ORIGINAL ARTICLE

CrossMark

Exploring the use of crowdsourced geographic information in defence: challenges and opportunities

Nikolaos Papapesios[1] · Claire Ellul[1] · Amanda Shakir[2] · Glen Hart[2]

Abstract

Geographic data are used by United Kingdom (UK) defence for purposes including peacekeeping, humanitarian aid and disaster relief, and fighting wars. The geographic extent of defence data covers the world, with greater focus directed towards areas considered to be of current interest. Traditionally, these data have been officially sourced, e.g. via National Mapping Agencies, but there is now increasing interest in the potential of crowdsourced geographic data to supplement authoritative data where they are not available, outdated or incomplete. Volunteered geographic information (VGI) and social media have the potential to provide this needed missing information. This paper presents initial work carried out in identifying the potential of crowdsourced geographic information in defence. We first provide a short description of the role of UK defence and review the existing literature on crowdsourced geographic information in defence, as well as generic VGI quality assessment methods. We then explore the potential of crowdsourced data in real-world applications: the conflation of VGI and social media with official data for effective decision-making in war zones, and the potential for crowdsourcing to increase effective collaboration between machines and humans in disaster situations. Based on our review, we outline specific research challenges for deploying crowdsourced geographic information in defence, focussing on data quality and fitness-for-purpose assessment. Defence-specific constraints include the need for rapid quality assessment processes and the need to communicate high-quality information effectively in situations where rapid decision-making is required. Ethical issues are also of fundamental importance.

Keywords VGI · Defence · Situational awareness · Crowdsourcing · Social media · Data usability

✉ Claire Ellul
  c.ellul@ucl.ac.uk

[1] Department of Civil, Environmental and Geomatic Engineering, University College London (UCL), Gower Street, London WC1E 6BT, UK

[2] Cyber and IS Division, Defence Science and Technology Laboratory, Porton Down, Salisbury, Wiltshire SP4 0JQ, UK

🙋 Springer

Abbreviations

DoD	Department of Defense (US)
DSTL	Defence Science and Technology Laboratory
DARPA	Defense Advanced Research Project Agency (US)
GIS	Geographic information systems
HACs	Human-agent collectives
JFC	Joint Forces Command (UK)
MOD	Ministry of Defence (UK)
OSM	OpenStreetMap
RN	Royal Navy (UK)
ISO	International Organization for Standardisation
iVGI	Involuntary geographic information
SDSR	Strategic Defence and Security Review
Social media	Collaborative and social computing
UAVs	Unmanned aerial vehicles
UGVs	Unmanned ground vehicles
VGI	Volunteered geographic information

1 Introduction

"The next war will be won in the future, not the past. We must go on, or we will go under (Mayfield 2011)."
– General of the United States (US) Army Douglas MacArthur, 1931 –

In an age where a considerable amount of geographic data are characterised as *big,* and the volume of data is predicted to continue to increase, a key question has emerged regarding how it can best be managed, collected and analysed. The availability of faster data collection methods, in combination with technological advances including Web 2.0, and low-cost tools such as Global Positioning System (GPS) devices and smartphones, enables the public to participate in the data collection process (Goodchild 2007; Girres and Touya 2010; Haklay 2010; Ali and Schmid 2014; Forghani and Delavar 2014; Neis and Zielstra 2014). The involvement of amateurs, individuals and volunteers in crowdsourcing, which was recently described as "*a type of participative online activity in which an individual, an institution, a non-profit organisation or a company proposes to a group of individuals of varying knowledge, heterogeneity, and number,* via *a flexible open call, the voluntary undertaking of a task*" (Estellés Arolas and González Ladrón-de-Guevara 2012, p. 197), generates location-based social networking and collaborative mapping (Franklin et al. 2013; Shanley et al. 2013). This new era, involving a set of geographic information systems (GIS) techniques and tools available to public and non-expert users, is characterised as Neogeography (Haklay et al. 2008) Neogeography as a concept appeared in 2006 and is defined by Turner (2006, p.3) as "*the sharing of location information with friends and visitors, helping shape context, and conveying understanding through knowledge of place*".

The spatial information produced by volunteers, usually characterised as "contributors" and occasionally as "prosumers" or "produsers", i.e. people who both produce and use data, is often referred to as volunteered geographic information (VGI). VGI is defined by Goodchild (2007, p.211) as the use of the web to "*create, assemble, and disseminate geographic data provided voluntarily by individuals*" (see Sect. 3.1 for a more detailed explanation of the terminology). VGI is a part of the Neogeography concept, together with techniques and tools such as Web 2.0, mashups, citizen science and public participation (Heipke 2010), and is considered as an inexpensive, up-to-date and plentiful source of data (Coleman et al. 2009; Fonte et al. 2015). It can be used as either a single source or to enrich and update authoritative data and products (Antoniou 2011). In particular, social media is considered an important source of these data, as it can rapidly inform users about what is happening in the world at any time and place. In the United States (USA), for instance, social media use rose sharply from 5% of internet users in 2005 to 70% in 2015 (Perrin 2015).The availability of georeferenced data through VGI and social media services and platforms such as OpenStreetMap (OSM),[1] Wikimapia and Twitter increases the potential of these data to be used in applications relating to areas as varied as disaster management, citizen science, pedestrian navigation and national security. The wide-ranging use of VGI in many domains also reveals its biggest challenge, i.e. the evaluation of its quality (Flanagin and Metzger 2008; Haklay 2010; Antoniou and Skopeliti 2015).

Defence, which is the potential VGI application area forming the focus of this paper, collects information from a wide range of official (authoritative) sources in order to inform decisions. The "geographic" aspect is a key component of this information given the importance of geography to defence and the worldwide remit of many defence forces (e.g. see Sect. 2 for UK Defence Roles). However, during emergency situations, or on-the-ground warfare, authoritative data are often ineffective or inadequate, particularly in areas of the world that are not well mapped or where the situation on the ground is rapidly changing. This paper examines the potential of crowdsourced geographic information (comprised of both volunteered and involuntary geographic information, see Sect. 3.1) to fill this gap, assessing current and future uses of these data in a defence context. While many challenges in this domain relate to the development of analytical and big data methodology and algorithms, we focus in particular on data quality. Understanding the quality and hence the fitness-for-purpose of a crowdsourced data set is fundamental when assessing the potential of crowdsourced geographic information as a data source for decision-making in defence. Based on our review, we propose a number of research challenges to enable crowdsourced data to fill the defence data gap.

The remainder of this paper is structured as follows: Sect. 2 describes the role of the United Kingdom (UK) Ministry of Defence (MOD), its general structure and military tasks, with the UK being selected as the main focus of this work due to

---

[1] OpenStreetMap (OSM) is one of the most popular examples of a VGI project. It is a collaborative mapping project founded in London in 2004 where citizens create freely accessible spatial data—in particular road networks and buildings—which can present an alternative to official data (Dorn et al. 2015).

the opportunity to (i) provide a clear understanding of the roles of a defence force and (ii) identify specific opportunities for crowdsourced data based on current and emerging projects, via project partners such as DSTL. DSTL is one of the MOD's executive agencies responsible for the development and application of Science and Technology across the MOD and within the context of UK security. However, this work is also applicable to other countries around the world. Section 3 provides a literature review of crowdsourcing related to military tasks, focussing in particular on national security and disaster response. Additionally, a review of current spatial data quality assessment methods is given. Section 4 presents two emerging UK defence applications, demonstrating the potential of crowdsourced geographic information, and therefore VGI, to improve situational awareness during a defence operation. Section 5 discusses the findings and the potential challenges and opportunities of using crowdsourced and voluntary geographic data in defence for the next decade, in particular relating to data quality assessment and fitness for purpose. The paper concludes with a summary of the findings and the main objectives for future work. Given the fact that many readers may not be familiar with defence terminology, we provide a short list of key terms at the end of the paper.

2 The roles of UK defence[2]

To provide a concrete focal point for the analysis in this paper, we first present a summary of the broad remit of a typical defence organisation, with the UK taken as an exemplar. The UK government department responsible for defence policy is the MOD. The security, independence and interests of the UK at home and in overseas territories are protected by the MOD, where the main objective is to ensure that the armed forces have the necessary training, equipment and support to fulfil their duties (UK HM Government 2015). It has both permanent and casual civilian personnel and works with 28 agencies and public bodies (Government Digital Service 2017). The Royal Navy (RN), British Army and the Royal Air Force constitute the UK regular forces, which are usually referred to as the British Armed Forces (Ministry of Defence UK 2015a, b; UK HM Government 2015).

Defence is always at a turning point, and the future, with its challenges, is difficult to predict. Nevertheless, every 5 years the MOD presents the Strategic Defence and Security Review (SDSR), which includes a vision for the future, the main threats and the required capabilities to address them (Government Digital Service 2017). The SDSR gives the UK government the information to understand the strategy and find a balance between the available resources, the MOD's national policy ambition, and real-world commitments. In the 2015 SDSR (SDSR 15), the MOD's defence policy was defined around eight missions presented in Table 1 (UK HM Government 2015), which update the seven military tasks (MTs) described in the 2010 SDSR (Brooke-Holland and Mills 2015).

