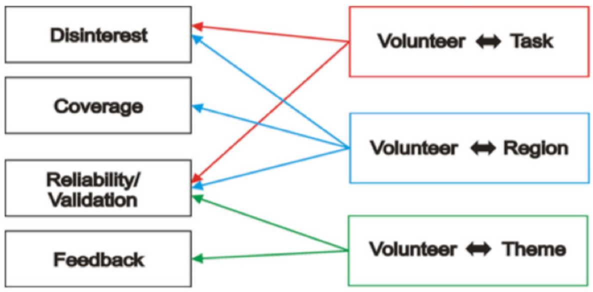


Fig. 8 - Problems and Challenges. Source: Matevelli *et al*., 2015; Davis Jr. *et al*. 2016.

2.3 SMGI application

As the “active” VGI studies didn't have a good result, we decided to invest in a “passive” SMGI application. An analysis of Instagram messages was performed using *LETICIA API*, a courtesy of LABIC (Laboratório de Imagem e Cybercultura - Laboratory of Image and Cyberculture of Federal University of Espirito Santo, Brazil). Messages from 22 days since the day/time of the incident were downloaded aiming at analysing values, behaviours and profiles of the posts. The expectation was to identify the *genius loci* (the essence of the place and its symbols) through the collective description and investigation of images, videos, text and geolocation of posts.

According to Borges, Jankowski and Davis Junior (2015a) the hashtag (#) is the main way to attach a category or a subject to a message in a social media environment. The authors point out that a semantic analysis should follow the potential words used by the group of people that is the research agenda's focus. This could also help by indicating the profile of people related to the certain topic.

Hashtags were selected in two ways: firstly, by researchers' intuition #mariana #riodoce and secondly using a platform called *Tagboard* (available at tagboard.com), a real-time search engine that favours the investigation of a hashtag, and a third tag was found: #sosriodoce. After a locational analysis of posts, from the three tags, the last was defined as most appropriate as they have a better rate of onsite posts. The download was done from November 5, 2015 to November 27, 2015.

From the posts, we could get information about the user ID (which allows to compare the posts sent by the same user, and construct studies about movements over time (AHAS & MARK, 2005), we could get the link to the picture posted and how many "likes" he received to that post (Figure 9). It was also possible to get the profile of the user, the time it was posted and its location. It's important to observe that many posts come without the location, which means that the user blocked the capture of location on his app (Figure 10). In this example, a media of 20-21% of the posts could be geo- referenced, and they were used in our case study. Further studies could be developed with the totality of data, analyzing aspects not related to spatial location. For this study, the geographic aspect was very important, which justifies only the use of data that presented information about latitude and longitude.

