

VGI DAY: REPORT

The concept.....	2
Motivation.....	2
VGI day: shake the community.....	2
The implementation.....	3
Preparing the VGI day.....	3
Seed influencers	4
Generate discussion: shake the community via interesting discussions	4
Track communication: VGI day hashtags.....	4
Behavior rules during the VGI day.....	4
Technical architecture.....	5
The results	6
Map of tweets	6
Participants	7
Density of tweets in time.....	8
Most used keywords	9
Sample of tweets	10

THE CONCEPT

The IC1203 ENERGIc has included as part of the dissemination activity a special event called "VGI day". The first VGI day has been organized the same day in some participating countries and institutions in the form of workshops and laboratories, panel discussions and twitting activity on the topic of VGI / Citizen Science in order to work on on-going ideas and challenges. ECIs and Phd students has been involved.

The COST Energi network (see www.vgibox.eu) has run a 2 day geolocated twitter chat, titled 'Volunteered Geographic Information Day' and the hashtag used is #VGIDay.

The conversation took place on 14th and 15th May 2015. Besides several activities have taken place in ENERGIc member countries to celebrate VGI. This will event has the aim to spread the topic and the Action's activities through the Twittersphere and beyond.

Discussions were started by @COST_Energi. Through this twitter handle, the aim was to share resources, results and ideas about the topic of VGI and geographic crowdsourcing. Members of Action and any potential VGI fan have been encouraged to join the discussions, bring ideas and links, and involve other contacts.

At the end of the experiment, a report has been produced on the generated discussion for our ENERGIc repository, and the dataset of tweets can be then used by researchers who want to visualise, analyse and try to do things with it. It might also end up as teaching material.

Joining was easy and required 3 steps:

1. Follow the @COST_Energi profile
2. Enable your phone to disclose your position – this will allow to geocode your tweets.
3. To participate to the discussion, use at least one of the dedicated hashtags in tweets:
#COSTEnergi, #VGIDay

MOTIVATION: VGI DAY: SHAKE THE COMMUNITY

Volunteered Geographic Information (henceforth: VGI) is becoming a hot topic in research, due to the fact that crowdsourcing itself has become an essential tool for solving problems that cannot be solved by machines.

People are voluntarily offering their work, for free, and since these workers are actually volunteers and not paid, works in the state of the art prove that the quality of their works is high.

However, many aspects of the topic have not been tackled yet, and there is enough room for discussion, looking for new solutions and new applications that can exploit volunteered work to enrich geographical digitalized data.

The concept behind the VGI day is to shake the VGI community and produce communication on the topic, via the Twitter microblog platform.

THE IMPLEMENTATION

In the following, we report details about the pipeline used to collect tweets generated by the VGI community during the VGI day event.

PREPARING THE VGI DAY

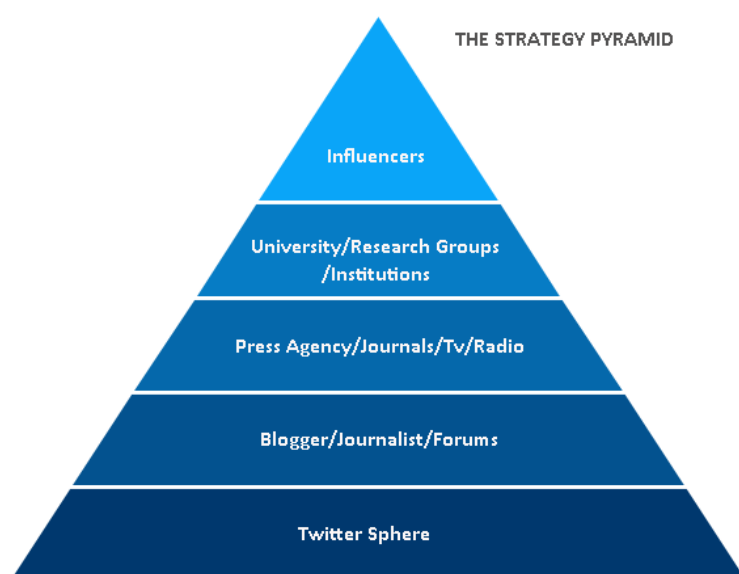
To get prepared to the collection of the content related to the VGI day, we needed a way to share and trace information and discussions through the twittersphere.