---

[2] The information presented here has been adapted from information provided on the UK Ministry of Defence website http://www.mod.gov.uk © Crown Copyright 2017.

Table 1 The 8 missions defined by the MOD in the 2015 SDSR. The first four are the basic missions, while the latter four are more relevant in times of crisis

No.	Definition of missions	Brief description of missions
1	Defend and contribute to the security and resilience of the UK and Overseas Territories	*Usually British Armed Forces missions including deterring attacks*: defending UK airspace, territorial waters and cyber space; countering terrorism at home and abroad; supporting the UK civil authorities in strengthening resilience; and protecting people overseas
2	Provide a nuclear deterrent	
3	Contribute to improved understanding of the world through strategic intelligence and the global defence network	UK's independent nuclear deterrent will remain essential to security today, and for as long as the global security situation demands. This mission includes contributing to improved understanding of the world through strategic intelligence and the global defence network, in particular close and enduring work with UK's allies and partners during peace and conflict
4	Reinforce international security and the collective capacity of our allies, partners and multilateral institutions	Including work to help shape the international security environment, and to strengthen rule-based international order including through conflict prevention, capacity building and counter proliferation
5	Support humanitarian assistance and disaster response, and conduct rescue missions	*British Armed Forces missions in times of crisis including*: major natural disasters overseas can put at risk the safety and security of British nationals, the Critical National Infrastructure and the supply chains such as energy, food and technology
6	Conduct strike operations	
7	Conduct operations to restore peace and stability	Working together with USA to support peace and stability in Europe's neighbourhood, the Middle East, Africa and Central Asia
8	Conduct major combat operations if required, including under NATO Article 5	A war-fighting division optimised for high intensity combat operations
		Article 5 states that if a NATO ally is the victim of an armed attack, each and every other member of the Alliance will consider this act of violence as an armed attack against all members and will take the actions it deems necessary to assist the ally attacked

By the end of 2025, plans assume a deployed force of around 50,000 drawn from a Maritime Task Group, an Army Division, an Expeditionary Air Group and the Joint Forces Command (JFC) (Ministry of Defence UK 2015a, b; UK HM Government 2015). The JFC provides the foundation and supports a framework for successful operations by ensuring the development and management of joint capabilities such as medical services, training, intelligence and cyber operations. Intelligence and understanding are amongst the advanced capabilities that need to be explored and managed. New technologies such as big data and open-source intelligence will improve the JFC's understanding of the world, allowing it to increase the speed and agility of military response (Ministry of Defence UK 2015b).

3 Literature review

In recent decades, countries and societies have been threatened by terrorism, cyber attacks and warfare. Government departments in charge of defence and national security are, therefore, searching for techniques to improve their data sets, and hence their ability to respond to or even prevent these activities, with up-to-date or even real-time information to obtain a better and faster understanding of critical situations. It is believed that crowdsourcing can be a key to improve defence decision-making (Greengard 2011; Parsons 2011; Franklin et al. 2013; Shanley et al. 2013).

This section presents an overview of previous work linking crowdsourcing, VGI and georeferenced social media data with defence cases, organised around some of the tasks relevant to UK defence (see Sect. 2). In particular, previous work highlights that VGI and open data can prove effective in civil emergency cases and increase the understanding of new environments and the strategic intelligence of the operations. Additionally, the concept of data quality is analysed, and previous work to assess data quality of crowdsourced data is summarised in order to underpin the subsequent assessment of how the fitness-for-purpose of crowdsourced data can be assessed in a defence context.

3.1 Clarification of crowdsourcing terminology

Crowdsourcing, as has been previously mentioned (see Sect. 1), is an open participative activity, where anyone can propose a specific task of data collection to a number of people while Goodchild (2007) introduced VGI as an "*umbrella*" term for the geographic information being created by everyday users, privately and voluntarily. More broadly, from a sociological perspective, collaborative and social computing (social media) (Roy et al. 2017) can rapidly provide up-to-date information, which in some cases is also geolocated. The real-time data produced by using social media platforms such as Twitter and Flickr can provide georeferenced data captured by individuals, although this is not specifically "volunteered" (See et al. 2016). Therefore, the data acquisition process to cover this type of VGI has been defined by Fischer as "*involuntary geographic information*" (iVGI) and can be also used for various activities (Fischer 2012). Both VGI and iVGI are considered in this paper.

As noted above, the availability of georeferenced data through VGI and iVGI services and platforms increases interest in this field for defence domains. Many authors in defence use crowdsourcing as the chosen term to describe the phenomenon of voluntary and involuntary citizen participation, and this paper follows the convention used in the defence context. We thus use the term "Crowdsourced Geographic Information" (See et al. 2016) to include both VGI and iVGI. As Fig. 1 shows, crowdsourcing is the main term used to characterise both spatial and aspatial data produced by passive and active users in participation activities, while the spatial data produced by active users or volunteers are referred to as VGI while the data produced by social media users (or passive users) are referred to as iVGI.
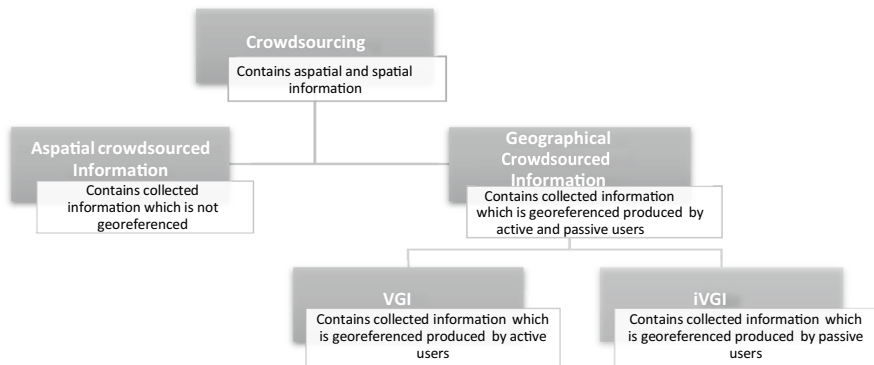
Fig. 1 Crowdsourced information can contain georeferenced and non-georeferenced information. This georeferenced information is referred to as crowdsourced geographic information and contains spatial crowdsourced information that is produced either by volunteers (VGI) or by social media or other users such as satellite navigation systems (iVGI). This figure has been compiled by the authors based on the review in Sect. 3.1

3.2 Crowdsourcing in defence

An extensive review has highlighted the fact that there is relatively little literature linking crowdsourcing and defence, perhaps due both to the relative novelty of crowdsourced data and to the fact that much defence research is, of necessity, confidential. The published literature has focussed on strategies to improve strategic intelligence, defend national territory, and the potential for crowdsourcing to provide additional information in times of crisis or disaster.

3.2.1 Crowdsourcing in defence and warfare

Many Western departments of defence have national defence and warfare preparedness as their primary objective. For instance, the first UK MOD task refers to the defence and contribution of the security and the resilience of the UK and overseas territories (see Sect. 2). War-fighting preparedness is also the first mission of the United States (US) Department of Defense (DoD), which has long understood the importance of GIS and highly accurate data for use during warfare operations (Franklin et al. 2013). Franklin et al. (2013) noted that data are now frequently co-produced, involving non-traditional stakeholders (non-DoD GIS professionals, Information Technology professionals and daily GIS users), and three partner types: clients, citizens and volunteers. They also explore the use of volunteered information for military operations and noted that these multiple producers have resulted in the US DoD increasingly facing challenges relating to data integrity and data security.

To build on this work, the US Defense Advanced Research Project Agency (DARPA) investigated various strategies by which they can improve their intelligence information. The importance of crowdsourcing in this context was examined via a funded challenge, i.e. the "DARPA Network Challenge", during which solutions and techniques were proposed by groups and individuals, and crowdsourcing

issues were examined (Greengard 2011; Hui 2015). The prize was either money or recruitment possibilities. The results showed that the intelligence of the agency can be improved by many groups, especially university teams but also individual experts. Motivational issues have also been investigated by DARPA during the "Red Balloon Challenge" (Hui 2015), which offered money as a reward for location corrections. The winning team solved the challenge by using crowdsourcing through social media.

Crowdsourcing has also been used to support warfare in an indirect manner. The US State Department organised a challenge to look for ideas of how crowdsourcing can support arms control transparency, giving as a prize, an amount of money (Hui 2015). The winner, by using visible light communications, improved arms control inspections.

3.2.2 Crowdsourcing for homeland security

Addressing challenges closer to home than traditional warfare, the Texan government in the USA has created the "Texas Virtual BorderWatch" (Tewksbury 2012; Hui 2015). As Texas is on the border with Mexico, numerous migrants try to cross the border every day, without authorisation. Due to illegal immigration and drug smuggling issues, additional techniques to improve homeland security have been investigated (Tewksbury 2012). A network of web-based surveillance cameras was created, and the people, transformed to "interactive citizen-soldiers" (Andrejevic 2006), reported any suspicious activities in inaccessible zones such as in rivers or in wooded areas. Although the immobility of the cameras was identified as one of the technical issues during this process, this was an effort where the awareness against criminality was improved by using crowdsourcing techniques (Tewksbury 2012). In a similar move, the US Department of Homeland Security (DHS) created the Neighbourhood Network Watch program to collect reports of suspicious online criminal activities (Hui 2015).