id	url	user	like	link
1127044502427581683_2138869142	https://scontent.cdninstagram.com/hphotos-xft1/t51.2885-15/s320x320/e35/12237589_1026965414000561_1506173115_n.jpg	vitorholetz	2	https://instagram.com/p/-kEQDtQLzz/
1127043031493807805_1063364386	https://scontent.cdninstagram.com/hphotos-xap1/t51.2885-15/s320x320/e35/12237110_985042814870737_1909716437_n.jpg	rosaspindola2	3	https://instagram.com/p/-kD6pyxl69/
1127042017034276492_1063364386	https://scontent.cdninstagram.com/hphotos-xaf1/t51.2885-15/s320x320/e35/12224232_1668552743359137_562368504_n.jpg	rosaspindola2	3	https://instagram.com/p/-kDr5ARI6M/
1127040251106463319_1063364386	https://scontent.cdninstagram.com/hphotos-xtp1/t51.2885-15/s320x320/e35/12269831_760426564063409_1792294123_n.jpg	rosaspindola2	5	https://instagram.com/p/-kDSMWxI5X/
1127039523445050938_1063364386	https://scontent.cdninstagram.com/hphotos-xft1/t51.2885-15/s320x320/e35/12237439_625261314281171_1809228275_n.jpg	rosaspindola2	2	https://instagram.com/p/-kDHmqxl46/
1127032466952560245_208400646	https://scontent.cdninstagram.com/hphotos-xaf1/t51.2885-15/s320x320/e35/12224216_528172514014193_208730126_n.jpg	ma_werneck	357	https://instagram.com/p/-kBg6zCh51/
1127032259053129045_1063364386	https://scontent.cdninstagram.com/hphotos-xpf1/t51.2885-15/s320x320/e35/12224491_843933249058259_2074435772_n.jpg	rosaspindola2	5	https://instagram.com/p/-kBd5LRI1V/
1127031205351689530_1063364386	https://scontent.cdninstagram.com/hphotos-xfa1/t51.2885-15/s320x320/e35/12298918_409414329257525_543405101_n.jpg	rosaspindola2	4	https://instagram.com/p/-kBOj1xl06/
1127029858585840908_1063364386	https://scontent.cdninstagram.com/hphotos-xtf1/t51.2885-15/s320x320/e35/11410726_1732400716972308_1667866887_n.jpg	rosaspindola2	6	https://instagram.com/p/-kA69kRI0M/
1127027116760927395_1063364386	https://scontent.cdninstagram.com/hphotos-xta1/t51.2885-15/s320x320/e35/12277465_1651265545146779_1441906957_n.jpg	rosaspindola2	7	https://instagram.com/p/-kATECxlyj/
1127022767087804517_5550759	https://scontent.cdninstagram.com/hphotos-xfp1/t51.2885-15/s320x320/e35/11176047_1746199788941518_840992522_n.jpg	karinadeluca7	9	https://instagram.com/p/-j_TxGC-Bl/
1127010794508129544_213943076	https://scontent.cdninstagram.com/hphotos-xpa1/t51.2885-15/s320x320/e35/12301344_1494241850871921_743696963_n.jpg	gabikiepert	15	https://instagram.com/p/-j8liwsBUI/
1126988900318089154_1063364386	https://scontent.cdninstagram.com/hphotos-xfp1/t51.2885-15/s320x320/e35/11925864_235727413427076_390294216_n.jpg	rosaspindola2	3	https://instagram.com/p/-j3m8NRI_C/
1126986877027782507_1063364386	https://scontent.cdninstagram.com/hphotos-xtp1/t51.2885-15/s320x320/e35/12299064_1641858546079356_877095624_n.jpg	rosaspindola2	5	https://instagram.com/p/-j3Jf3xl9r/
1126984169604550411_1063364386	https://scontent.cdninstagram.com/hphotos-xpf1/t51.2885-15/s320x320/e35/12237225_175396786141174_293360101_n.jpg	rosaspindola2	4	https://instagram.com/p/-j2iGYxI8L/
1126946850110986715_53966813	https://scontent.cdninstagram.com/hphotos-xft1/t51.2885-15/s320x320/e35/12299014_1500723910227880_278650898_n.jpg	listentosilva	949	https://instagram.com/p/-juDB5nFnb/
1126938452745491262_515197019	https://scontent.cdninstagram.com/hphotos-xpf1/t51.2885-15/s320x320/e35/12269723_532444920243419_1892748840_n.jpg	amore_amorinh	22	https://instagram.com/p/-jsl1Pn5M-/
1126910185953147616_1265485109	https://scontent.cdninstagram.com/hphotos-xfp1/t51.2885-15/s320x320/e35/12237420_845771522210250_377472245_n.jpg	jackelinetomaz	3	https://instagram.com/p/-jltfvgbrg/
1126900387592210599_1349778009	https://scontent.cdninstagram.com/hphotos-xtp1/t51.2885-15/s320x320/e35/12277552_1646257878972059_2073613864_n.jpg	doeagua2	31	https://instagram.com/p/-jje6Trwin/
1126891648237311899_2293320644	https://scontent.cdninstagram.com/hphotos-xaf1/t51.2885-15/s320x320/e35/12224433_1663577920547157_1262083818_n.jpg	pipoca.pop	15	https://instagram.com/p/-jhfvJiheb/
1126882801939641339_183301787	https://scontent.cdninstagram.com/hphotos-xtf1/t51.2885-15/s320x320/e35/12299062_522865684539358_580229995_n.jpg	ruivoinova	14	https://instagram.com/p/-jffAZG_P7/
1126882545728692385_191959039	https://scontent.cdninstagram.com/hphotos-xtp1/t51.2885-15/s320x320/e35/12256776_741100249329420_940350125_n.jpg	dudulealgv	35	https://instagram.com/p/-jfbRxvyCh/
1126873017893617366_2292805350	https://scontent.cdninstagram.com/hphotos-xaf1/t51.2885-15/s320x320/e35/12224092_1693898767490525_1583170284_n.jpg	sos.riodoce	4	https://instagram.com/p/-jdQoShX7W/
1126871719102780170_1903925899	https://scontent.cdninstagram.com/hphotos-xtp1/t51.2885-15/s320x320/e35/12256736_199751587028399_1319133925_n.jpg	pierriocupaciona	6	https://instagram.com/p/-jc9usjhMK/
1126866260295317280_1658009508	https://scontent.cdninstagram.com/hphotos-xaf1/t51.2885-15/s320x320/e35/12276865_1043287482382886_1772559809_n.jpg	radiontsalvador	11	https://instagram.com/p/-jbuSyRncg/
1126864374265169690_506956936	https://scontent.cdninstagram.com/hphotos-xfa1/t51.2885-15/s320x320/e35/12298980_604666079672885_1626115199_n.jpg	cacalacacala	34	https://instagram.com/p/-jbS2SHkca/
1126856674750776278_179684879	https://scontent.cdninstagram.com/hphotos-xpt1/t51.2885-15/s320x320/e35/12237282_499437583550949_887654174_n.jpg	darlantrainerx	21	https://instagram.com/p/-jZizjTEPW/
1126855704766052916_377318325	https://scontent.cdninstagram.com/hphotos-xap1/t51.2885-15/s320x320/e35/12224444_1692612900950514_2092885266_n.jp	menotecordeiro	63	https://instagram.com/p/-jZUsLs_o0/
1126845387303819699_828875517	https://scontent.cdninstagram.com/hphotos-xfp1/t51.2885-15/s320x320/e35/12224111_981878651834997_2099261666_n.jpg	novezeronove	48	https://instagram.com/p/-jW-jTCBmz/
1126833391329447670_358938938	https://scontent.cdninstagram.com/hphotos-xaf1/t51.2885-15/s320x320/e35/12301297_1082672081745757_1957779019_n.jpg	gabrielapossatto	89	https://instagram.com/p/-jUP_LPY72/
1126825657476847781_173883333	https://scontent.cdninstagram.com/hphotos-xfp1/t51.2885-15/s320x320/e35/12277405_195923670745484_1840919094_n.jpg	fabriciozucoloto	10	https://instagram.com/p/-jSfcdtBil/

