

The Value of Social Media Data: Integrating Crowd Capabilities in Evidence-Based Policy

Forthcoming article in Government Information Quarterly

Panos Panagiotopoulos

*Senior Lecturer in Information
Management*

Queen Mary University of
London

P.Panagiotopoulos@qmul.ac.uk

Frances Bowen

*Professor of Innovation
Studies*

Queen Mary University
of London

F.Bowen@qmul.ac.uk

Phillip Brooker

Lecturer in Sociology
University of Liverpool

P.D.Brooker@liverpool.ac.uk

*This is the authors' version. The final version will be made available from the journal
website <http://www.sciencedirect.com/science/journal/0740624X?sdc=1>*

Abstract

Social media have been widely embraced by governments for information dissemination and engagement but less is known about their value as information sources. Crowdsourced content from social media can improve inclusivity in policy development but it is not always clear how it can form part of policy evidence. The paper builds on the conceptual framework of crowd capabilities to examine the value of social media data in evidence-based policy. Acquisition and assimilation – the two elements of crowd capabilities – drive our exploratory case analysis in the context of agricultural policies in the UK. The study combined qualitative data from interviews and workshops with an analysis of networks of farmers on Twitter. Policy makers were broadly positive about the immediacy, cost-effectiveness and diversity of useful input that can be sourced from online sources. Limitations were identified in terms of representation and inclusion of participants in large datasets that are sourced from open platforms. We compare social media data to traditional sources of evidence and further reflect on the new capabilities that can support the needs of policy makers in this endeavor.

Keywords: policy crowdsourcing, environment and farming, absorptive capacity, big data, case study.

1. Introduction

Governments are increasingly integrating data from digital interactions in new approaches to policy and regulation. Social media users broadcast information about their personal and professional activities, provide feedback on policy topics or engage in discussions that can potentially inform policy development. As a result, there are increasing signals about tools

being used in the public sector to source, aggregate, filter, analyze and visualize content from social media (Bekkers, Edwards, & de Kool, 2013; Panagiotopoulos, Shan, Barnett, Regan, & McConnon, 2015; Williams et al., 2013).

Social media monitoring or simply social monitoring is gaining significant attention as organizations are learning how to harness and exploit contributions from social media users (Benthaus, Risius, & Beck, 2016; Culnan, McHugh, & Zubillaga, 2010; Dong & Wu, 2015; Katona & Sarvary, 2014). Due to the technical and organizational complexities involved in social monitoring, such practices can be considered as a more ad hoc form of policy crowdsourcing (Prpić, Tæihagh, & Melton, 2015). Compared to active crowdsourcing platforms, contributions sourced from social media spaces have usually been posted with different original intentions and without an explicit purpose to contribute to a policy topic.

Whether labeled as crowdsourcing or social monitoring, progress with understanding and harnessing the value of social media data in policy has been rather limited. Government departments have been focusing on managing their digital media presence, disseminating information to their respective audiences and, to a varying extent, considering how to stimulate and manage interactions. Academic studies have respectively outlined the strategic and operational benefits (Criado, Sandoval-Almazan, & Gil-Garcia, 2013), developed models to map these new types of interactions (Mergel, 2013) or looked at how social media adoption processes are unfolding within government agencies (Mergel & Bretschneider, 2013).

When social media data are collected and visualized, such exercises are likely to take place using monitoring platforms that are based on content metrics developed to assist decisions in a commercial communications context (Mergel, 2012). Monitoring trends and collecting reactions can be useful for tasks like emergency management but might not align with the needs of policy makers to develop evidence-based approaches (Cartwright & Hardie, 2012; Head, 2008; Parsons, 2002; Sanderson, 2002). This is because, despite being a plentiful source, social media data differ from current sources of public and stakeholder evidence such as surveys and consultations. Furthermore, harnessing the value of social media data might require new capabilities for policy makers to address challenges such as understanding of how social conversations evolve and who the key stakeholders are (Castelló, Morsing, & Schultz, 2013; Janssen & Helbig, 2016).

