Exploring the Use of #OpenData in UK Open Government Data Community

Ramine Tinati

School of Electronics and
Computer Science,
University of Southampton,
Southampton. England
rt506@ecs.soton.ac.uk

Leslie Carr

School of Electronics and
Computer Science,
University of Southampton,
Southampton. England
lac@ecs.soton.ac.uk

Susan Halford

School of Social Sciences,
University of Southampton,
Southampton.
England
sjh3@soton.ac.uk

Catherine Pope

Faculty of Health Sciences,
University of Southampton,
Southampton.
England
cjp@soton.ac.uk

1. INTRODUCTION

The UK Open Government Data (OGD) community has become a vibrant and expanding community during recent years [12]. Comprising of individuals, businesses, organisations, academia, and government, the community has helped establish a common set of principles that aim to decrease the barriers of citizens-to-government engagement, increase economic growth, and inspire a culture of data reuse and openness [2][8].

The boundaries of Open Government are not limited to just the UK, it is happening worldwide. Supported by the Open Government Partnership (OGP), over 47 countries have committed to achieving transparency and accountability through the opening up of public data in free, re-usable formats. Building upon the success of early adopters of open data, countries are able to share and learn from each other's experiences, and improve their chances of achieving a complex socio-technical agenda.

The agenda of an Open Government extends beyond the promise to make public processes more transparent, it is an agenda bound by a social and technical change. Within the UK, the commitment towards openness and access to unpublished government data has been realised by the launch of data.gov.uk in 2010, which represents the national Web portal for all of the UK government's data. Data.gov.uk embodied the commitments towards a nationwide scheme, driven partly by government, and also by the social and technical activities of the community that support Open Data [12].

Since the launch of data.gov.uk, the number of datasets published has increased, and as of June 2013, just fewer than 10,000 unique datasets have been published. Originally rolled out by central government, the variety of datasets – in size and topic – in data.gov.uk has increased since local government, councils and local authorities have committed to publishing their data [8].

Open Government is a socio-technical activity [13]; it requires the support and cooperation of both online and offline activities in order to succeed. Whilst offline activities such as 'hackathons' were essential in bringing together a community of interested individuals, the use of social networking Web platforms such as Twitter has become the active channel for communication and dissemination of news and information [14][12].

The adoption of Twitter and the #opendata hashtag has been recognised as a backchannel for those within the community to interact with each other, and has featured within the title of the UK government latest OGD publication; "*Unleashing the potential of #Opendata*" [8].

Twitter has provided another forum for discussion and interaction for citizens and government to interact, however, the extent to which the platforms have helped and supported the agenda and mobilization of OGD is currently unknown. In order to explore the role and impact of Twitter within the OGD community, this paper will examine its use as a tool for engaging and supporting the growth and activities of the community

2. THE RISE OF SOCIAL NETWORKING: MORE THAN COMMUNCATIONS

Over the last decade, social networking has become a prominent Web activity; with current records indicating that three of the top five most visited Websites are social networking platforms[1]. Services such as Facebook and Twitter are providing new ways for humans to communication and engage, and as a result of the Web's global reach, individuals are able to become part of a network without the physical restrictions of location or time. With certain platforms boasting figures of over 1 billion active users [7], or 400 million daily messages per day [5], social networking offers a unbound resource for communicating at a global scale.

Indeed, Social networking is not only providing a platform for humans to communicate and connect with each other, it is also a platform for information to be disseminated, shared amongst these temporal networks which are formed through the alliances and associations between humans [1]. Mediated through the capabilities and features of a platform; information, in the representation of videos, images, URLs and files can be shared virtually instantaneously.

Platforms such as Twitter occupy a dual purpose; functioning as both as social networking site and a new broadcasting service[9]. Amongst many of the uses of Twitter which include the detection of disease [11], the classification of political opinions [6], or the classification of influential individuals [3], Twitter is a prominent tool for enabling communities around specific online and offline activities to communicate, debate, and share information [4].

Furthermore, individuals are no longer limited to the boundaries of a single technological platform or service [10]. As a consequence of this, other Web activities are able to harness the wealth of information available to improve and support their own growth. An example of this is the online activities of the UK Open Government Data activities, where those committed have adopted Twitter as a way to establish networks, communicate and share news and information.

---

[1] http://www.alexa.com/topsites (Accessed on 20th June 2013)

These technologies are helping reshape traditional organisational structures of citizen-to-government engagement [14], and have been shown to be fundamental to the growth and mobilisation of the stakeholders committed to its success [12].