Fig. 9 – Table collected from Instagram posts: id, url, like, link. Source: The authors.

local	local_id	lat	long	filter	time	perfil
				Normal	1448574196	https://igcdn-photos-c-a.akamaihd.net/hphotos-ak-xft1/t51.2885-19/11201730_71585576185708
				Normal	1448574020	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Normal	1448573899	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Crema	1448573689	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Normal	1448573602	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Normal	1448572761	https://igcdn-photos-e-a.akamaihd.net/hphotos-ak-xfa1/t51.2885-19/s150x150/11356620_92505
				Normal	1448572736	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Normal	1448572610	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Normal	1448572450	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Normal	1448572123	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Reyes	1448571605	https://igcdn-photos-a-a.akamaihd.net/hphotos-ak-xpa1/t51.2885-19/11934606_1634812530108
				Normal	1448570177	https://igcdn-photos-c-a.akamaihd.net/hphotos-ak-xap1/t51.2885-19/s150x150/12224160_14949
				Normal	1448567567	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Clarendon	1448567326	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
				Normal	1448567003	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/11373733_86319
Regência, Espirito Santo, Brazil	225046346	-19,6	-39,8167	Normal	1448562555	https://igcdn-photos-b-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/s150x150/12135477_16664
				Normal	1448561554	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xfp1/t51.2885-19/s150x150/11939496_93594
				Normal	1448558184	https://igcdn-photos-a-a.akamaihd.net/hphotos-ak-xpa1/t51.2885-19/11910201_1472833846353
				Normal	1448557016	https://igcdn-photos-g-a.akamaihd.net/hphotos-ak-xfa1/t51.2885-19/s150x150/11939640_64064
MMurad FGV	214607849	-20,3055191	-40,2954407	Normal	1448555974	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xta1/t51.2885-19/s150x150/10666210_13078
MMurad FGV	214607849	-20,3055191	-40,2954407	Normal	1448554919	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xfp1/t51.2885-19/s150x150/11313617_45429
Governador Valadares, Minas Gerais	226366478	-18,85	-41,9333	Normal	1448554889	https://igcdn-photos-f-a.akamaihd.net/hphotos-ak-xfa1/t51.2885-19/s150x150/10954793_10436
				Gingham	1448553753	https://igcdn-photos-b-a.akamaihd.net/hphotos-ak-xtf1/t51.2885-19/s150x150/12256665_15147
				Normal	1448553598	https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xtp1/t51.2885-19/11324940_4556487046170
				Normal	1448552948	https://igcdn-photos-a-a.akamaihd.net/hphotos-ak-xap1/t51.2885-19/s150x150/925593_146710
Centro Cultural Da Justica Federal	887965477	-23,01169667	-43,3101815	Normal	1448552723	https://igcdn-photos-a-a.akamaihd.net/hphotos-ak-xfa1/t51.2885-19/10520160_165393871815
				Normal	1448551805	https://igcdn-photos-c-a.akamaihd.net/hphotos-ak-xaf1/t51.2885-19/10727335_373929209449
				Normal	1448551689	https://igcdn-photos-e-a.akamaihd.net/hphotos-ak-xpf1/t51.2885-19/s150x150/11430388_50255
				Normal	1448550459	https://igcdn-photos-f-a.akamaihd.net/hphotos-ak-xat1/t51.2885-19/11887015_88635919810511
Regência, Espirito Santo, Brazil	225046346	-19,6	-39,8167	Normal	1448549029	https://igcdn-photos-f-a.akamaihd.net/hphotos-ak-xap1/t51.2885-19/s150x150/12080562_14781
				Slumber	1448548107	https://igcdn-photos-h-a.akamaihd.net/hphotos-ak-xft1/t51.2885-19/11809992_1668477220033