Addressing another homeland security issue, i.e. civilian unrest in 2011 in England, where thousands of people rioted as a result of a citizen's fatal shooting by police, the authorities used a combination of social media platforms (Flickr, Facewatch ID and Twitter) to recognise the offenders' faces (Suleyman 2017). Thousands of photographs were shared via Flickr and sorted according to postal codes, and the authorities were informed of any recognised rioter's face via Facewatch ID. As a final point, a hashtag was used by authorities to collect all the needed information regarding the looters.

3.2.3 Use of social media to increase awareness and assist short-term predictions in a crisis

It is important during military operations, where the environment is usually aggressive and dangerous, to decrease the level of unknowns, and with the diffusion of terrorism and the expansion of military operations around the world, the analysis of crowdsourced data, especially social media data, are often the only easily available source of information. Thus, social media can reduce the level of uncertainty in new

environments and also increase knowledge within areas of command responsibility (Mayfield 2011). As a consequence, efficient use of social media data may lead to a better understanding during crisis situations, in collaboration with allies, partners and multilateral institutions (see Sect. 2) for peacekeeping purposes.

Social media can improve situational awareness by examining not only the environment as a whole, but also specific target users, and a number of commercial tools have been developed to support this task. Rapid Information Overlay Technology is an extreme-scale analytics system with defence objectives and can "spy" on a user's activities and habits (Gallagher 2013). The data can be mined from the most popular social media platforms such as Facebook and Twitter, and statistics of the user's daily habits can be collected and predicted. Another example is Wikistrat, which is an analytical services consultancy that makes use of crowdsourcing to improve client awareness, and has provided predictions for a military operation for US Africa Command and a prediction of activities of the Islamic State (IS) in Middle East by analysing social media data (Hui 2015).

Having as an aim to improve the accuracy of short-term and middle-term events, the Aggregative Contingent Estimation System (ACES) was a predictive tool that used the answers of participants, who were asked questions relating to several scenarios, to assess the opinions of the crowd (Parsons 2011). Some of the scenarios investigated included the decision of the US military to repeal "don't ask don't tell" regulations (with regard to LGBT personnel) and the resignation of Yemen's president Ali Abdullah Saleh. The ACES tool, with the assistance of 1800 participants, correctly predicted both results.

3.2.4 Real versus fake news

In 2006, Hezbollah forces posted a number of videos and photographs through social media to "promote" the war against Israeli forces (Mayfield 2011). Using this technique, the Islamist militant group and political party of Lebanon was able to influence the national troops and citizens and managed to win an unexpected and difficult battle. Moreover, in 2009, an innocent woman was shot during the protest gatherings of the Iranian presidential election, and through social media, several versions of the video that captured the incident, some of them fake, were posted online, creating a wave of global reactions and a negative climate that took the Iranian government weeks to control. These two examples demonstrate that social media can be used for both positive and negative manipulation of the crowd during warfare or a national crisis. Social media is open to all and can include terrorist websites, which could be used by extremists and terrorists (Hope and McCann 2017). Crowdsourced information, and especially social media, does not always have a positive impact because criminal activities can exploit it for their purposes (Johnson 2014).

Taking advantage of the effect that social media can have on the crowd, terrorist organisations try to promote their objectives and manipulate social media users by spreading fake news via well-known platforms. According to the RAND Corporation, after analysing 23 million tweets posted in Arabic by 771,327 Twitter users, it was discovered that ISIS supporters produced approximately 50% more Tweets per day than their opponents in the period July 2014 to April 2015 (Bodine-Baron et al.

2016). This is one of the first examples of a terrorist organisation that has successfully used a social media platform to promote its message, stimulate and recruit new fighters and spread its propaganda. This has driven the US DoD to explore new techniques to decrease ISIS influence, and Twitter to continue its campaign of account suspensions. The research showed that the number of supporters was reduced as a result of Twitter's account suspensions.

More recently, additional attention has been paid to the growing amount of fake news in the media. For example, the recent story of the girl trapped in the catastrophic earthquake in Mexico, which was based on the evidence of a witness, and where the incident was reproduced multiple times in social media. This resulted in the Mexican aid forces spending hours of efforts on the rescue before realising that they did not have any valid proof that the incident was real (Associated Press 2017).

3.2.5 Crowdsourcing and cyber attacks

Cyber space is a rather new field where crowdsourcing is also relevant (Hui 2015). Defending and securing cyber space is one of the UK MOD's missions (see Sect. 2). Johnson (2014) has explored a number of cyber attacks (Estonia-2007, Belarus-2008, Lithuania-2008, Georgia-2008, China-2009, W32.STUXNET-2010) where the criminals shared malware through social media and networks, investigating the threats to human–machine interaction in multi-layered networks. After explaining the difficulties that commercial and government organisations have in predicting the potential groups that take part in cyber attacks, he identified four ways in which social media could be linked to cyber attacks: social networks motivate participation through crowdsourcing; social networks are used to target individuals via phishing; "disposable service models" associated with social networks aid coordination of an attack; use of anti-social networks supports botnets and associated criminal infrastructure) where social networking can block the attribution of cyber attacks.

Research regarding the advantages of crowdsourcing for cyber security was completed by the US DHS and the Center for Risk and Economic Analysis of Terrorism Events (CREATE) (Hui 2015). It appeared that encouraging individuals and institutions to voluntarily cooperate to secure cyber space works effectively, and an example of a group of volunteer experts from various institutes and the China's Ministry of Information Industry is presented. By combining simple techniques, a challenging computer virus named "Conficker" was overcome and its creator identified (Hui 2015).

3.3 Crowdsourcing to support civil disasters and emergencies

Another field where crowdsourcing has already proved useful for military operations is humanitarian aid and disaster relief (Roy et al. 2017). Supporting humanitarian assistance and disaster response and conducting rescue missions in times of crisis falls under UK defence policy (see Sect. 2). When traditional methods and reference data are not available, when conditions are uncertain, or there is an excessive delay

until a data collection exercise can be initiated, then the required information can be collected via crowdsourcing and social media. Indeed, crowdsourcing is an efficient way to collect near real-time data in disaster response (Ortmann et al. 2011), especially when the population is vulnerable and crisis mapping is required (Shanley et al. 2013). Related terms such as "digital volunteerism" (Shanley et al. 2013) and VGI (Goodchild 2007; Filho 2013) are used in this context, highlighting the role that volunteers can play in improving situational awareness for rapid response to natural disasters and complex humanitarian emergencies (Shanley et al. 2013).

3.3.1 Use of VGI and iVGI in disaster response

Mills and Chen (2009)explain the benefits of using Twitter during emergency situations including low cost, ease of use, scalability, rapidity and the use of visualisation instead of more traditional methods such as messaging or other platforms such as Facebook. The wide range of Twitter use in disaster emergencies (fires, ice storms, earthquakes, hurricanes, cyclones) makes it a low-cost alternative for collecting geolocated information easily and rapidly. An alternative idea for implementing crowdsourced data in disaster response is the "Tweak the Tweet", where the platform asks users to create new hashtags during a disaster, making them more machine-readable (Finin et al. 2010).

Shanley et al. (2013) have presented examples of where volunteerism was used in disaster management and emergency response conditions, citing occasions where US organisations including the US Geological Survey and the US Federal Emergency Management Agency, used VGI when it was almost infeasible to collect data from authoritative sources. Different techniques were created by projects such as "Do-It-Yourself", "Did you feel it" and "Civil Air Patrol volunteers", where non-expert users collected information for natural disasters (earthquakes and hurricanes) via open-source map platforms such as OSM or via cameras (attached to balloons, kites and unmanned aerial vehicles—UAV) when there was no possibility to approach the area (Parsons 2011; Shanley et al. 2013).

Becker and Bendett (2015) also presented several examples (Camp Roberts, DCMO challenge, STAR-TIDES Network) of how crowdsourcing can be implemented in disaster response cases. The "TIDES" program, implemented by the National Defense University Center for Technology and National Security Policy and the US DoD, investigates the use of open-source knowledge to assist in cases such as disaster response where the population is under crisis (Becker and Bendett 2015), aiming to provide decision makers with the necessary knowledge. Providing end-users with needed information can be beneficial in civil emergency because human lives are under threat, and rapid and effective decisions need to be taken. The authors also note that VGI can provide close to real-time information, citing, in particular, the example of the Haitian earthquake in 2010, when 640 volunteers in the Humanitarian OSM Team (HOT) collected urgently needed cartographic information inexpensively and very rapidly.

It is essential that crowdsourced information can be collected as quickly as possible for disaster response. The research project "Evolution of Emergency Copernicus services" (E2mC) has created a new component called "Witness" for the Copernicus

Emergency Management Service (EMS), which can reduce the time needed to integrate crowdsourced data after a disaster event (Havas et al. 2017). Witness architecture includes data acquisition, storage, management and analysis, and graphical user interface components, which have been used in several recent disaster cases such as in the Central Italy earthquake (2016) and the Haiti hurricane (2016).