Given these trends and perspectives, this paper aims to explore:

- (1) *How do policy makers perceive the value of social media data as policy evidence?*
- (2) *Which capabilities can support the needs of policy makers to integrate input from social media?*

To address these questions, we draw on the conceptual framework of crowd capabilities by Prpić et al. (2015) that is based on broader work on absorptive capacities (e.g. Culnan et al., 2010; Zahra & George, 2002). Crowd capability refers to an organization's ability to source a large amount of contributions from an unknown digital audience (acquisition) and develop appropriate internal information flows to transform and exploit this content to meet organizational aims (assimilation). By understanding how policy makers perceive the acquisition and assimilation potential of social media data, we establish the link between absorptive capacities and the value of social media data as policy evidence.

We conducted an exploratory study with policy-making teams in the UK Department for Environment, Food and Rural Affairs (DEFRA). The study shows that there is good scope for sourcing social media content as long as: (1) it is appropriately summarized and (2) includes useful information about the composition of digital audiences and the context in which the data have been produced. The more these conditions were met the more policy makers could recognize social media data as a useful source of evidence. Despite the potential to broaden the diversity of input from the public, there were important technical and methodological limitations (e.g. lack of network-feedback tools, inclusion and representation of social media users). The paper concludes by discussing the challenges of integrating social media data in evidence-based policy using the lens of crowd capabilities.

2. Conceptual background

Social media include a large ecosystem of networking and information sharing platforms with diverse functionalities (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). Inevitably, there is a level of audience fragmentation and distribution of activities across platforms. Advanced social media users in organizations need to manage social interactions across platforms, devices and multiple users as well as find useful ways to monitor content. These needs have led to the development of social media management and monitoring tools with features like administration and oversight of multiple accounts, scheduling and planning content (e.g. updating blogs), audience engagement indicators (e.g. number of retweets or likes), real-time monitoring of trends, searches based on keywords (e.g. product feedback), reporting and exporting options, dashboards and other visualizations that help users navigate and filter information ¹.

These features can offer powerful support to an organization's ability to promote content, monitor and engage, including public sector organizations that also face challenges in managing digital interactions. Nevertheless, these features and their underlying interfaces have been mostly developed and optimized for brand engagement, reputation-based applications and other business decisions where counting social impressions is important (e.g. Benthaus et al., 2016; Fan & Gordon, 2014; Hoffman & Fodor, 2010; Schniederjans, Cao, & Schniederjans, 2013). The needs of policy makers are likely to extend beyond snapshots of public opinion though popularity measures.

2.1. Social media data in government

Social monitoring in government has been driven by two important trends. First, policy makers are seeking alternative, more cost effective and less direct means of developing policy while adhering to higher standards of openness and transparency (Obama, 2009; UK Government, 2015). Crowdsourcing and digital engagement activities can support this agenda, although traditional challenges like procedural fairness and feedback-outcome links remain. For example, this was the case with the UK government's Red Tape Challenge that

¹ Although a review of commercial monitoring tools is outside the scope of this research, an indicative list of applications would include: HootSuite, Topsy, Radian6, SproutSocial, Social360, Pulsar, Brandwatch, Engagor, Trackur, Sysomos, Nielsen BuzzMetrics, Buzzient, Social Radar, Social Mention, Klout, Seesmic and many others.

crowdsourced regulatory input between 2010 and 2012 (Lodge & Wegrich, 2014). Second, there have been significant advances in the areas of analytics and big data research both in terms of technical capabilities and awareness about the importance of data in society (Janssen & Helbig, 2016; Schintler & Kulkarni, 2014).

On the empirical side, related work has focused on intelligence gathering during emergency events or policing (e.g. Sutton, 2009; Williams et al., 2013). Another stream has looked at the predictive and preventive ability of social monitoring in public health (e.g. Brownstein, Freifeld, & Madoff, 2009; Kostkova, Szomszor, & St. Louis, 2014). A study by Bekkers et al. (2013) in the Netherlands suggests that organizations with established surveillance mechanisms like the police are more willing to consider social media sources in comparison to policy teams in other departments that prefer the monitoring of closed spaces (e.g. forums). In the UK, there is some indication of social monitoring taking place to proactively identify conversations of interest in incidents of food safety and campaigns about food hygiene (Panagiotopoulos et al., 2015). Charalabidis et al. (2014) examine the functional and policy requirements of social monitoring platforms for policy makers. Ferro et al. (2013) evaluate such a platform that visualizes content from multiple social media sources – insights were found to be meaningful mainly when cross-examined with traditional forms of data like surveys.