Fig. 10 – Table collected from Instagram posts: local, lat, long, time, profile. Source: The authors.

Posts were registered all over the World, but, as expected, they were more concentrated in the axis between Belo Horizonte and Vitória, the two main cities in the states where the disaster happened (Figure 11).

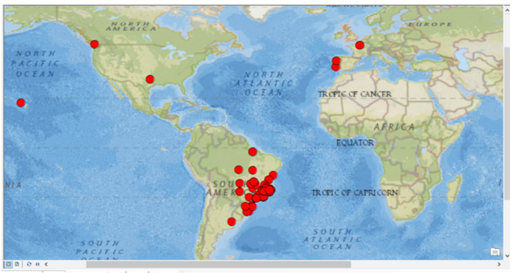


Fig. 11 – Posts all over the World. Source: The authors.

We could observe the posts in Brazil and compare them with the limits of the Rio Doce Basin and its insertion in the bigger basin unit. It was possible to see, in this general capture, smaller scale, the concentration of posts forming a curve along the Rio Doce (Figure 12).

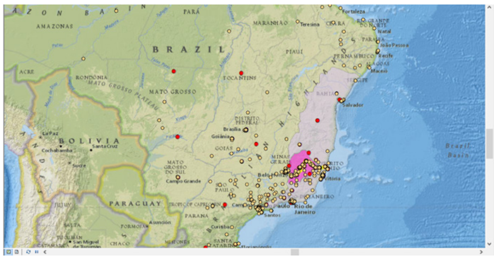


Fig 12 – Posts in Brazil, concentrated from Belo Horizonte to Vitória. In pink, the Rio Doce Basin. Source: The authors.

To make sure we were analyzing posts that were published from people that were in the area, real pictures from the place, real residents or witnesses of the scene, we decided to cut the data sample just in a buffer zone. Posts were collected and selected within a 10km distance from the affected area (Rio Doce) (Figure 13).

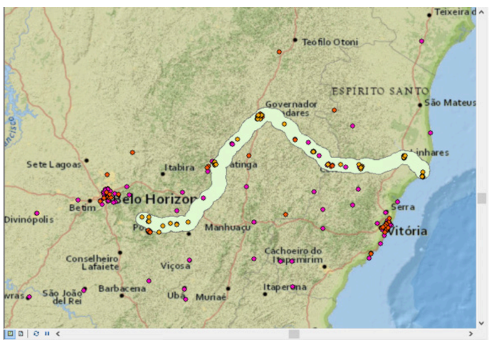


Fig. 13 – Posts in a buffer zone of 10km distance from the Rio Doce. Source: The authors.

The goal was to perform an image analysis separating drawings or memes, landscapes and everyday life images. We were more interested in pictures from daily life, from people that were in the area. Images representing local landscapes, before and after the disaster (Figure 14, Figure 15). But we were also interested in pictures in which people were registering their life at the river, before and after the disaster (Figure 16).

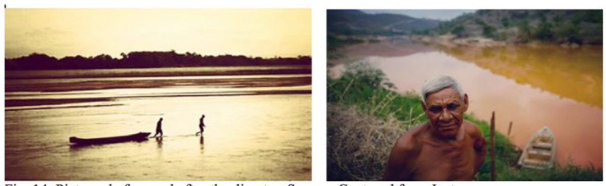


Fig. 14 - Images before and after the disaster.

Fig. 15 - Examples of landscape images. Source: Captured from Instagram.

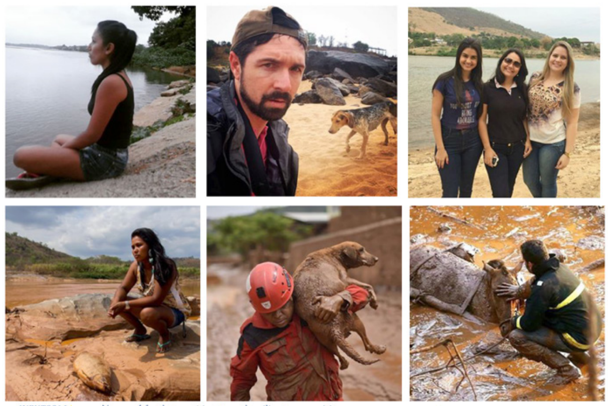


Fig. 16 - Examples of Everyday life images. Source: Captured from Instagram.