3.4 Data quality and crowdsourcing

From a producer's perspective, data quality refers to how well the data produced, for a specific purpose, conform to a representation or abstraction of the real world (Devillers and Jeansoulin 2006; Harding 2006). However, the end-user's perspective differs, and as crowdsourced data can be open data, they can end up being used for purposes beyond which they were originally collected for. Even so, it is important that the producer can understand as far as possible the end-user's needs (Harding 2006), and in order for any data set to be used appropriately, the user needs to have sufficient understanding of its quality, i.e. how "fit for purpose" the data are. This is particularly the case in defence, where decisions made both in warfare and in disaster management can directly impact human lives. Initial discussions with respect to the need to describe the data quality of spatial data first started towards the end of the 1980s (Goodchild and Gopal 1989), focussing specifically on the accuracy of the data (Chrisman 2006), with accuracy being split into geometric accuracy and semantic accuracy. When the need to measure the quality increased due to an increasing variety of data sources and data sets, it was realised that accuracy only forms one component of quality, and that issues such as currency, completeness and others are also relevant. A key challenge thus relates to the need to find a concept that can represent data quality overall.

3.4.1 General approaches to measuring data quality

Standards provide specific information about the quality of a data set, and this information is derived by measuring the quality of the data itself. Quality can be categorised as internal and external quality. Internal quality refers to the level of the similarity between the data that have been produced and that the data should have been produced, referred to as "perfect" data, while external quality refers to the level of agreement between the produced data and the user needs. External quality has more recently, and more correctly, been described as "fitness for use", "fitness for purpose" or "fitness for use for a certain purpose" (Devillers and Jeansoulin 2006; Dorn et al. 2015; Fonte et al. 2015).

Data quality assessments can also be extrinsic or intrinsic. When the measurement of a data set is assessed in comparison with the previously mentioned authoritative or reference data, the approach is extrinsic (Girres and Touya 2010; Haklay 2010; Helbich et al. 2012; Zielstra and Hochmair 2013; Barron et al. 2014). To evaluate quality using an intrinsic approach, only one data set is needed (Barron et al. 2014).

3.4.2 Measuring the quality of crowdsourced geographic information

As noted in Sect. 1, the technological evolution and low-cost tools enable public participation in the data capture process. However, as end-users of the resulting data are not aware of the accuracy level of the devices and/or of the expertise or the training level of the volunteers, the necessity for quality evaluation prior to use becomes even more important (Ali and Schmid 2014). Both intrinsic and extrinsic measures can be applied to crowdsourced geographic data, depending on the availability of "ground truth" data. While geometric evaluation is important, semantic inconsistency is also a major issue in VGI (Ali and Schmid 2014), e.g. a road in OSM can be marked as "key: highway", referring to one of the most important OSM road tags (Pourabdollah et al. 2013). However, there are other OSM values that can clarify the type of road such as "motorway", showing that the highway key is not always relevant.

Basing the analysis on available quality standards, a number of studies have compared VGI data sets with authoritative data sets (Ali and Schmid 2014; Antoniou and Skopeliti 2015). These assume that available authoritative data can be characterised as reference data, which means that the quality level of the authoritative data sets is high (Antoniou and Skopeliti 2015) and such data sets are assumed to be "correct".

Barron et al. (2014) have created a framework to evaluate VGI quality intrinsically by presenting various methods and indicators. Additional examples include Dorn, Törnros, and Zipf (2015), who present a new extrinsic land use quality approach; Barron, Neis and Zielstra (2014) who compared various VGI projects and platforms for measuring quality and presented a number of OSM trends; Fonte et al. (2015) who combined existing quality assessment methods for validating land cover maps; and Girres and Touya (2010) who evaluated the OSM quality intrinsically and extrinsically for a number of elements and indicators at a national level.

However, due to licensing issues and restrictions related to some authoritative data sets, it is not always possible to evaluate the VGI quality by comparison with authoritative sources (Barron et al. 2014; Antoniou and Skopeliti 2015). Therefore, alternative evaluation methods have been proposed using intrinsic analysis, with the new measures described as quality indicators (Antoniou and Skopeliti 2015). A list of these indicators was presented in Fonte et al. (2017), who also proposed indicators specific to VGI, expressing the view that either existing standards need to be updated and take into account new sources and reports, or new standards need to be developed. In addition, Degrossi et al. (2017) summarised 13 quality methods in a systematic literature review, which can be used for crowdsourced geographic sources when authoritative ones are not available. However, some of these are not suitable for social media, and the necessity for further research is indicated. Likewise, Senaratne et al. (2017) outlined 30 methods that can be used to assess the VGI quality of maps, images and text, proposing data mining as an autonomous approach for estimating VGI quality.

3.4.3 Standard approaches to documenting and communicating spatial data quality

Once assessed, the quality of a spatial data set can be documented in many ways, from an online webpage to a PDF document, with the main objective being to increase the level of confidence in the spatial data products. To facilitate interoperability and ease of comparison of data sets, a number of standardised metadata structures have also been created (Antoniou and Skopeliti 2015; Dorn et al. 2015). One of the most well-known standards organisations is the International Organization for Standardisation (ISO), and several principles and guidelines have been proposed by the ISO to assess data quality (Barron et al. 2014; Antoniou and Skopeliti 2015). ISO 19113 and ISO 19114 describe quality principles for geographic information (Van Exel et al. 2010; Barron et al. 2014; Neis and Zielstra 2014; Dorn et al. 2015). ISO 19157 provides a more updated approach and replaces the previous versions (Barron et al. 2014; Neis and Zielstra 2014; Antoniou and Skopeliti 2015). It defines a list of six quality elements: completeness, logical consistency, positional accuracy, temporal quality, thematic accuracy and usability (Barron et al. 2014; Antoniou and Skopeliti 2015). The first five elements are focused on the producer (internal quality), while the last one (usability) is focused on user needs (external quality) (Fonte et al. 2017). To measure data quality, ISO's principles and assessments assume that geographic data can be compared authoritative or reference data.

4 Deploying crowdsourced geographic information in UK defence: analysing the potential of crowdsourced data in real-world applications

This section aims to investigate the current progress and potential benefits of crowdsourced geographic information, and in particular VGI deployment, in defence. Two UK defence studies are presented, both relating to current research. In the first study, crowdsourced geographic information is explored from a theoretical perspective in an effort to increase the strategic intelligence of defence during a warfare mission. The second project relates to civil emergency organisations in times of crisis where VGI has been used.

4.1 Improving decision-making and situational awareness[3]

In Spring 2017, the Royal Navy, which is part of the British Armed Forces, in conjunction with DSTL, initiated the exercise *"Information Warrior"*, in order to test and develop their information warfare capabilities (BiP Solutions Limited 2017).

---

[3] The information presented here has been adapted from information provided on the UK Ministry of Defence website http://www.mod.gov.uk © Crown Copyright 2017.

The main objective is to achieve an advantageous position in comparison with their opponents during a conflict, and the project has a specific focus on how the information about what is happening at a particular location can change the course of a battle. The five main themes of the exercise are as follows (BiP Solutions Limited 2017; Navy 2017):

i.   Artificial Intelligence (AI)—by using AI technology, the Royal Navy can further develop fleet intelligence (BiP Solutions Limited 2017). Thus, the efficiency of the fleet will be increased, allowing autonomous, fast and complex decision-making.

ii. Command, Control, Communication and Computers—employing computer systems to integrate a comprehensive infrastructure to meet the challenges during peacetime and warfare. The operational capabilities developed through this computing technology and Unmanned Aerial Systems will increase the efficiency of the Royal Navy.

iii. Cyber, Electromagnetic and Space Activity—Cyber attacks are new threats that can cause several problems to ships. Cyber protection needs to be constant and comprehensive to avoid any damage.

iv. Intelligence, Surveillance and Reconnaissance—developing and using unmanned systems can decrease the risk to humans during a mission and enhance a commander's decisions.

v.   Intelligence Exploitation—combining multiple intelligence sources (Open-Source Intelligence, Satellite Imagery) to improve the intelligence picture and reduce the workload of data analysts. Concepts such as Adaptable Tactical Information eXploitation (ATIX)[4] and Every Platform A Sensor projects (EPAS) utilise data from all available sources and then combine and analyse them in order to improve a commander's understanding in complex situations and enable faster, informed decision-making.

These themes demonstrate that defence interest is moving beyond traditional warfare techniques to include Information Technology, with a potential objective to improve the UK armed forces responsiveness during a battle situation. In particular, it is suggested that intelligence exploitation and open-source intelligence may increase the success of a mission and ATIX will try to utilise and analyse various data sources, aiming to improve end-users' decision-making. Indeed, the main focus of ATIX is data analytics, making use of all available data sources to improve a commander's understanding in complex situations (BiP Solutions Limited 2017). DSTL notes that the system could work in 4 stages, and each stage interacts with a Virtual Knowledge Store where the data are analysed comprehensively. First, the data are collected. Tracks, UAV information, mobile messages and open-source information, when available, are combined and sent to the Virtual Knowledge Store

---

[4] These concepts are under development by DSTL. EPAS aims to maximise collected information by using all possible sensors, and ATIX aims to use exploited information to best judge when to change a course of action.

where the data are managed. The data are then extracted, analysed and compared against planned courses of action. Any deviations from these are identified and sent to the decision maker for action as appropriate.