STEP 1

As a first step, we tackled the problem of sharing information and generating discussion through Twitter. Usually, identifying *influential entry points* in Twitter is a way of diffusing information effectively, allowing one to spread communication through the target community. These entry points are nodes in the Twitter graph (or, simply, *influential users* [henceforth: *influencers*]). Typically, if we involve influencers in the communication, other users that were not previously aware of the VGI day are in turn involved, thus propagating the discussion through the network. This happens because influential users produce visible and relevant content, and the large network observing them every day becomes aware of that content as soon as it is published. The larger the network around the set of influential users, the larger the diffusion degree.

The strategy pyramid and the @Profile identification:

- **400 @profiles:**
 - influencers;
 - stakeholders;
 - institutions;
 - university;
 - research groups;
 - researchers;
 - professors;
 - journals;
 - Tv;
 - radio;
 - national and international press agency;
 - journalist;
 - bloggers;
 - forums;
 - friends.
- **27 countries**
- **2 days** of tweeting activities
- More than **100 tweet** sent



STEP 2

As a second step, we tackled the problem of tracking the communication through space and time. In order to do so:

- Every involved user had to enable geotagging on the produced posts

- Every produced post had to be recognized as belonging to the VGI day discussion

In the following, we described how we built the initial set of influencers, how we generated discussion over the network, how we tracked posts belonging to the VGI day discussion and how we instructed participants on how to behave during the VGI day.

SEED INFLUENCERS

In order to build the initial set of influencers (i.e., relevant users in the ENERGIc and international VGI community that could help us in diffusing information over the network) we:

- Collected the Twitter profiles of the COST Energic participants (via Basecamp)
- Manually identified other profiles that were followed by the previously identified influencers, and that are somehow related to the VGI field

Then, we asked the influencers to share the news about the beginning of the VGI day, some days before the event itself started.

GENERATE DISCUSSION: SHAKE THE COMMUNITY VIA INTERESTING DISCUSSIONS

In order to generate discussions and make the community participate, we created the *@COST_Energic* Twitter profile and collected a set of recent and relevant articles / publication graphs / discussion hints that could be launched as provocative tweets.

Then, we asked the influencers (by mentioning them on Twitter *before* the event started) to follow *@COST_Energic* and to suggest to their followers to do the same. This allowed us to have a concrete visibility level through the community.



TRACK COMMUNICATION: VGI DAY HASHTAGS

In order to track content meant to answer to any discussion related to the VGI day, we asked the participants to use one of the following hashtags in *each post* they produced during the event:

- #VGIday
- #COSTEnergic

BEHAVIOR RULES DURING THE VGI DAY

We asked to each participant to:

- Enable geotagging of produced content
- Enrich as much as possible the content (via multimedia content and external content)
- Follow *@COST_Energic* Twitter profile
- Use the *#VGIday* or the *#COSTEnergic* hashtags to track their participation to the conversation

The behavior rules were communicated via Twitter (using the @COST_Energic Twitter profile), via Basecamp (targeting the COST Energic community) and via influencers (who retweeted the rules so as to make their friends see them).

Some results:

- 1) Tweet from @COST_ENERGIC Twitter profile:
 - **600** tweets visualized (number of times users have seen the tweet on Twitter)
 - **300** visualizations per day
 - **1%** Interaction Rate (number of times users have interacted with the tweet. Covers all click anywhere in the tweet)
 - **3** click per day on links contains in tweets

Impression	Interaction	Interaction Rate	retweet	click on user profile	url clicks	click on hashtag
49	2	0.040816	0	1	1	0
16	1	0.0625	0	0	1	0
48	4	0.083333	0	3	1	0
12	1	0.083333	0	0	1	0
34	1	0.029412	0	0	1	0
46	5	0.108696	0	1	3	0
34	3	0.088235	0	1	2	0
27	2	0.074074	0	0	2	0
209	7	0.033493	1	1	1	3