However, it was possible to notice the great number of pictures with memes. Meme is a unit of information that multiplies from brain to brain, which somehow self-propagates. In the case of using images, someone works in a drawing or image treatment in order to produce a criticism, a satire, a social manifestation, an opinion. In the case study, we noticed many memes criticizing the companies involved in the disaster (Figure 17) as their logo was used and repeated in several memes.

Fig. 17 -Examples of memes – images of social criticism. Source: Captured from Instagram.

We decided to construct an algorithm to separate the pictures into3 categories, because there were too many to be separated manually, and because we were searching for a logic to be reproduced in other studies, on a large scale. We imported photos into *ArcGIS* 10.2 and divided the photos into 5 different colors. The reason why we chose five was based on studies of mental map's semiology logic applied to cartography, that states the human capacity to understand and read differences (BERTIN, 1977). After dividing into 5 colors, the image was converted to polygons, the area of each polygon was calculated, and the standard deviation of the polygons' areas. At that point each image had a value of standard deviation, and they were classified per ranges of standard deviation. This was done to all the images using Model Building in *ArcGis* (Figure18).

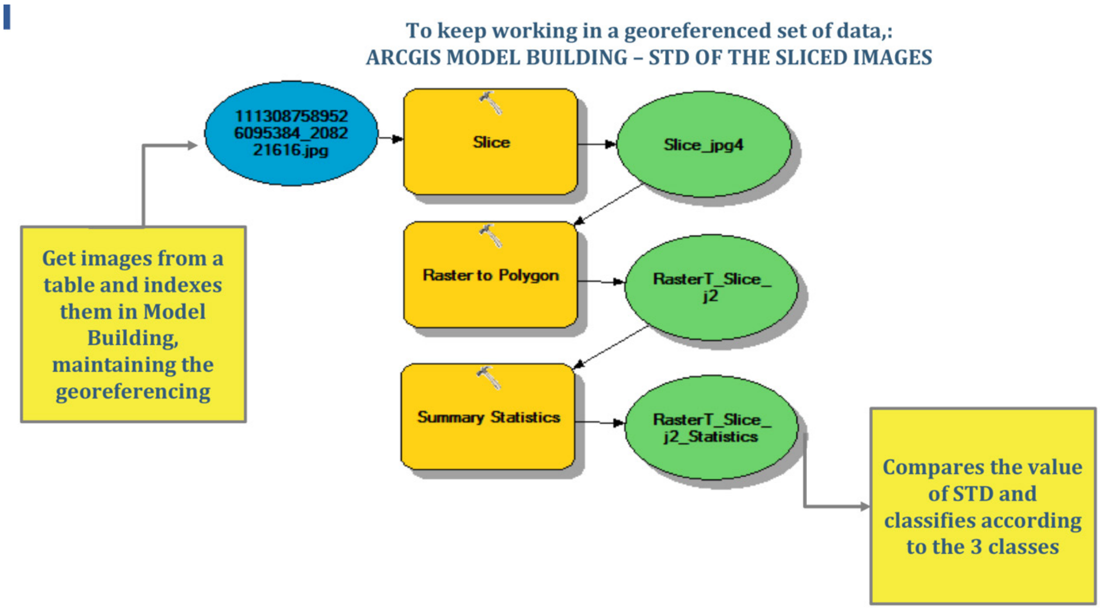


Fig. 18 – Model Building performing steps: dividing, converting from raster to polygons, calculating polygons' areas, calculating standard deviation from areas, classifying the pictures by a range of standard deviation. Source: The authors.

The next step was a statistical analysis to show the standard deviation overall of polygons' areas. We observed that memes tend to have a higher standard deviation (SD) among the areas of the polygons, as they are composed as drawings and tend to have little fragmentation of shapes (Fig. 19).

Fig. 19 – Image transformation – divided into 5 colors – vectorized in polygons. Memes – very different sizes of polygons.

We also observed that everyday life pictures were generally composed by profusion of information and details, and results in low standard deviation (SD) among polygons (Figure 20).