4.1.1 Examining the potential for crowdsourced geographic information (VGI–iVGI)

MOD's *"Information Warrior"* and similar defence programmes that focus on developing information warfare capabilities provide an ideal opportunity to examine the possibility of improving the information intelligence through citizen or volunteer participation. The role that crowdsourcing can play during conflict, homeland security and generally in crisis situations when military needs to intervene, was previously explained (see Sect. 3.2), and the information required by ATIX could be enhanced by including both VGI, e.g. maps such as OSM, and iVGI such as Tweets, uploaded videos, posts to blogs and social media. The location of where things are happening, i.e. the "geography" in VGI, is the only way to integrate these very diverse sources of information.

The relevance of crowdsourced (VGI/iVGI) information is particularly important when taking into consideration that a huge amount of defence missions are operated in overseas territories (see Sect. 2). In this context, rapid knowledge of the area of responsibility can be of primary importance, and crowdsourced geographic information provides a way to access up-to-date situational information for specific geographic locations, both prior to and during a mission. Feeding this into the ATIX system may improve understanding of potential courses of action of opposing forces.

Beyond immediate decision-making, situational awareness of new environments can also be enhanced by examining the socio-cultural factors of the local population, which can in turn be explored via social media, where the local population can express its thoughts and beliefs. For instance, recent research used Flickr data to verify whether knowledge of new environments can be useful for defence cases (Lambio and Lakes 2017). By examining approximately seven million geolocated Flickr photographs collected for France and Afghanistan, respectively, it was noted that tags in English in both countries showed a peaceful and touristic area (France) on one side and a war environment (Afghanistan) on the other. The large amount of war-related tags was explained due to the vast amount of military operations in the fight against terrorism. However, by examining the national language tags (French, Afghan), a different message was found. Many Afghan tags referred to the cultural and historical heritage of the country. The researchers noted that English tags in Afghanistan were posted mostly by soldiers (USA, New Zealand, Australia, Canada, Ireland and UK), while the Afghan posts were uploaded mostly by the local population, which is potentially more interested in cultural activities and in the country's national monuments than the war. Therefore, findings from social media about new environments can increase situational awareness.

4.1.2 Data quality considerations for situational awareness and decision-making

Although in theory it may be possible to make use of VGI and iVGI in ATIX, a full understanding of its quality ("fitness for purpose") is necessary prior to including it, due to the generally variable quality of the data and issues such as fake news. It is thus important to understand the provenance of the data, and whether it can be trusted. Appropriate quality measures need to be automatically and very rapidly generated given the need for up-to-date situational awareness and data feeds into systems such as ATIX.

Moreover, the use of standardised protocols and guidelines that can be followed before the collection process can enhance VGI quality and especially its usability (Minghini et al. 2017). These standardised methods have been successfully implemented in citizen science. By improving VGI usability, the focus moves to user needs (see Sect. 3.4), so that the end-users can be more confident to include crowdsourced geographic information in their decision-making process because the validity and the accuracy of the collectors will be improved (Cohn 2008). Mooney et al. (2016) developed a generic protocol for VGI vector data collection that can increase the quality of VGI and thus the fitness-for-purpose for either a specific VGI project or a future one, and can guide new or experienced users. Minghini et al. (2017) then applied this protocol to real-world applications (one for VGI vector data and one for geotagged photographs).

4.2 Disaster response

During a natural disaster, rapid situational awareness is required to assist first responders in having a rapid overview of the disaster in a "new" environment (Endsley and Garland 2000; Ramchurn et al. 2015). The Orchid project involves a close collaboration between universities and industrial partners to develop new technologies and new science related to defence, and tries to close a technology readiness gap by moving research into real products and systems (Orchid 2017). Transforming machines, or agents, from passive (waiting for the user to decide a potential move), to energetic (acting autonomously and intelligently), is one of its main objectives (Jennings et al. 2014). The "energetic" machines need to deal with the volume, variety and pace of the information. Therefore, the balance between the relationship of the agents, also known as highly inter-connected components, and the users of the system is one of Orchid's core challenges.

Human-agent collectives (HACs) model this relationship in Orchid and are characterised as open and social, making use of AI to address problems in domains including disaster response and national security (Orchid 2017). Orchid's project "HACs in action" is aimed to support any natural disaster operation such as responding to an earthquake, a flood or a conflagration (Ramchurn et al. 2015; Orchid 2017). "HACs in action" collects and manages crowdsourced data in a system (CrowdScanner), increases the collaboration between humans and machines in an effective way (e.g. Multi-UAV coordination), improves human planning coordinated by a single

agent (AtomicOrchid) and finally tracks the data provenance in a system to reduce the probability of mistakes and the uncertainty level (Provenance Tracker).

CrowdScanner is a subsystem of the Android application named AppScanner (Amini et al. 2012), which collects, interprets and fuses both trusted and crowd-sourced information (Huynh et al. 2015). After a natural disaster, information needs to be collected by human teams, including volunteers and citizens, as quickly as possible. The information will be collected from various sources such as devices and social media. Evaluating the destruction of buildings and realising the number of human losses and first aid needs can improve the situational awareness of a potential operation. This heterogeneous information is gathered by CrowdScanner, and the information from the crowd is then compared to expert evaluation (Amini et al. 2012). Importantly, the system checks for issues such as noise, uncertainty, accuracy and trust levels as part of this process. Once the crowd's field information is combined and checked by the experts, tasks are prioritised and categorised according to urgency, and a machine (e.g. a UAV) will be sent to verify the situation and collect additional information. Multi-UAV coordination allows numerous machines to be coordinated intelligently, with the result that autonomous unmanned aerial and ground vehicles (UAVs and UGVs) can be used to collect information in an efficient and rapid way without the necessity of human control (Bethke et al.; Ramchurn et al. 2015).

4.2.1 Examining the potential for VGI to aid defence in disaster response

Unlike the situational awareness study above, Orchid directly addresses the use of VGI, both in terms of identifying sources of data, and in its awareness of the importance of the provenance of such data. Additionally, a wide range of software tools already exist to capture the required data, making use of devices such as mobile phones to provide the location-enabled content vital for integration and interoperability. For example, the Red Cross have developed an application that allows users to submit a status report during a disaster (Weaver et al. (2012), i.e. the CROSS system, for gathering data and for routing volunteers to affected areas (with centralised command and control monitoring) (Chu et al. 2012). Platforms such as Ushahidi, Sahana and Crows (Greengard 2011; Hui 2015) can be used for real-time information, especially after a crisis such as a natural disaster or bioterrorism attacks (Greengard 2011). These tools, in turn, highlight the fact that both VGI (e.g. OSM maps derived from the latest satellite images) and iVGI (social media such as Twitter, or uploaded videos) could be useful in this context. Additionally, lower technology approaches such as Short Message System (SMS) messages can also be useful since internet connectivity is not guaranteed. The Ushahidi platform makes use of this approach, taking advantage of the fact that access to mobile phones is increasing (Heinzelman and Waters 2010). Mobile phones, even if not enabled with GPS, can still provide approximate positional information by triangulation from mobile signal masts.

4.2.2 Data quality considerations for disaster response

Many of the data quality issues highlighted above (Sect. 4.1) apply here. In particular, it is important for the provenance of the data to be clearly identified before they are incorporated into the Orchid system, and issues relating to "fake news" are again relevant, e.g. an exaggeration of the number of people requiring assistance so as to focus efforts in a particular area or region.

An important consideration is the requirement to fully automate the quality assessment process as far as possible to enable a rapid flow of information and hence deployment of secondary machines (UAVs). While some provenance-related information, e.g. date, time, GPS accuracy, device type (mobile, operating system) and the software platform used for data capture, could be captured automatically, this requires that existing platforms be adapted to both capture and publish the information in a machine-readable format, and that this format is standardised across all platforms so that data from any source can be incorporated into Orchid. Current metadata standards (see Sect. 3.4.1) do not provide the required level of granularity. Additional information that may also feed into issues relating to trust in the data, e.g. operator name, operator skill level/experience, may also be captured semi-automatically. A balance also needs to be struck between having fully open and free text entry in the platforms to accommodate a wide variety of situations and issues that people may wish to report versus the need for systems such as Orchid to automatically parse the incoming data against a pre-defined schema. The underlying assumption of "always on", i.e. that devices will somehow be connected to a network and can thus upload data, may also be untrue in such situations.

5 Discussion: wider opportunities and challenges for deploying crowdsourced geographic information in UK defence

Building on the examples above and the broader opportunities highlighted by the literature review, this section further examines the opportunities for crowdsourced data use in UK defence (and hence in other similar defence forces), considering both benefits and constraints around its deployment with particular reference to data quality and fitness for purpose. The outputs of this research are focused on UK defence because of its global role in this domain. However, these potential opportunities are very much applicable more broadly. In particular, as described in Sect. 2, collaboration via NATO is a key component of the 8 missions defined by the MOD.

Examining the literature review first, Fig. 2 shows that the majority of research papers cited in Sect. 3 are linked to case studies based in the USA, reflecting a greater volume of work on this topic undertaken by US defence. In particular, US-based papers dominate in Sect. 3.2, where three quarters of the authors cite more than two US-based cases per paper. In Sect. 3.3, this percentage is reduced to 50% but is still high. While UK and US defence forces are close collaborators, these numbers highlight the need for more UK-focussed researches into the potential of crowdsourced geographic data for UK defence, with particular focus on the eight tasks outlined in Sect. 2.