Undoubtedly, social networking technologies are providing new ways for communities to interact at scale. However, is it possible to comprehend the benefits that platforms such as Twitter (directly or indirectly) have on communities such as UK Open Government Data? What, if any, impacts do technologies such as these have, and how do they translate to improving the current state of the UK Digital Economy?

3. EXPLORING THE ROLE OF SOCIAL NETWORKING FOR SUPPORTING OPEN GOVERNMENT DATA

To explore the role and impact that social networking technologies have on the success of the UK Open Government Data community, the following analysis will explore the relationship between the use of Twitter's #opendata communication channel and the access and downloads of the published datasets on data.gov.uk.

The following analysis uses a corpus of 398,435 tweets made by 101,321 unique Twitter users, which contain the #opendata hashtag, collected during the period of July 2012 to June 2013. Based upon previous studies [12], tweets containing the #opendata hashtag have been identified as a communications channel to discuss the variety of activities that involve the UK Open Government Data community.

As the #opendata hashtag is used as a worldwide channel to discuss OGD, this analysis will focus on an extracted subset of the full #opendata tweets, selected based on their relevance to the activities of the UK Open Government Data community. This sub-set of tweets contains 10,432 messages and 4,933 unique users.

The analysis also uses the daily dataset deposit available at data.gov.uk which contains a metadata log of the complete list of datasets available, including relevant metadata such as department responsible for publishing the data, and format of the data. In addition to this, the metrics of the data.gov.uk website will be used, which contain the number of views, visits, and downloads between July 2012 and June 2013 ('Download' metrics only available since January 2013).

3.1 #OpenData Twitter Communications Channel

The frequency of #opendata tweets between July 2012 and June 2013 varied considerably during the 11-month period, both within the full #opendata communications channel, as well as the communications of UK OGD. Exploring the communications of the subset of tweets further, by looking at the data with a finer level of granularity retweets and mentions accounted for 66% of and 30% of the entire tweet stream, respectively. In respect to such high amounts of communications (only 4% of tweets were not communications or shared messages); the analysis is an indication that the #opendata channel is frequently being used by the Open Data community to communicating and interacting with each other. Both retweets and mentions offer a way for individuals to either share information to potentially new audiences and mentions can be used to conduct conversations between users.

In order to examine if and how Twitter is being used as a platform to engage with other Web activities, the sub-set of #opendata

tweets was further examined in terms of the content within the tweets, extracting messages containing external links. 8,613 (83%) of the 10,432 tweets identified contained an external link (URL in the form of HTTP/HTTPS), and within this set of tweets, 6288 (72%) of them were identified as retweets.

This initial analysis of the Twitter #opendata dataset has illustrated that the proportion of retweets and mentions in comparison to 'plain' tweets indicate that the #opendata communication channel is providing a forum for individuals to connect, communicate, and share information with each other. Furthermore, the high proportion of external links (83%) embedded within the tweets is an indication that the communications made are more than just internal messages within a community. As Table 1 shows, the #opendata stream is being used to point to external sources, including data.gov.uk.

In an attempt to explore this further, the analysis will examine the metrics of the number of published datasets, their formats, and the number of views and downloads within data.gov.uk.

3.2 Data.gov.uk Metrics, Views, Downloads and Datasets

As Figure 1 illustrates, the number of views (the number of times an individual URL was requested) and visits (unique requests) to data.gov.uk between July 2012 and June 2012 have varied, with a total of 5,696,729 total page views and 1,674,843 visits made within the 11-month period. In addition to this, as Figure 2 illustrates, the number of available datasets published in data.gov.uk and the number of downloads have increased since the beginning of 2013. As of current, there are 9590 datasets available, which have received a total of 249,724 downloads since January 2013.

Examining the available datasets in data.gov.uk at a finer level of granularity, the types of data formats published can be exposed; 65 formats have been identified, and CSV, XLS, PDF and HTML account for 96% of the total published datasets.

3.2.1 Correlation Between #Opendata Communications and data.gov.uk Metrics

As a way to explore whether the communications of #opendata (the subset of UK OGD tweets) has any impact on the usage of data.gov.uk, an analysis of the correlation between the number of page views and the volume of communications was conducted. Calculating the spearman rank correlation between views and tweets presents a correlation of 0.55, which is an indication that online communications have an effect on the usage of data.gov.uk.

Referring back to the top external sources embedded within the #opendata tweets shown in Table 1, peaks in activity identified during October 2012 and March 2013 corresponded to the publication of new datasets, which was widely discussed on Twitter, as well as other news websites and discussion forums. During this period (March 2013) there was also a rise in the number of downloads of datasets.