TABLE 1. TWEET ANALYTICS & FEEDBACK FROM MENTIONED USER'S PROFILE - SAMPLE

Sources: Twitter Analytics

TECHNICAL ARCHITECTURE

We collected tweets via a Streaming Twitter crawler (provided via the Twitter API), which collected automatically and in real-time tweets:

- Either containing the hashtag(s) #VGIday and/or #COSTEnergic
- Or produced by one of the influencers

The tweets were downloaded with the full data structure provided by Twitter, in JSON format, and stored in a MongoDB database. The collected tweets contain references to:

- The text of the tweet
- (In case attached) Links to the multimedia objects attached to the tweet
- (In case enabled) Position of the user when he/she created the tweet
- Data about the user (screen name, profile image)

THE RESULTS

We analyzed the content in the database, filtering out all the tweets that were not containing the #VGIday/#COSTEnergetic hashtags, and we extracted a report of the data we collected.

MAP OF TWEETS

Since users were asked to enable geotagging on the produced content, we were able to extract positions from the produced content and show it in a map:

FIGURE 1. #EUROPE4X4 - TOP DOMAINS AND SHARE OF POSTS



FIGURE 2. VGI DAY TOP LOCATIONS AND DEMOGRAPHICS

Country	Where in the world the post are originating from (%)
Italy	83%
USA	6%
UK	3%
Germany	1%
Ireland	1%
Spain	1%
Portugal	1%
Greece	1%
Austria	1%
France	1%
Switzerland	1%

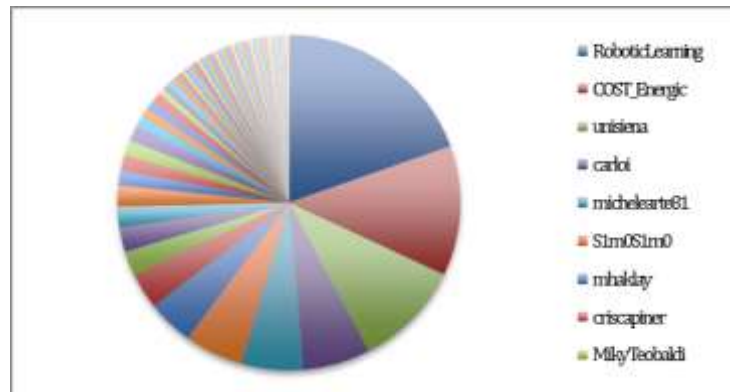
Most of the geotagged tweets were produced in Tuscany (Italy) since during the first day of the VGI event there was a public event involving people from the VGI community in Siena, that started using the #VGIday hashtag to describe this public (local) event. Most of the content that

was produced that day is, in fact, in Italian language, since participants of this event were all Italians.

PARTICIPANTS

A total of 133 people participated to the VGI event by publishing content containing either #VGIday or #COSTEnergic.

In the following, we show the most active users:



On average, each user published 5.44 tweets in total. However, a few users produced a large number of tweets, while the majority produced just one VGI-related tweet. The top-5 users are the following:

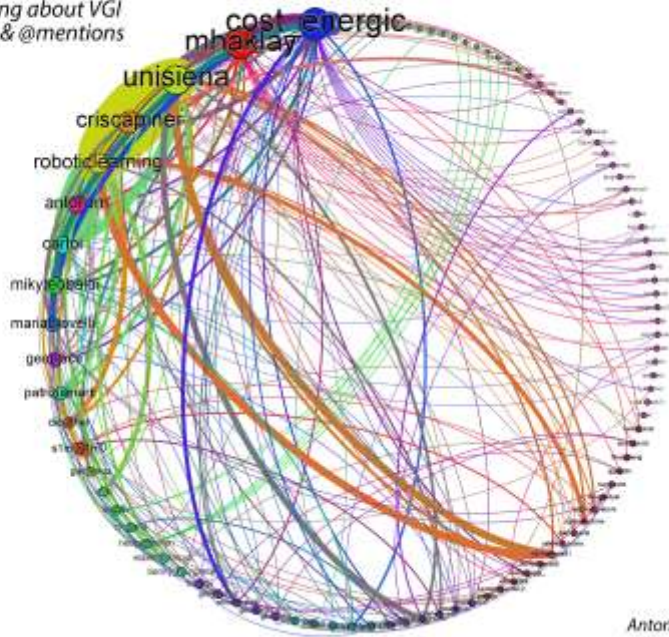
1. @COST_Energic, with 73 tweets
2. @RoboticLearning, with 67 tweets
3. @unisiena, with 35 tweets (in Italian, related to the local event taking place in Siena)
4. @carloi, with 22 tweets
5. @michelearte81, with 20 tweets