TOTAL NUMBER OF SOURCES

SECTION 3.2
WARFARE/NATIONAL SECURITY

General
25%

US
75%

General
38%

US
62%

SECTION 3.3
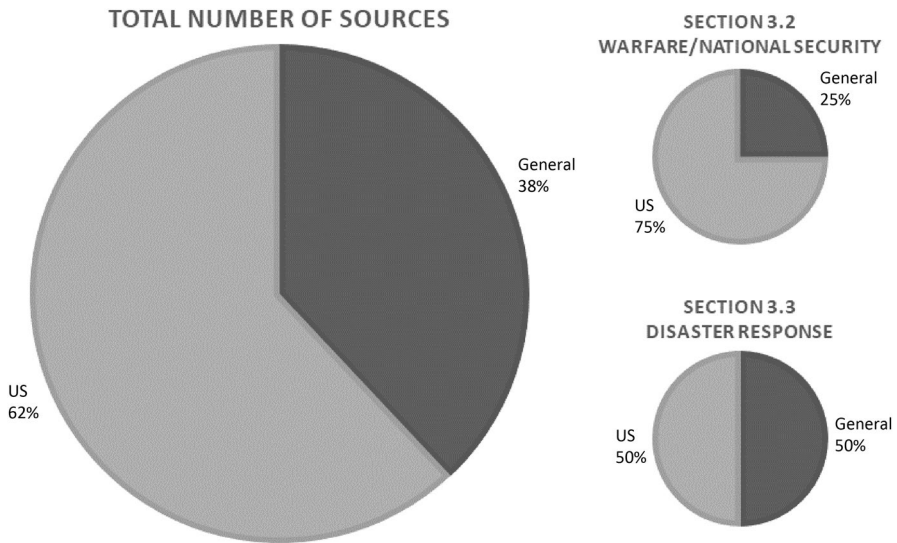DISASTER RESPONSE

US
50%

General
50%

Fig. 2 The pie chart on the left shows the total influence of papers from the USA. More than 60% of identified papers relate to the US DoD. For the papers in Sect. 3, only one quarter of the papers are not US DoD focussed when referring to warfare and national security cases (pie chart top right), while the percentage splits equally in disaster response crowdsourced activities (pie chart bottom right)

In Sect. 4, two studies from a UK defence background are presented. The first one is applicable to warfare conditions (UK defence mission tasks 1, 3, 4, 7, 8), aiming to improve their information capabilities to gained advantage against opponents. The role of intelligence exploitation is examined in an early stage, and the potential of crowdsourced geographic information deployment is explored. The second case relates both to disaster response cases but can also be applied in national security (UK defence mission tasks 5, 6, 7). By using a mixture of available technologies, machines acquire an energetic role so that their autonomy can be increased, and further information gathering tasks executed automatically. The studies illustrate the interest in information and strategic intelligence within defence, and while not specifically focussed on crowdsourcing, do highlight the potential of this data source for defence. In the first study, ways to assess the potential of open-source intelligence as an input into rapid decision-making processes need to be investigated, with focus on provenance, trust, validated information, and the challenges of merging this information with more "official" sources where these are available. In the second study, transforming machines such as UAVs, from acting passively (waiting for orders) to acting autonomously and intelligently, can improve the level of efficient data collection.

More specifically, both studies relate to the need to improve the situational awareness of the end-users in rapidly changing situations, and to reduce the time needed for the data assessment element that forms part of any decision-making process; in both studies, decisions must be made within a very short time-span. The literature review additionally confirms that, in emergency situations, crowdsourced geographic

information can have a role in the decision-making process by increasing the situational awareness of the user's area of responsibility. Therefore, VGI and iVGI could potentially fill an information gap observed in the two UK defence cases, providing up-to-date on-the-ground information in situations where alternative approaches to information gathering are not possible (Greengard 2011; Hui 2015).

Both the crowdsourced nature of the data (i.e. the fact that it is collected on the ground, in the area of interest, by people with local knowledge) and the geographic nature of the data (relating to a place, which provides the opportunity to integrate this information with "official" data sources already held by the various defence forces) are crucial to success.

Even given this potential, both the studies and the literature review highlight the fact that there are a number of challenges to be addressed to provide decision makers with assurance about the quality and fitness-for-purpose of the data. These are discussed below, with a number of recommendations for further research associated with each challenge.

5.1 Data quality and usability

Existing quality standards such as ISO 19113 are heavily biased towards internal elements (completeness, logical consistency, and positional, temporal and thematic accuracy), which refer to producer-oriented needs (Antoniou and Skopeliti 2015). They take the approach that quality can be documented as structured metadata to be used with "official" data sets, but it takes time to create, is complex to read and understand, is often not fully populated, requires manual creation, and even where it exists, is often held separately from the data and thus ignored. It is not adaptable to the fact that a data set may be reused in many different ways.

Additionally, standards-based metadata are very rarely available for crowdsourced geographic data. Users capturing the data may not be aware of the standards, and in the case of iVGI, they do not necessarily know that their data will be reused.

A standards-based approach to data quality documentation is fundamental to be able to automatically ingest data into a system such as that proposed by Orchid (Sect. 4.2) Full information about a data set's provenance, such as that provided by metadata, is also needed to ensure that a decision maker is aware of any limitations in the data set and can evaluate its fitness for purpose. This in turn will increase trust in the resulting analysis and decision-making processes.

Research challenges:

- *What are the data quality information needs of defence end-users, who for the most part are not expert GIS users or aware of existing standards? Do these maps to elements of existing standards (e.g. positional accuracy, currency) or should new elements and standards be investigated?*
- *What data quality information can be automatically captured as part of a VGI or iVGI process?*

- *How can data quality of crowdsourced data be automatically and rapidly assessed, in particular where no comparative data sets are available?*
- *How should the outcomes of any data quality assessment process be documented for:*
  - *Human evaluation, in particular in a rapid decision-making process?*
  - *Machine-based evaluation for automated data ingestion?*
- *What is the relationship between usable data quality descriptions and trust in a data set and in the outcomes of any analysis based on this data set?*
- *Can data quality descriptions be designed to allow automatic or semi-automatic assessment of fitness-for-purpose of a data set?*
- *Can we express the characteristics of a data set such that its fitness-for-purpose can be determined without an* a priori *understanding of all potential end-users?*

5.2 Data analysis and combining crowdsourced and other data sets

Big data analytics methods, used with social media sources such as Twitter or Facebook, are becoming increasingly commonplace, and machine learning techniques form an important part of predictive systems such as ATIX (Sect. 4.1). However, some caution is required when making use of such data, as different social media platforms are used by different groups of people, and hence, the outcomes of any analysis should not be taken to represent the population as a whole.

Additionally, both case studies highlight a situation that is becoming increasingly common when working with geographic information, i.e. the need to merge two or more very heterogeneous data sources to answer real-world problems. However, the data quality assessment methods described in Sect. 3 only focus on single data sets.

Research challenges:

- *Can analytical methods be developed to automatically take account of any awareness that the source data are not representative of an entire population, and to ensure that this is clearly described in any analytical results?*
- *Can analytical methods be developed to handle input data sets having variable quality and provenance?*
- *Can data quality information be used to recommend whether to use two data sets in combination?*

5.3 "Fake news"

Every decision made within a military context has a direct impact on human lives, and it is therefore important that any data underpinning decision-making is trusted. However, as noted above, fake news is on the increase, in particular where crowdsourced data are sourced from social media (i.e. an iVGI context). It is important that this is identified prior to input into the decision-making chain.

Research challenges:

- *Can data quality information, especially provenance, be used to identify "fake news" versus real information?*
- *Can the analytical techniques developed to merge crowdsourced and other data sets (Sect. 5.2) be adapted to identify fake news by comparing multiple sources of information about the same location, taking advantage of the power of GIS to integrate disparate, heterogeneous information?*

5.4 Privacy, legal, ethical issues and data ownership

The crowdsourced geographic information produced by a number of social media and VGI platforms includes automatic geolocation capabilities that can be used for different purposes, sometimes without the producer's knowledge (Shanley et al. 2013). In particular, as noted in Sect. 3.2, this information can, in turn, lead to the ability to predict an individual's movements.

The Virtual Knowledge Store machine (see Sect. 4.1), where whole data sets are analysed comprehensively, might need to inform the public that the whole process is used only for defence reasons and the safety system will be highly restricted. In theory, the level to which a volunteer or a social media user is willing to share the data can differ and can be configured by users if necessary. Some users may, in fact, be more open to sharing location information if informed that it will contribute to defence and personal safety (Boyd 2011). However, for other users, this may be an issue, whereas they may be more willing to share data in response to a natural disaster.

Crowdsourced geographic information and especially VGI is the subject of several privacy and ethical issues (Mooney et al. 2017). Ethical considerations when capturing data are important since both warfare and disaster situations pose a high degree of risk to human beings. Sometimes volunteers can participate without risking their lives, e.g. digitising data from a satellite image, but in many situations, this is not possible.

