Exploring big 'crisis' data in action: potential positive and negative externalities

Rachel L. Finn
Trilateral Research & Consulting
rachel.finn@trilateralresearch.com

Hayley Watson
Trilateral Research & Consulting
hayley.watson@trilateralresearch.com

Kush Wadhwa
Trilateral Research & Consulting
kush.wadhwa@trilateralresearch.com

ABSTRACT

As growing emphasis is placed on engaging with big 'crisis' data, including data from social media, GPS, and satellite, adequate policies and measures must be in place in order to use this data in an ethically and legally responsible manner. The current working paper introduces the BYTE study, which is working towards identifying and understanding the various positive and negative externalities, or impacts, associated with the use of big 'crisis' data. This insight paper provides a preliminary discussion of various externalities that may be encountered in this study. By doing so, the authors highlight the need for additional research in this area to promote ethically and legally responsible crisis data practices.

Keywords

Big data, crisis data, social media, humanitarian aid, social impacts

INTRODUCTION

The term 'big data' has become one of the most used buzzwords in the technology world in recent years. In April 2013, a Google search for the term 'big data' revealed more than 30 million hits (Beyer et al., 2013). As of the 15 January 2015, a search for the same term on Google.co.uk revealed approximately 768 million hits. But what exactly do we mean by 'big data'? For most, big data is characterised by the original Gartner definition and encompasses data that is high volume, high velocity and high variety (Laney, 2001). Others have added more criteria, including the ability to evaluate the veracity of the data, the ability to visualise the data outputs (Networked European Software and Service Initiative, 2012), or the ability to enhance insight or improve decision-making (Australian Department of Finance and Deregulation, 2013) to this definition. Within crisis management, engaging with large data sets relevant to managing crises, what we refer to as big 'crisis' data, has the potential to significantly enhance preparedness, response and recovery activities.

By big 'crisis' data we mean data stemming from various sources that can be used within crisis management practices. While this involves integrating traditional crisis data about the physical environment, (GIS data, seismic sensors, and others) large volumes of high-velocity social media data is also increasingly being combined with this to make large data sets (UN Global Pulse, 2012). As an increasing number of groups involved in crisis management come to the realisation that value can be gained from engaging with big 'crisis' data, understanding how to use this data in an ethically and legally responsible manner is essential. Consequently, within this paper we explain how the BYTE project

(The Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities), funded by the European Commission, aims (in part) to do just that.

Briefly, BYTE is a multidisciplinary study that seeks to identify possibilities for capturing the positive externalities and diminishing the negative externalities associated with big data in order to support European innovation. In order to reach this aim, project partners are conducting exploratory research into the economic, legal, social, ethical and political externalities that are in evidence across different sectors. The research methodology combines desk-based research and primary research focusing on seven sectors with different big data contexts. In addition to crisis management, the sectors include: environmental, smart cities, cultural, energy, health and transport data. Alongside examining current practices, BYTE also uses foresight analysis to consider how big data will develop to the year 2020. Combined, BYTE will use these findings to develop a research and policy roadmap that provides incremental steps necessary to support industry, scientists, practitioners and members of the public to address these externalities. One of the key aims of BYTE is to assist stakeholders in improving their position with respect to innovation and socio-economic gains in Europe, and establishing a socially and ethically sound big data ecosystem is crucial to this endeavour.

Within the current working paper, and of relevance to the use of IT in preparing for, responding to and recovering from crises, our focus is on presenting an initial view of some of the externalities, also referred to as impacts, that we expect to emerge from a case study on the use of social media and GIS data to produce "crisis maps". While the case study will examine a number of externalities, this paper is focused on the following potential impacts: improved relief delivery, increasing public trust, violating privacy and introducing sampling bias. The paper will conclude by discussing how the BYTE project aims to intervene to assist in capturing the benefits of big data whilst addressing the potential negative impacts First however, the following section provides further information on the rationale behind the focus of the case study on crisis data and the intended methodology for this analysis.

BYTE: CRISIS INFORMATICS CASE STUDY

Understanding potential externalities associated with the potential uses of big data is essential. Within BYTE, the following definitions of externalities have been adopted: Positive externalities occur when a product, activity or decision by an actor causes positive effects or benefits realised by a third party resulting from a transaction in which they had no direct involvement. Negative externalities occur when a product, activity or decision by an actor causes costs (or harm) that is not entirely born by that actor but that affects a third party, e.g., citizens (Business Dictionary, 2014).