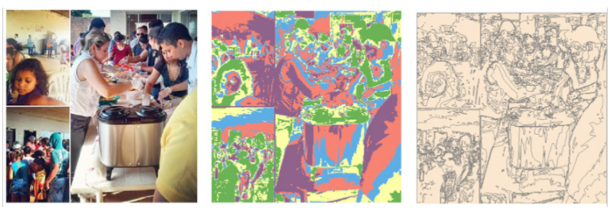


Fig. 20 – Image transformation – divided into 5 colors – vectorized into polygons. Everyday life: many small polygons.

The landscape images were similar to everyday life images, but we observed size differences in the polygons, they were not as small as in the everyday life images and not as big as in memes (Figure 21).

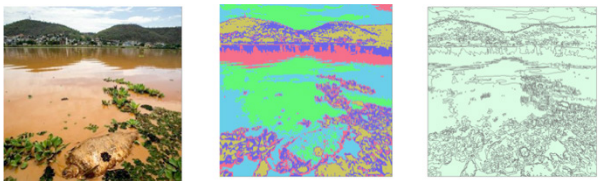


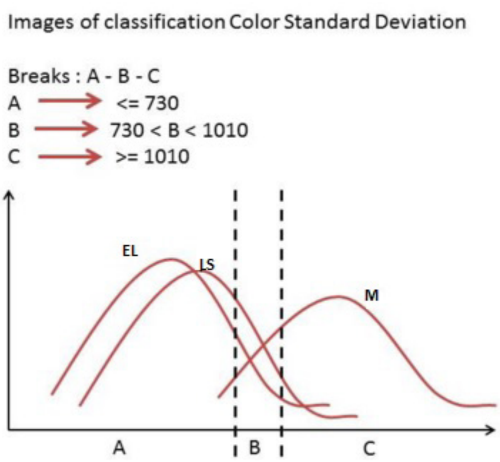
Fig. 21 – Image transformation – divided into 5 colors – vectorized in polygons. Landscape: small and big polygons, but polygons in all kind of sizes. Source: The authors.

3. RESULTS AND DISCUSSION

As we observed, the everyday life images resulted in many small polygons and the standard deviation among the shapes' areas is low. The landscape photo segmentation resulted in large shapes, but also in smaller shapes in some parts, which results in medium standard deviation of polygon's' areas. The drawing memes resulted in very large areas in contrast with very small areas in the segmented representation, which presents a high standard deviation among the shapes' areas.

The images selected were automatically identified by the model (everyday life, landscape and meme) and confirmed by a manual visual classification. Then the statistical analysis was performed to confirm the results (Figure 22).

The curves "EL", "LS" and "M" represent respectively 'everyday life', 'landscape' and 'meme' distribution of standard deviation of the polygon's areas (Figure 22). We can observe that everyday life and landscape tend to have a more similar performance, changing region of media in the distribution. But it is very difficult to separate the pictures from these two groups. On the other hand, the performance of memes were very different from the other ones. The graph shows three ranges presented as "A", "B" and "C". "A" shows that we have an interval in which most images from everyday life and landscape are distributed, and it represents an interval of confusion, in which it was not possible to separate the types. And "C" an interval in which most meme images were distributed.



Break	Total images	Images meme	Images landscape	Images everyday life
A	114	19	42	53
B	20	8	7	5
C	55	36	12	7

Notes on break A: p + vc (A1 + A2)
A2 → <= 400
A1 → 400 < A1 < 730

Landscape and everyday life is the standard deviation very close.

The classification method is valid, whereas the percentage of error in each interval is 10%.

Fig. 22 - Statistic results using *Minitab* (www.minitab.com, accessed in April 2016). Standard deviation performance of polygons' areas, where EL is everyday life, LS is landscape and M is meme.

The quantitative analysis of the procedure applied at #sosriodoce dataset sample shows that the results confirm our first manual analysis with an overall confusion of 10%. Landscape was confirmed as the most challenging class as it has the highest mistake rate. Landscape images essentially mixes results with everyday life. This confirms the perception that a second filter should be proposed to separate the everyday life and landscape categories.

Analyzing the posts in a general sense, using this method of calculating the standard deviation of the polygons in each image, it's possible to identify the same or very similar images. This is very useful to recognize those images that were proposed by one user and replicated by others. The more a post is repeated, the more it was chosen by the users as a representative of a feeling or a way of thinking. This was very strong in the memes posts, because some of them were expressively repeated: the one with the mud, the one with an eye and the mud, and the one that associates Brazilian's disaster with the terrorist attack in Paris, as they happened in the same time (Figure 23).

Fig. 23 - The most repeated memes. Source: Captured from Instagram.