Issues such as data ownership, usage rights and interoperability can cause conflicts between the owners and the users (Shanley et al. 2013). Copyright and terms of use can be used to protect the data and give needed safety to the producers, reducing their concerns.

Taking into consideration that VGI is a relatively new research field, Mooney et al. (2017) note that the gap in understanding these issues is still large and there is scope for additional research. Moreover, privacy in VGI is not related only to individual contributors but also to organisations and institutions, while ethics and legal issues must be examined from both data producer and user perspectives. Standardised protocols and guidelines (see Sect. 4.1) can also improve the knowledge of volunteers regarding privacy and security. Some of the research challenges to be addressed in the near future related to defence include:

- *How can users of crowdsourcing platforms be made aware of potential defence uses of their data?*

- *Should there be a clear "opt-in" or "opt-out" approach to data sharing with defence?*
- *How can UK defence balance privacy and ethical considerations with the need for up-to-date information and better situational awareness?*
- *How should the ownership of any "derived data" or analytical results be handled?*

5.5 Next steps

The above research challenges are not, by any means, specific to the defence context or indeed to the UK or other defence forces. Indeed, a comparison with related literature outlined in Sect. 3.4 highlights the similarity of these challenges to those reviewed. However, it can be argued that there are some differences imposed by the defence context, in particular the need for rapid decision-making in situations that are in a state of flux, i.e. where the data used to underpin decisions are changing very rapidly. This means that when addressing any of the challenges above, any solutions must provide results in a very short time-frame.

It must also be clearly stated that the above list does not represent the full scope of crowdsourcing research as relevant to defence, in particular given the focus on data quality rather than on analytical algorithms. We argue that automating the assessment of data quality, and developing a better understanding of how this can be communicated to both defence analysts and downstream to decision makers is a fundamental first step in enabling the use of crowdsourced data in defence, increasing trust in the resulting data and analytical outputs through a better understanding of the provenance of data used in decision-making.

It is important to remember that every decision made within the defence context has direct impact on human life and thus, it must be based on the best information available. The need for a full audit/provenance trail and a comprehensive and detailed understanding of data quality, and the need to rapidly perform the required assessments and communicate this complex quality information succinctly to decision makers are two potentially contradictory requirements, which means that research into crowdsourcing in defence could be particularly challenging.

Given these challenges, it is therefore of fundamental importance to work closely with end-users of defence data to understand, in detail, the context in which such data are deployed. An additional challenge is that a fully embedded user-centred design approach (where the researcher observes operational activity) may not be possible in this context for security reasons, meaning that secondary approaches such as questionnaires, interviews and meetings may be required. The possibility of setting up challenges such as the "DARPA Network Challenge" (Sect. 3) may provide a more interactive way to address this challenge and engage both data quality experts and end-users.

6 Conclusions

The number of national security risks and threats has increased sharply in recent years, and the rise of terrorism has resulted in an increased number of warfare and national security operations. These incidents, in combination with the need to assist in disaster response activities worldwide, pose particular challenges to UK defence. Information from the crowd can potentially assist with these tasks significantly, but several issues and challenges will need to be addressed to allow users to have increased trust in this information, both in terms of its source and overall quality and subsequently, as a data component forming part of analytical output, and hence input into the decision-making process. Decisions made in this context have direct impact on human lives, so ensuring the fitness-for-purpose of any data used is fundamental.

This paper provides an overview of how crowdsourced data could be used in defence, analysing two UK defence cases where VGI and iVGI could be applied. It highlights, in particular, the need for appropriate methods to both rapidly determine the quality of crowdsourced data and then to communicate this quality to the users of the data, i.e. the analysts and the decision makers by providing sufficient information about quality and provenance to engender trust while also ensuring that the quality determination process is as rapid as possible and the information is not too extensive so as to slow down any decision-making process. This overarching challenge frames the research agenda for the use of crowdsourced data in defence research, which will only be successful if carried out in close collaboration with the end-users, i.e. defence analysts and decision makers.

References

Ali AL, Schmid F (2014) Data quality assurance for volunteered geographic information. In: GIScience: international conference on geographic information science, pp 126–141. https://doi.org/10.1007/978-3-540-87473-7

Amini S, Lin J, Hong J, Lindqvist J, Zhang J (2012) Towards scalable evaluation of mobile applications through crowdsourcing and automation. CMU-CyLab-12-006, Carnegie Mellon University

Andrejevic M (2006) The discipline of watching: detection, risk, and lateral surveillance. Crit Stud Media Commun 23(5):391–407. https://doi.org/10.1080/07393180601046147

Antoniou V (2011) User generated spatial content: an analysis of the phenomenon and its challenges for mapping agencies. Dissertation. University College London, UK

Antoniou V, Skopeliti A (2015) Measures and indicators of VGI quality: an overview. In: ISPRS annals of photogrammetry, remote sensing and spatial information sciences remote sensing and spatial information sciences, vol 2, pp 345–351. https://doi.org/10.5194/isprsannals-II-3-W5-345-2015

Associated Press (2017) Anger in Mexico as officials admit Frida Sofia, the "girl trapped in earthquake rubble" does not exist. In: Telegraph media group limited. http://www.telegraph.co.uk/

news/2017/09/22/anger-mexico-officials-admit-girl-trapped-earthquake-rubble/. Accessed 21 Sep 2017

Barron C, Neis P, Zipf A (2014) A comprehensive framework for intrinsic OpenStreetMap quality analysis. Trans GIS 18(6):877–895. https://doi.org/10.1111/tgis.12073

Becker D, Bendett S (2015) Crowdsourcing solutions for disaster response: examples and lessons for the US government. Procedia Eng 107:27–33. https://doi.org/10.1016/j.proeng.2015.06.055

Bethke B, Valenti M (2008) UAV task assignment automation. IEEE Rob Autom Mag 15:1

BiP Solutions Limited (2017) Defence contracts online. In: Bipsolutions. https://www.contracts.mod.uk. Accessed 11 Sep 2017

Bodine-Baron E, Helmus TC, Magnuson M, Winkelman Z (2016) Examining ISIS support and opposition networks on Twitter. RAND Corp 1:494. https://doi.org/10.7249/RR1328

Boyd D (2011) "Real names" policies are an abuse of power. In: Apophenia. http://www.zephoria.org/thoughts/archives/2011/08/04/real-names.html. Accessed 8 Aug 2018

Brooke-Holland L, Mills C (2015) The 2015 strategic defence and security review. House of Commons, Briefing paper number 7462, pp 1–42

Chrisman N (2006) Development in the treatment of spatial data quality. In: Devillers R, Jeansoulir R (eds) Fundamentals of spatial data quality. Wiley, Hoboken, USA, pp 21–30

Chu ET, Chen YL, Lin J-Y, Liu JWS (2012) Crowdsourcing support system for disaster surveillance and response. In: Proceedings of 15th international symposium on wireless personal communications, pp 21–25

Cohn JP (2008) Citizen science: can volunteers do real research? AIBS Bull 58:192–197. https://doi.org/10.1641/B580303

Coleman DJ, Georgiadou Y, Labonte J (2009) Volunteered geographic information: the nature and motivation of produsers. Int J Spat Data Infrastruct Res 4(1):332–358. https://doi.org/10.2902/1725-0463.2009.04.art16

Degrossi LC, Albuquerque JP de, Rocha R dos S, Zipf A (2017) A framework of quality assessment methods for crowdsourced geographic information : a systematic literature review. In: Proceedings of 14th ISCRAM conference, pp 21–24

Devillers R, Jeansoulin R (2006) Spatial data quality: concepts. In: Devillers R, Jeansoulin R (eds) Fundamentals of spatial data quality. Wiley, Hoboken, USA, pp 31–42

Dorn H, Törnros T, Zipf A (2015) Quality evaluation of VGI using authoritative data—a comparison with land use data in Southern Germany. ISPRS Int J Geo-Inf 4(3):1657–1671. https://doi.org/10.3390/ijgi4031657

Endsley MR, Garland DJ (2000) Situation awareness, analysis and measurement. CRC Press, Boca Raton, USA

Estellés Arolas E, González Ladrón-de-Guevara F (2012) Towards an integrating crowdsourcing definition. J Inf Sci 32(2):189–200. https://doi.org/10.1177/016555150000000

Filho JV (2013) Starting a spatially enabled society: a web system for collecting volunteered geographic information in the area of public security. In: GEOProcessing 2013: the fifth advanced geographic information systems, applications, and services, pp 151–154

Finin T, Murnane W, Karandikar A, Keller N, Martineau J, Dredze M (2010) Annotating named entities in twitter data with crowdsourcing. In: Proceedings of NAACL HLT 2010 work creating speech and language data with Amazon's mechanical turk, pp 80–88

Fischer F (2012) VGI as big data. A new but delicate geographic data-source. GeoInformatics 15(3):46–47

Flanagin AJ, Metzger MJ (2008) The credibility of volunteered geographic information. GeoJournal 72(3–4):137–148. https://doi.org/10.1007/s10708-008-9188-y

Fonte CC, Bastin L, See L et al (2015) Usability of VGI for validation of land cover maps. Int J Geogr Inf Sci 29(7):1269–1291. https://doi.org/10.1080/13658816.2015.1018266