Alongside their potential value, social media data raise a series of ethical, privacy and risk management issues beyond the ones summarized by Picazo-Vela et al. (2012, p. 508) for the broader use of social media in the public sector. Bekkers et al. (2013) identify strategic dilemmas involved in social monitoring by agencies when it comes to balancing responsiveness and surveillance. They suggest that social media sources can be insightful as policy input and even to proactively address citizen's concerns. Transparency is highly advised by the authors, especially for forums that their users would perceive as private. These are indeed some of the main complexities of social media data: (1) users' contributions are usually posted with different intentions and target audiences and (2) it is not feasible to seek every individual user's consent when aggregating large amounts of publicly available contributions (Boyd & Crawford, 2012). In their guidelines on addressing such issues in social media research, Evans, Ginnis, & Bartlett (2015) suggest that ethical considerations and good practice have to be continually reviewed to meet changes in technology, legislation and user expectations.

Notwithstanding these challenges, there is a lot more to know about the value of social media data as part of the policy making cycle and evidence gathering exercises. It is important to look further into these aspects because social media can generate large amounts of data that do not resemble traditional sources (e.g. administrative or public engagement data) and can contain new insights on how groups of the public think about policy topics. We now discuss social media data alongside the principles of evidence-based policy.

2.2. Evidence-based policy, public engagement and social media data

Evidence-based policy refers to the systematic use of evidence to inform public policy decisions; an approach that originates from evidence-based medicine (Howlett, 2009; Marston

& Watts, 2003; Parsons, 2002; Sanderson, 2002). A key moment in the popularization of evidence-based policy was the 'Modernising Government' white paper by the British government (1999) that stimulated much interest in knowledge management in policy making and the analytical capacities of social science research (Parsons, 2002; Sanderson, 2002).

Policy evidence can take both scientific (e.g. statistics, experiments) and social science forms (e.g. surveys, focus groups) depending on the nature of the topic, collection opportunities and the capacities of policy makers (Howlett, 2009; Parsons, 2002). Governments need a certain level of policy analytical capacity to be able to perform tasks that include systematic evidence although recent work shows that this is not always the case in practice (Newman, Cherney, & Head, 2017).

An important principle of evidence-based policy emphasizes the need to build evidence bases through public and stakeholder engagement, which constitute an important source of social science evidence (Head, 2008). Engagement with the public and stakeholder understanding are usually seen as part of establishing long-term relationships (Kasabov, 2008; Sanderson, 2002) and developing public engagement capabilities (Pang, Lee, & DeLone, 2014). When this is achieved through traditional sources such as consultations or surveys, the extent of input might either be too limited, too costly, skewed by dominating actors or might simply take too long to be collected. This is where social media data emerge as a plentiful and immediate source of information for important stakeholders and the general public.

Policy makers collect, filter, translate and interpret evidence to design and implement policies based on both systematic and idiosyncratic assessments (Ingold & Monaghan, 2016). Many of these assessments about information flows and public input are not new, but the novelty, volume, complexity and diversity of social media data provide a challenging new landscape. The needs and requirements of policy makers in integrating social media data commonly extends beyond snapshots of public opinion and requires a more advanced understanding of how social conversations evolve, who the key stakeholders are and how specific forms of input reflect the state of communities or professional groups (Castelló et al., 2013; Kasabov, 2008; Sanderson, 2002). This raises questions about who these new audiences are, and which forms of social media content can be regarded as useful policy evidence.