Referring back to the memes, another big group of posts were the ones that criticize the company Vale. It's interesting to observe that none of the posts talked about BHP Billiton (Broken Hill Proprietary Company, associated with Billiton Company), the company that was operating the area with Vale Company. Only the name "Vale" was associated to everything, probably because of a previous social way of thinking, past disasters, or because it's more known in Brazil.

Fig. 24 – The most repeated memes. Source: Captured from Instagram.

The analysis of everyday life images allows us to see that people were very concentrated in helping the local society, sending water and basic supplies, organizing food and accommodation. The difficulties that need to be faced, the loss of important working conditions were also very present in the images (Fig. 25).

Fig. 25 – Everyday life images. Source: Captured from Instagram.

The analysis of the landscape images allowed us to understand that the river was undoubtedly the main visual reference in the area. People present many views from the area, but it was also possible to recognize that the ones more replicated were those taken from the top, in aerial view. Those were not taken by the local people, but were accepted as the feeling they had, and were repeated a lot. The one that became the icon was the image of the mud arriving in the sea (Fig. 26). We were very interested in the pictures takenby people that were in the area, and the point of view they provide. Based on that, we suggest a case study of calculating visual axis from the river and from some main points to manage the land transformation and recovery (Fig. 26).

Fig. 26 – The main and landscape images. Source: Captured from Instagram.

4. CONCLUSION

Passive SMGI is useful for understanding citizens' values and increase knowledge for spatial planning. It represents a rich universe of information for research purposes as it needs less motivation efforts to get people involved as well as less investment in publicity.

A connection between SMGI passive posts and the level of affection to shared content shall be further investigated using a broader sample. The right choice of a hashtag (a dataset primary characteristic) can determine the success of *genius loci* catching. An analysis of number of "likes" could also be carried out by image categories and correlating them to location.

Further investigations into image classification methods should be done. Image classification was very efficient to separate drawings from the other two categories of everyday life and landscape. Confusion between everyday life and landscape could be explained by the fact that it is common to have people in landscape pictures. A second filter and/or image classification should be considered. We can suggest color pattern analysis as a possible path to follow.

The *genius loci* (the essence of the place, and its symbols) can be further investigated through the collective description of images, videos, text and geo-location of posts. The automatic process of the image classification can benefit the exploration into the values and behaviors of citizens involved in SMGI passive posts.

It is clear, especially in Brazil, that social media is used as a mechanism to manifest social and cultural issues. Memes were broadly shared with impact images or drawings with a critique. They represent the collective thought. Quantitative identification (using same area's standard deviation value) and qualitative investigation of memes repetition should lead to good insights about collective values and thoughts. However, in the directly affected area most posts refer to reality, meaning that it is also used as a mechanism to register occurrences.

In passive analytics, the aim is to capture information on society's values and behaviors. That involves other issues that should be further analyzed: Does the sample correspond to reality? Does the profile of posters refer to local people? Can this investigation be performed in different geographic areas? Is this methodological proposal efficient in identifying collective behaviors? Or is it only a registration mechanism? Or maybe it is both?

ACKNOWLEDGEMENTS

The Project "Geodesign and Parametric Modeling of Territorial Occupation: new resources of geo-technologies to landscape management of Pampulha Region, Belo Horizonte", was made possible in part by the support ofCNPq, Call MCTI/CNPQ/MEC/CAPES Nº 22/2014, Process: 471089/2014-1, and to the Project "ProgramaPesquisadorMineiro – PPM IX", Process TEC – PPM – 00059-15.

Contribution to CNPq project "Geodesign e Modelagem Paramétrica da Ocupação Territorial: Geoprocessamento para a proposição de um Plano Diretor da Paisagem para a região do Quadrilátero Ferrífero-MG", Process 401066/2016-9, Universal Call 1/2016.

We kindly thank the courtesy of LABIC (Laboratório de Imagem e Cybercultura - Laboratory of Image and Cyberculture of Federal University of Espirito Santo, Brasil) for the use of the API Leticia.

REFERENCES

AHAS, R.; MARK, U. Location based services: new challenges for planning and public administrations? **Futures**, 2005, n. 37, p. 547–56.

AHAS, R.; SILM, S.; JARV, O.; SALUVEER, E.; TIRU, M. Using mobile positioning data to model locations meaningful to users of mobile phones. **Journal of Urban Technology**, 2010. v. 17, n. 1, p. 3-27.

BAO, J.; ZHENG, Y.; WILKIE, D. Recommendations in location-based social networks: a survey. **GeoInformatica**, v. 19, n. 3, p. 525-565, 2015.

BAO, J.; ZHENG, Y.;WILKIE, D.; Mokbel, M. F. **Geoinformatica**, v. 19, n.3, 2015, p. 525-565.