Fonte CC, Antoniou V, Bastin L, Bayas L, See L, Vatseva R (2017) Assessing VGI data quality. In: Foody G, See L, Fritz S et al (eds) Mapping and the citizen sensor. Ubiquity Press, London, pp 137–163

Forghani M, Delavar M (2014) A quality study of the OpenStreetMap dataset for Tehran. ISPRS Int J Geo-Inf 3(2):750–763. https://doi.org/10.3390/ijgi3020750

Franklin AL, Mott T, Williams THL (2013) Coproduction in the U.S. Department of Defense: examining how the evolution of geographic information systems (GIS) expands non-traditional partner engagement. Policy Internet 5(4):387–401. https://doi.org/10.1002/1944-2866.POI345

Gallagher R (2013) Software that tracks people on social media created by defence firm. The Guardian, pp 13–16

Girres JF, Touya G (2010) Elements of quality assessment of French OpenStreetMap dataset. Trans GIS 14(4):435–459

Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69(4):211–221. https://doi.org/10.1007/s10708-007-9111-y

Goodchild M, Gopal S (1989) The accuracy of spatial databases. CRC Press, Boca Raton, pp 107–113

Government Digital Service (2017) GOV.UK. In: Crown copyright 2017. https://www.gov.uk. Accessed 12 Sep 2017

Greengard S (2011) Following the crowd. Commun ACM 54(2):20–22. https://doi.org/10.1145/1897816.1897824

Haklay M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environ Plan B Urban Anal City Sci 37(4):682–703. https://doi.org/10.1068/b35097

Haklay M, Singleton A, Parker C (2008) Web Mapping 2. 0: the neogeography of the GeoWeb. Geogr Compass 2(6):2011–2039. https://doi.org/10.1111/j.1749-8198.2008.00167.x

Harding J (2006) Vector data quality: a data provider's perspective. In: Devillers R, Jeansoulir R (eds) Fundamentals of spatial data quality. Wiley, Hoboken, USA, pp 141–158

Havas C, Resch B, Francalanci C, Pernici B, Scalia G, Fernandez-Marquez JL, Van Achte T, Zeug G, Mondardini MRR, Grandoni D, Kirsch B, Kalas M, Lorini V, Rüping S (2017) E2mC: improving emergency management service practice through social media and crowdsourcing analysis in near real time. Sensors 17:2766. https://doi.org/10.3390/s17122766

Heinzelman J, Waters C (2010) Crowdsourcing crisis information in disaster. US Institure of Peace, Special Report 252

Heipke C (2010) Crowdsourcing geospatial data. ISPRS J Photogramm Remote Sens 65(6):550–557. https://doi.org/10.1016/j.isprsjprs.2010.06.005

Helbich M, Amelunxen C, Neis P (2012) Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. In: Proceedings of GI_Forum, pp 24–33. https://doi.org/10.1080/13658816

Hope C, McCann K (2017) Google, Facebook and Twitter told to take down terror content within two hours or face fines. In: Telegraph media group limited 2017. http://www.telegraph.co.uk/news/2017/09/19/google-facebook-twitter-told-take-down-content-within-two/. Accessed 22 Sep 2017

Hui JY (2015) Crowdsourcing for national security. Nanyang Technological University, pp 1–15

Huynh TD, Ramchurn SD, Moreau L (2015) HAC-ER: tracking provenance in disaster response. In: Proceedings of 2015 international conference on autonomous agents and multi-agent systems international foundation for autonomous agents and multi-agent systems, pp 533–541

Jennings NR, Moreau L, Nicholson D, Ramchurn S, Roberts S, Rodden T, Rogers A (2014) On Human-agent collectives. Commun ACM 57(12):80–88

Johnson CW (2014) Anti-social networking: crowdsourcing and the cyber defence of national critical infrastructures. Ergonomics 57(3):419–433. https://doi.org/10.1080/00140139.2013.812749

Lambio C, Lakes T (2017) Placing of photos on the internet: critical analysis of biases on the depictions of France and Afghanistan on FLICKR. Geoforum 82:21–25. https://doi.org/10.1016/j.geoforum.2017.03.016

Mayfield A (2011) A commander's strategy for social media. Networks 1st quarte 1(60):79–83. https://doi.org/10.2217/rme.11.82

Mills A, Chen R (2009) Web 2.0 emergency applications: how useful can Twitter be for emergency response? J Inf Priv Secur 5(3):3–26. https://doi.org/10.1080/15536548.2009.10855867

Minghini M, Antoniou V, Costa CC, Estima J, Olteanu-Raimond AM, See L, Laakso M, Skopeliti A, Mooney P, Arsanjani JJ, Lupia F (2017) The relevance of protocols for VGI collection. In: See L, Fritz S, Mooney P, Olteanu-Raimond A-M, Fonte CC, Antoniou V, Foody G (eds) Mapping and the citizen sensor. Ubiquity Press, London, pp 223–247

Ministry of Defence UK (2015a) SDSR 2015: Defence key facts

Ministry of Defence UK (2015b) SDSR 2015: Defence fact sheets

Mooney P, Minghini M, Laakso M et al (2016) Towards a protocol for the collection of VGI vector data. ISPRS Int J Geo-Inf 5:217. https://doi.org/10.3390/ijgi5110217

Mooney P, Olteanu-Raimond A-M, Touya G, Juul A, Alvanides S, Kerle N (2017) Considerations of privacy, ethics and legal issues in volunteered geographic information. In: See L, Fritz S, Mooney P, Olteanu-Raimond A-M, Fonte CC, Antoniou V, Foody G (eds) Mapping and the citizen sensor. Ubiquity Press, London, pp 119–135

Navy R (2017) Royal Navy. In: Crown copyright 2017. https://www.royalnavy.mod.uk. Accessed 10 Sep 2017

Neis P, Zielstra D (2014) Recent developments and future trends in volunteered geographic information research: the case of OpenStreetMap. Futur Internet 6(1):76–106. https://doi.org/10.3390/fi6010076

Orchid (2017) The ORCHID project. In: Orchid project. www.orchid.ac.uk. Accessed 5 Sep 2017

Ortmann J, Limbu M, Wang D, Kauppinen T (2011) Crowdsourcing linked open data for disaster management. In: Proceedings of terra cognita 10th international semantic web conference, pp 11–22

Parsons B (2011) U. S. Government turns to crowdsourcing for intelligence. Nat Defense 12:26–27

Perrin A (2015) Social media usage: 2005–2015. Pew Research Center, Washington, DC, USA, pp 52–68

Pourabdollah A, Morley J, Feldman S, Jackson M (2013) Towards an authoritative OpenStreetMap: conflating OSM and OS OpenData national maps' road network. ISPRS Int J Geo-Inf 2(3):704–728. https://doi.org/10.3390/ijgi2030704

Ramchurn SD, Fischer JE, Ikuno Y, Wu F, Flann J, Waldock A (2015) A study of human-agent collaboration for multi-UAV task allocation in dynamic environments. In: IJCAI international joint conference on artificial intelligence 2015–January, pp 1184–1192

Roy SE, Kase EK, Bowman H (2017) Crowdsourcing social media for military operations. Association for Computing Machinery, pp 23–27

See L, Mooney P, Foody G, Bastin L, Comber A, Estima J, Fritz S, Kerle N, Jiang B, Laakso M, Liu HY, Milčinski G, Nikšič M, Painho M, Pődör A, Olteanu-Raimond AM, Rutzinger M (2016) Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS Int J Geo-Inf 5(5):55. https://doi.org/10.3390/ijgi5050055

Senaratne H, Mobasheri A, Ali AL, Capineri C, Haklay Ml (2017) A review of volunteered geographic information quality assessment methods. Int J Geogr Inf Sci 31(1):139–167. https://doi.org/10.1080/13658816.2016.1189556

Shanley LA, Burns R, Bastian Z, Robson ES (2013) Tweeting up a storm the promise and perils of crisis mapping, pp 865–879. https://doi.org/10.2139/ssrn.2464599

Suleyman C (2017) A moment that changed me: walking home through the London riots in 2011. In: The guardian. https://www.theguardian.com/commentisfree/2017/aug/04/a-moment-that-changed-me-london-riots-2011. Accessed 15 Sep 2017

Tewksbury D (2012) Crowdsourcing homeland security: the Texas virtual borderwatch and participatory citizenship. Surveillance Soc 10(3/4):249–262

Turner AJ (2006) Introduction to neogeography. O'Reilly Media, Inc, Sebastopol, USA

UK HM Government (2015) National security strategy and strategic defence and security review 2015: a secure and prosperous United Kingdom

Van Exel M, Dias E, Fruijtier S (2010) The impact of crowdsourcing on spatial data quality indicators. In: Proceedings of the sixth International conference on geographic information systems (GIScience 2010), pp 1–4

Weaver AC, Boyle JP, Besaleva LI (2012) Applications and trust issues when crowdsourcing a crisis. In: 2012 21st international conference on computer communications and networks, ICCCN 2012—Proceedings, pp 1–5

Zielstra D, Hochmair HH (2013) Positional accuracy analysis of Flickr and Panoramio images for selected world regions. J Spat Sci 58(2):251–273. https://doi.org/10.1080/14498596.2013.801331