Current research has not yet paid adequate attention to the composition of digital audiences and how sourced content can help policy makers gain new insights. Although there are clear opportunities to engage with new demographics on social media (Murthy, Gross, & Pensavalle, 2015), the relationship between traditional and social media audiences can be particularly challenging for organizations (Kidd, 2011; Panagiotopoulos et al., 2015). The pluralism of social media inevitably leads to audience fragmentation and distribution of activities across channels where dynamic audiences come together around specific interests (e.g. Highfield, Harrington, & Bruns, 2013). Policy makers have to make assumptions about who are the people they are listening to and assess the usefulness of the collective input produced by social media audiences. Such processes of understanding the public usually have high influence on engagement practices and, in turn, drive responses from the public about important policy decisions (Barnett, Burningham, Walker, & Cass, 2012; Walker, Cass, Burningham, & Barnett, 2010).

In terms of the relevance of social media data in policy, we cannot assume that large amounts

of aggregated contributions will always contain useful insight (Boyd & Crawford, 2012; Davies, Nutley, & Smith, 2001). The realities of information use in policy making suggest that large datasets come with biases that differ significantly across activities and stages of policy making (Schintler & Kulkarni, 2014). ICTs are generally expected to provide solid evidence through “hard” and “big” data stored in information systems (Misuraca, Codagnone, & Rossel, 2013; Wastell, 2006). However, as Kum et al. (2015) illustrate using empirical cases in child welfare, there are inherent complexities in translating big datasets of administrative data into policy evidence. Similarly, new digital interactions might expand rather symbolic capacities for engagement where an increased volume of conversations that appears relevant does not necessarily translate to new insight (Zavattaro & Sementelli, 2014).

Therefore, there are both practical and theoretical challenges of sourcing content from social media audiences and the new capabilities to support this endeavor. We now turn to the concept of crowd capabilities as a useful way to frame social media data as policy evidence and understand the capabilities required by policy makers to utilize such data as evidence.

2.3. Crowd capabilities and the needs of policy makers

In the organizational literature, the concept of absorptive capacity generally involves a set of capabilities for sourcing, monitoring, filtering and exploiting information within organizations (Cohen & Levinthal, D, 1990; Zahra & George, 2002). Consistent with dynamic capabilities and the resource-based view of the firm, the concept refers to internal processes that lead to competitive advantage through external knowledge acquisition, processing and eventually exploitation.

Absorptive capacity has been widely applied in information systems research (e.g. Roberts, Galluch, Dinger, & Grover, 2012) and also directly used to describe an organization’s ability to absorb social media content that can be exploited in ways that can create value (Culnan et al., 2010; Ooms, Bell, & Kok, 2015; Schlagwein & Hu, 2017). Culnan et al. (2010) note that digital interactions in social media platforms set new challenges for content monitoring and reporting as part of organizational learning. Schlagwein and Hu (2017) explore how different types of organizational use of social media can support organizations’ absorptive capacity. Ooms et al. (2015) more specifically explore the link between social media sources and absorptive capacities for inbound open innovation in companies. These perspectives altogether make evident that the link between sourcing social media data and creating organizational value involves new tools for sourcing content and new analytical techniques like data visualization.

In response to these challenges, the concept of crowd capabilities is a novel approach for aligning contributions from digital interactions to an organization’s processes (Prpić, Shukla, et al., 2015; Prpić & Shukla, 2014). Prpić et al. (2015) frame a crowd capability as an organization’s core part in acquiring crowd capital or new resources through crowdsourcing. Specifically, crowd capabilities refer to decisions within an organization on how to “(1) obtain resources dispersed in a crowd and (2) align crowd contributions with its existing internal processes.” (Prpić, Shukla, et al., 2015, p. 81). This two-phase distinction comes directly from the two necessary and complementary components of absorptive capacity

identified by Zahra and George (2002): acquisition and assimilation of external knowledge.

In the evidence-based policy context, *acquisition* refers to activities of sourcing content from social media that generates two challenges: (1) understanding the types of digital interactions on social media that can be potential useful (e.g. audiences, platforms, frequency and scope of monitoring) beyond measuring social impressions or regular monitoring of a set of keywords and (2) choosing the appropriate infrastructure to source content and enable incoming information flows (e.g. monitoring tools).