Within crisis management, the growing popularity of integrating social media and (other) digital content (e.g., photographs, videos and satellite imagery) provides a unique source of information for crisis managers, humanitarian organisations and others involved in responding to a crisis. In the last five years we have seen countless examples of studies and practitioners engaging with data stemming from social media, in combination with crowdsourcing activities and satellite data (among other sources) in order to support response and recovery efforts. Examples include (but are not limited to): an examination of the supporting of post-event reconnaissance via the collection and re-use of data from social media (and other sources) following the September 2014 Colorado floods (Dashti et al., 2014); the crowdsourced activities concerning the collection, sharing of video, image and text data following the 2013 Boston marathon attacks (Tapia, LaLone and Kim, 2014) and an examination of the role of Facebook and Twitter for crisis communication during the 2011 South East Queensland Floods (Bruns et al., 2012).

As time progresses and crisis managers, humanitarian organisations and others' engagement and activities and interaction with big 'crisis' data grows, it will be necessary to implement appropriate measures and policies to manage, protect and optimise the value that can come from engaging with this data. These data are generation and processing activities are dynamic and disruptive, and as such, a

thorough understanding of the externalities associated with big 'crisis' data, although desirable, is difficult to circumscribe.

BYTE partners are working with the Qatar Computing Research Institute (QCRI), Crisis Mappers, humanitarian organisations and other groups in order to identify and understand the impacts associated with the automated processing of big 'crisis' data. Innovation tools used by QCRI and others combines human computing (crowd sourcing) and machine computing (artificial intelligence) to evaluate citizens' needs during or immediately after crises. This includes mining "open" social media data, including text feeds, images, videos, location and temporal information to gather information, identify needs and assess damage. The case study enables BYTE to examine issues related to economic and innovation impacts associated with open source software for big data analytics, political issues around cross-border data collection and legal issues, including intellectual property and privacy issues related to the use social media data and the validity of high velocity data. The research questions guiding the examination include:

1. Which positive and negative societal externalities are associated with the use of big data in crisis management?
2. Who are the (positively and negatively) affected parties?
3. How might potential positive impacts captured, and how might challenges associated with negative impacts addressed?

For BYTE, externalities are related to processes (i.e., production, service, use) and not to the product itself. That is, it is not big data per se that causes a particular externality, but rather, it is the social processes employed via big data that can produce externalities. Furthermore, these externalities may result from the direct collection or processing of data (e.g., privacy infringements), as well as the opportunities and risks that may arise as a result of the existence of the data (e.g., linking data sets). In addition, as externalities may have unexpected effects on third parties, a central task in BYTE is the identification of the involved processes, their effects as well as the potential affected parties.

In order to conduct this examination, the case study involves some preliminary desk-based research to develop an understanding of the overall context, technological setting and the processes involved in collecting, storing and using big 'crisis' data. Subsequently, we will organise and run a series of in-depth semi-structured interviews and focus groups to examine the externalities associated with the capturing and use of big 'crisis' data. Whilst this study has not yet fully begun, in preparation for the case study, in the section below we discuss some of the potential impacts that we expect to examine.

POTENTIAL EXTERNALITIES

The use of automated software tools for mining social media content will have both positive and negative potential impacts on crisis response. While the potential positive impacts will certainly have constructive consequences for authorities, humanitarian organisations, relief workers and members of the public, these groups will have to devise ways of managing the potential negative impacts. In either case, these specific externalities result from the intersection between data mining and social media, and may be compounded by their interaction. This section examines a selection of potential positive and negative impacts associated with the mining of social media data for disaster relief within this case study. Beyond these, additional impacts certainly exist, and the significance of specific impacts may augment and diminish in particular contexts.