Notice that if a user was talking about VGI day (respecting the rules we suggested), but her tweets were private (due to privacy constraints on the user profile), her tweets could not be captured by the stream crawler, and thus could not end up in the final report.

THE NETWORK OF MAIN PARTICIPANTS

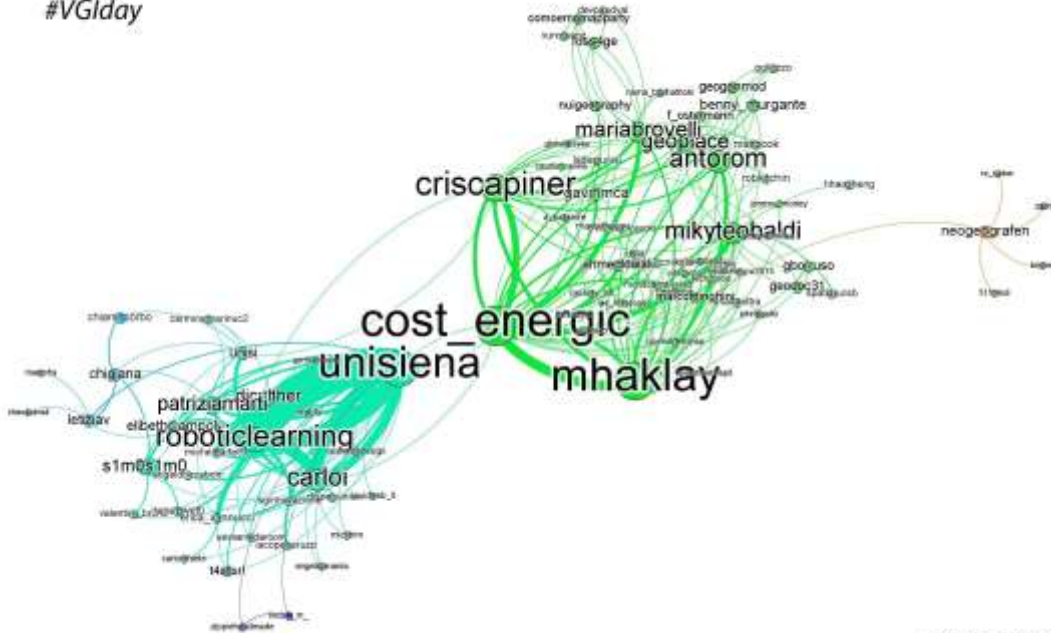
#VGIday Network

Tweeting about VGI
@reply & @mentions



Antonello Romano
@Antorom

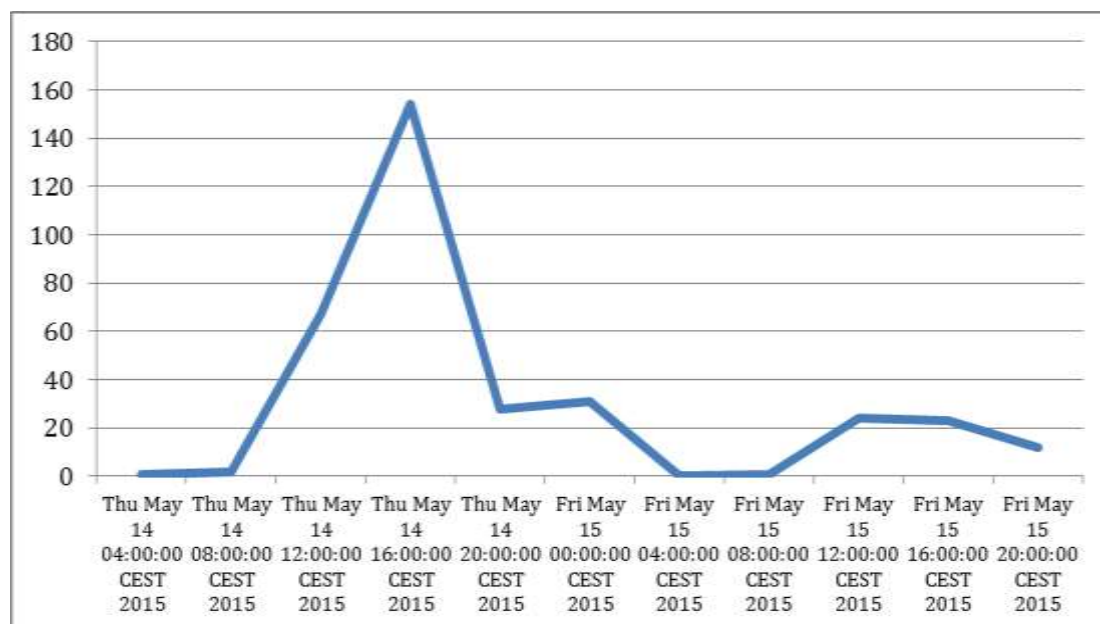
#VGIday



Antonello Romano

DENSITY OF TWEETS IN TIME

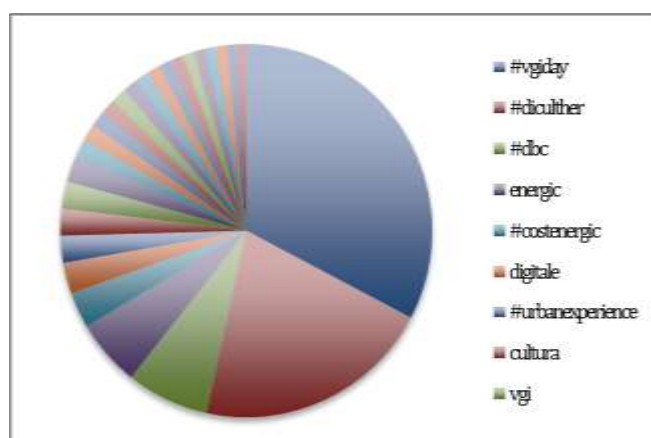
In the following, we show a density graph of the produced tweets over time.



The participation to the event was high during the first day, due to the previously mentioned local event that was occurring in Siena. Here, most of the content we captured was Italian and did not follow the main discussion produced via the @COST_Energic official Twitter profile: discussion was mainly focused on what was going on during this local event and was self-moderated by the participants.

MOST USED KEYWORDS

In the following, we report the most used keywords and hashtags in the retrieved tweets. Since content was produced in both Italian and English, keywords are mixed in two languages.



The top-5 hashtags that were used during the event were:

1. #VGIday, with 389 occurrences
2. #diculther, with 246 occurrences (mostly during the first day)
3. #dbc, with 85 occurrences (mostly during the first day)
4. #COSTEnergic, with 37 occurrences
5. #urbanexperience, with 28 occurrences (mostly during the first day)

The top-5 keywords that were used during the event were:

1. digitale (Italian), with 32 occurrences (mostly during the first day)
2. cultura (Italian), with 28 occurrences (mostly during the first day)
3. VGI, with 28 occurrences
4. culturale (Italian), with 27 occurrences (mostly during the first day)
5. spatial, with 19 occurrences

SAMPLE OF TWEETS



Finally the Storify tool was used to create a report on the web interaction and they are available here:

<https://storify.com/IolandaIacono/walkabout>

<https://storify.com/EleonoraCiceri/vgi-day>

A WORDLE OF THE @COST-ENERGIC INTERESTS

A wordle of the ENERGIIC interests