4. CONCLUDING REMARKS

This paper has examined the relationship and impact of online social networking technologies such as Twitter in relation to the activities and growth of the UK Open Government Data community. The analysis of the #opendata communications stream revealed that services such as Twitter offer more than a forum for discussion; they provide a service to share news and

information, which ultimately helps the engagement of the community.

By exploring the external links from the #opendata communications stream, the analysis illustrated how the use of embedded objects such as URLs within tweets are related to the number of visits and potentially the number of download that data.gov.uk obtains. The analysis has shown that specific events within the #opendata communication channel, such as the publication of new data or the launch of new services can be responsible for this. However, in order to explore this further, additional data is required which contains data.gov.uk website metrics at a finer level of granularity. By exposing this data will help expose the content of the embedded links within the #opendata communications channel.

Ultimately, the analysis in this paper has demonstrated that services such as Twitter are critical for the communications and activities of other Web communities such as UK Open Government Data. Social networks are a platform for discussion and functions as a means to share news, resources and networks to help improve community awareness and engagement.

5. REFERENCES

1. Boyd, D., Golder, S., and Lotan, G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *HICSS-43*, IEEE Comput. Soc (2010).
2. Cameron, D. Letter to Government departments on opening up data. *Number10.gov.uk*, 2010. http://www.number10.gov.uk/news/statements-and-articles/2010/05/letter-to-government-departments-on-opening-up-data-51204.
3. Cha, M. and Gummadi, K.P. Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social*, (2010).
4. Choudhury, M. De. Tie Formation on Twitter : Homophily and Structure of Egocentric Networks. *Distribution*, (2011).
5. CNET. *Twitter hits 400 million tweets per day, mostly mobile*. New York, NY, USA, 2012.
6. Conover, M.D., Ratkiewicz, J., Francisco, M., Gonc, B., Flammini, A., and Menczer, F. Political Polarization on Twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, AAAI (2011), 89–96.
7. Facebook. Facebook - Statistics. 2011. https://www.facebook.com/press/info.php?statistics.
8. HM Government. *Open Data White Paper, Unleashing the Potential #opendata*. London, 2012.
9. Kwak, H., Lee, C., Park, H., and Moon, S. What is Twitter, a social network or a news media? *19th international conference on the World Wide Web*, (2010), 591–600.
10. Nielsen. *State of the Media: The Social Media Report 2012*. 2012.
11. Sadilek, A. and Kautz, H. Modeling spread of disease from social interactions. *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, (2012), 322–329.
12. Tinati, R., Carr, L., Halford, S., and Pope, C. Exploring the Impact of Adopting Open Data in the UK Government. *Digital Futures 2012*, (2012).
13. Tinati, R., Halford, S., Carr, L., and Pope, C. Mixing Methods and Theory to Explore Web Activity. *Proceedings of the 3rd Annual ACM Web Science Conference (WebSci '12)*, ACM (2012).
14. Wigand, F.D.L. and Ph, D. Twitter Takes Wing in Government : Diffusion , Roles , and Management. *Social Influence*, (2010), 66–71.

Table 1 External Links Embedded within #OpenData UK OGD Tweets

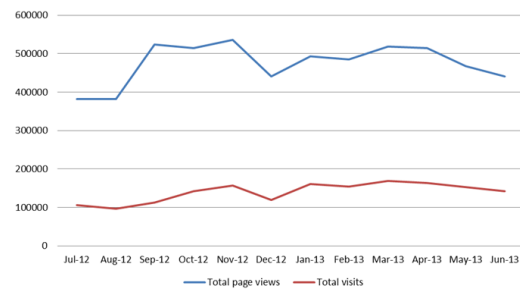| External URL	Percentage of External Links (%)
www.guardian.co.uk	12
www.economist.com	10
radar.oreilly.com	10
www.theodi.org	8
www.gov.uk	8
data.gov.uk	8
www.eventbrite.com	6

Figure 1 Volume of Page Views and Visits for data.gov.uk during July-12 and June-13.
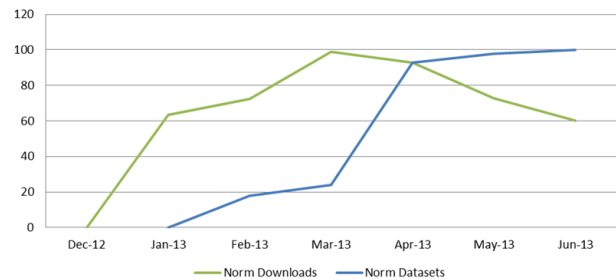

Figure 2 Volume Available Datasets and Dataset Downloads (Normalized) in data.gov.uk during Jan-13 and June-13.