One instance where the intersection of social media and data mining might produce positive impacts is in relation to increasing citizen protection and trust through more interactive and responsive relief actions. The use of citizen-produced information by governments, NGOs and relief organisations provides the opportunity to respond to needs communicated directly by citizens, which can originate from areas that are difficult to reach and survey for damage after specific types of disasters. For instance, following Typhoon Yolanda (Haiyan) in the Philippines in 2013, the Standby Task Force and GIS Corps (members of the Digital Humanitarian Network) were mobilised in order to collate data in the form of images from social media in order to support UN OCHA in assessing the damage caused by the storm (Pierson, 2015). While the mining of social media data would not increase the rate at which regions become accessible, it might enable relief efforts to be planned in advance and mobilised immediately. In addition, the "open source" nature of these tools can also assist in providing better

citizen protection and security by enabling these tools to be widely accessible to response organisations (Huijboom, 2009). The combination of human and machine learning, especially to evaluate the veracity of information, can assist in building trust between those who provide relief and those who provide data, where relief organisations can trust the data coming in. Finally, citizens providing data may report increased trust that the government and other agencies are addressing the needs communicated by citizens into its plans as it is based on the automated processing of information rather than subjective decisions by organisations.

Social media also provides a rich and detailed information resource. The data being inputted by individuals can include moving and still images, location information, temporal information descriptions of needs as well as other information. Combining this rich information with automated processing capabilities can result in an increased efficiency in the planning and use of resources. For example, the automated processing of image information, especially of landmarks and other significant buildings and sites, can contribute to examining the overall levels of damage in particular locations. These damage levels can then be compared with other locations to identify areas where search and rescue or relief efforts need to be prioritised. It may also enable an identification of what types of resources are necessary. For example, damage to roads might be associated with damage to the water supply, indicating that drinking water needs to be provided.

Yet these potential positive impacts of mining social media information also raise potential negative impacts. Privacy and data protection are often-discussed issues in relation to social media, and the mining of social media and other citizen-produced data on the Internet can lead to a range of privacy and data protection infringements (within the use of social media for crisis management, see for instance; Palen et al. 2010).

Social media data certainly includes personal data (data relating to an identified or identifiable individual), and in most countries and regions the processing of such data is subject to a number of controls and obligations.[1] Whilst most jurisdictions allow for exceptions to data protection obligations in moments of crisis or disaster, this is founded upon a "trade-off" model of privacy and security, whereby individuals are expected to "trade" some rights (privacy and data protection) for others (security) (Lieshout, et al., 2012). Instead, a solution that respects both privacy and security ought to be the aim. In this case study, privacy and data protection issues are partially addressed by using open social media data (i.e., data from Twitters API) volunteered by citizens and by automatically processing that data to provide aggregated information. Yet, social media users are not asked to consent to the collection, storage and processing of this information by governments and other powerful organisations for this purpose, despite their rights to privacy and to the protection of their personal data. This could lead to further, compounding impacts. Specifically, the collection and processing of social media data by governments and other authorities could lead to a situation where individuals become concerned that the government has access to their social media data and is using it outside of emergency or crisis scenarios. This may discourage individuals from interacting with the government or organisations deemed to be linked with the government, which could undermine some of the potential positive impacts above, especially those related to fostering trust between citizens and governments.

Another potential negative externality is the possibility that sampling bias has been introduced into data processing activities. While the data coming in from social media channels is certainly "big data" by virtue of it being high velocity, high volume and taking a number of different forms (variety), big data is often mistakenly associated with representativeness (boyd and Crawford, 2012). Instead, data coming from social media users is always inherently biased towards young, educated and physically able populations (White and Selwyn, 2013; TNS Opinion and Social, 2011). These groups are better able to communicate their needs via social media, and relying upon automated processing of this information – e.g., to create maps or other data representations– could result in a

[1] For example in the European Union, under the Data Protection Directive, processing of personal data must be processed under one of the lawful conditions of processing, and data subjects have specific rights related to that data.

disproportionate provision of aid to some locations and populations. Whilst most aid workers and governments are already well aware of the need to consider the local context and the differing needs of vulnerable populations during crises, there is a continuing requirement to recognise, manage and implement appropriate measures into action (Telford and Cosgrave, 2007). As such, against this backdrop the use of these tools and automated processes might further remove vulnerable groups and individuals from relief efforts, as including those who do not use social media would require extra work outside the automated processes.