BERTIN, J. **A neográfica e o tratamento gráfico da informação**. Curitiba, Editora da Universidade do Paraná, 1986. 273p.

BORGES, J. L. C; JANKOWSKI, P.; DAVIS JR, C. A. A Study on the Use of Crowdsourced Information for Urban Decision-Making. **Revista Brasileira de Cartografia**, Rio de Janeiro, v.68, n. 4, 2016, p. 695-703.

BORGES, J.; JANKOWSKI, P.; DAVIS JR, C. Crowdsourcing for Geodesign: Opportunities and Challenges for Stakeholder Input in Urban Planning. In.: SLUTER, C. R.; CRUZ, C. B. M.; MENEZES, P. M. L. **Cartography - Maps Connecting the World**. Switzerland, Springer, 2015. p. 361-373.

BORGES, J.; JANKOWSKI, P.; DAVIS JUNIOR, C. Crowdsourced information from Tweets during the WorldCup in Brazil: A theme search. **Changing Cities II**, Spatial, Design, Landscape & Socio-economic Dimensions, 2015. p. 1511-1519.

BRUNS, A.; BURGESS, J.; CRAWFORD, K.; SHAW, F. #qldfloods and @QPSMedia: crisis communication on Twitter in the 2011 South East Queensland Floods. **ARC Centre of Excellence for Creative Industries and Innovation (CCI)**, Brisbane, 2012. 57p.

CAMPAGNA, M. Social Media Geographic Information: Why is special when it goes spatial? **Energic Book**, Ubiquity Press. 2016. p.45 - 54.

CAMPAGNA, M.; FLORIS, R.; GIRSHEVA, A.; IVANOV, K.; MASSA, P. The role of Volunteered Geographic Information (VGI) and Social Media Geographic Information (SMGI) in Spatial Planning. **Workshop "Role of Volunteered Geographic Information in Advancing Science: Effective Utilization"**, Eighth International Conference on Geographic Information Science, TU Vienna. September 23, 2014. 3p.

CAMPAGNA, M.; KUDINOV, A.; GIRSHEVA, A.; IVANOV, K.; KOPNOV, M.; FALQUI, R. Place, I

CHEN, Z.; FU, R.; ZHAO, Z.gMission: a general spatial crowdsourcing platform. **PVLDB**, v. 7, n. 13, 2014, p. 1629-1632.

CHENG, T.; WICKS, T. (2014) Event Detection using Twitter: A Spatio-Temporal Approach. **PLoS ONE**, 2014, v. 9, n. 6, e97807. Disponível em <https://doi.org/10.1371/journal.pone.0097807>

CRAGLIA, M. Volunteered geographic information and spatial data infrastructures: when do parallel lines converge? **Workshop on Volunteered Geographic Information**, Santa Barbara, California, USA, 2007, 3p.

DAVIS Jr, C. A.; MORO, M. M.; VEZULA, G. M.; MACHADO, N. G. Contribuições Voluntárias: Impactos Potenciais dos Cidadãos On-Line e seus Dispositivos Móveis. In.: MOURA, A. C. M. (ed.) **Tecnologias de geoinformação para representar e planejar o território urbano**. Rio de Janeiro, Interciência, 2016, p.23-34.

EARLE, P.; BOWDEN D.; GUY, M. Twitter earthquake detection: earthquake monitoring in a social world. **Annals of Geophysics**, 2011, v. 55, p. 708–713.

GOODCHILD, M. Citizens as sensors: the world of volunteered geography. **GeoJournal**, 2007, v. 69, p. 211–221.

JANKOWSKI, P; ANDRIENKO, G.; ANDRIENKO, N.; KISILEVICH, S. Discovering Landmark Preferences and Movement Patterns from Photo Postings.**Transactions in GIS,** 2010, v. 14, n. 6, p. 833–852.

MATEVELLI, G. V.; MACHADO, N. G.; MORO, M. M.; DAVIS JR, C. A. Taxonomia e Desafios de Recomendação para Coleta de Dados Geográficos por Cidadãos. **Anais do XXX Simpósio Brasileiro de Bancos de Dados**, Sociedade Brasileira de Computação (SBC), 2015.

NASA Earth Observatory images by Joshua Stevens, using Landsat data from theU.S. Geological Survey. Caption by Kathryn Hansen. Instrument(s): Landsat 8 – OLI.

PUCCI, P.; MANFREDINI, F.; TAGLIOLATO, P. Mapping Urban Practices Through Mobile Phone Data. **PoliMISpringerBriefs**, 2015. 90 p.

REUTERS (2015, November 9) Brazil mine dam burst endangers water supply far downstream. (Accessed November 13, 2015).