CONCLUSION

This preliminary examination of some potential positive and negative externalities associated with big 'crisis' data demonstrates that like all technology interventions, the processing of crisis data such as that stemming from social media needs to be continually assessed as part of a larger set of legal, ethical and social processes. As part of the empirical case study research, the BYTE team will continue to examine how the implementation of these tools has occurred in practice, and how humanitarian organisations have attempted to capture these benefits whilst minimising the potential negative impacts. The first output of this BYTE examination will be a series of research needs to track the potential positive and negative impacts of big crisis data and implement a more socially responsible system. The second output will be a series of policy steps that need to be implemented to enable European practitioners to take better advantage of the opportunities offered by big data and increase their competitiveness in relation to other countries and regions. Both of these outputs will be relevant for big crisis data and will assist in addressing some of the impacts outlined in this paper.

ACKNOWLEDGMENTS

This paper is based on research emanating from the European Commission-funded BYTE project (The Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities) under grant agreement number 619551. The views in this paper are those of the authors and are in no way intended to reflect those of the European Commission.

REFERENCES

1. Australian Department of Finance and Deregulation (March 2013) *Big Data Strategy — Issues Paper*, Australian Government.
2. Beyer, M. A., Lapkin, A., Gall, N., Feinberg, D., and Sribar, V.T. (2011) *'Big Data' Is Only the Beginning of Extreme Information Management*, Gartner Inc., Stamford CT. https://www.gartner.com/doc/1622715/big-data-beginning-extreme-information [Accessed: 15 January 2015].
3. boyd, d., and Crawford, K. (2012) "Critical questions for big data", *Information, Communication & Society*, 15 (5), 662-679.
4. Business Dictionary (2014), *Web Finance Inc*. http://www.businessdictionary.com [Accessed: 15 January 2015].
5. Bruns, A., Burges, J.E., Crawford, K., and Shaw, F. (2012) *#qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods - Research Report*. ARC Centre of Excellence for Creative Industries and Innovation, Queensland University of Technology, Brisbane QLD Australia. http://eprints.qut.edu.au/48241/
6. Dashti, S., Palen, L., Heris, M., Anderson, K., Anderson, S., and Anderson J.T. (2014) "Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods". In: *Proceedings of the 1th International ISCRAM Conference*, Pennsylvania, USA.
7. Huijboom, N., van den Broek, T., Frissen, V., Kool, L., Kotterink, B., Meyerhoff Nielsen, M. and Millard, J. (2009) *Public Services 2.0: The Impact of Social Computing on Public Services*, JRC European Commission, EUR24080.
8. Laney, D. (2001) "3-D Data Management: Controlling Data Volume, Velocity and Variety". http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [Accessed 26 January 2015].
9. Networked European Software and Service Initiative. (December 2012) *Big Data White Paper*. http://www.nessi-europe.com/Files/Private/NESSI_WhitePaper_BigData.pdf [Accessed 26 January 2015].

10. Palen, L., Anderson, K.M., Mark, G., Martin, J., Sicker, D., Palmer, M., and Grunwald, D. (2010) "A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters". *British Computer Society*, 8 (1), pp. 8-12.
11. Pierson, L. (2015) "Using reliability engineering to ensure the success of digital humanitarian (technology) developments – here's how", *Data-mania.* http://www.data-mania.com/blog/using-reliability-engineering-ensure-success-digital-humanitarian-technology-deployments-heres/ Accessed 26 January 2015].
12. Tapia, A., LaLone, N. and Kim, H-W. (2014) "Run Amok: Group Crowd Participation in Identifying the Bomb and Bomber from the Boston Marathon Bombing". In *Proceedings of the 11th International ISCRAM Conference*, Pennsylvania, USA.
13. Telford, J. and Cosgrave, J. (2007) "The international humanitarian system and the 2004 Indian Ocean earthquake and tsunamis", *Disasters*, 31(1), pp.1-28.
14. TNS Opinion and Social. (2011) *Special Eurobarometer 359: Attitudes on Data Protection and Electronic Identity in the European Union*, Special Eurobarometer, European Commission. http://ec.europa.eu/public_opinion/archives/eb_special_359_340_en.htm [Accessed 26 January 2015]
15. UN Global Pulse. (May 2012) *Big Data for Development: Challenges & Opportunities*, United Nations, New York.
16. van Lieshout, M., Friedewald, M., Wright, C. and Gutwirth, S. (2012) Reconciling privacy and security, *Innovation: The European Journal of Social Science Research,* 26 (1-2).
17. White, P. and Selwyn, N. (2013) "Moving on-line? An analysis of patterns of adult internet use in the UK, 2002–2010", *Information, Communication & Society*, 16 (1), 1-27.