Assimilation refers to internal transformation processes where crowd contributions are aggregated, filtered, visualized or otherwise transformed to enable useful information flows within the organization. In a policy context, this relates to incorporating social media analysis as input to different stages and activities of policy making. These transformation processes are not always as simple as summarizing large datasets to answer a particular question (e.g. public opinion towards a new policy). As Prpić et al. (2015) note, some forms of crowd engagement require the aggregation of collective contributions while others are more analytically-oriented and require multiple levels of filtering.

As crowd capabilities are applicable to a wide range of processes in which a crowd can be potentially involved (e.g. crowdfunding, competitions, crowd voting) it is important to further develop and expand the concept as a motivating framework for this study. Crowd capabilities can provide a systematic consideration of the new capabilities that can enable the integration of social media in evidence-based policy. At the same time, the concept provides an analytical structure to address the gap in our understanding of how policy makers perceive the value of social media data and how they can be used alongside established sources of evidence. Drawing these contemporary theoretical and practical strands together, this study asks: (1) how policy makers perceive the value of social media data as policy evidence, and (2) which capabilities can support their needs to integrate input from social media.

3. Study context and methodology

We conducted an exploratory case study to shed light on the new capabilities that can facilitate the link between social media data and evidence-policy (Seawright & Gerring, 2008; Yin, 2014). The UK Department of Environment, Food and Rural Affairs (DEFRA) provided a useful setting for this study as a government department that draws heavily on public engagement and the use of external stakeholder input in policy decisions. DEFRA's policies relate to a variety of topics where crowdsourced data are relevant but the size and composition of involved audiences differs significantly (e.g. from farming reform and flood protection to specialized environmental issues like forestry, chemicals and pesticides). Another important reason for the selection of DEFRA is the organization's commitment to evidence-based methods and the implementation of initiatives that aim at positive behavioral change (e.g. sustainable consumption, energy labeling, reducing food waste) (Collier et al., 2010; DEFRA, 2015).

We focused on farming and agricultural policy as a representative illustration of how policy makers understand the value of social media data within DEFRA's remit. Farming and

agricultural networks on social media are an unexplored field that involves a set of challenges to policy and regulation.

The study was designed based on a mixed methods perspective. As Urquhart and Vaast (2012) and Whelan et al. (2016) explain, building theory related to social media in organizations involves combining insights from digital interactions with traditional sources of research data. First, we collected and analyzed a wide range of documentary evidence from a selection of policy topics including consultations, response to consultations, social media posts and evaluation reports. Then we conducted several analyses of how Twitter conversations can reflect the state of the farming and agricultural communities in the UK. Finally, we used these analyses in two workshops and a series of interviews with policy makers to probe their perceptions about the value of social media data. All data collection activities took place between November 2014 and March 2015.

3.1. Workshops and social media analysis

We collected and reviewed different accounts, common keywords and hashtags associated with British farming on Twitter². The aim was to explore how collective content by farmers on Twitter could be suitable as input in policy activities and the forms of analysis that could support this input. Data from Twitter were sourced and analyzed using Chorus Analytics, a research-based social media analysis suite designed to support social science projects (Chorus, 2017).

Here we report on one example of this analysis based on a dataset of 85,436 tweets that were captured retrospectively over a period of six months from a random selection of 700 users. This group was sampled from the followers of @NFUTweets, the main Twitter account of the National Farmers Union of England and Wales. As @NFUTweets is one of the major hubs of British farming on Twitter, the aim was to gain an exploratory snapshot of main conversations around farming. The dataset was summarized and visualized with the help of the cluster mapping technique that produces tree-structured maps based on keyword frequency and co-occurrence within tweets (Cribbin, 2009; Westerman & Cribbin, 2000). The cluster map was available via the Chorus Analytics platform that was also used to capture the dataset and produce other forms of exploratory analysis (e.g. timeline of tweets, sentiment analysis, ad hoc filtering based on keywords) (Chorus, 2017). Microsoft Excel and NVivo were further used to store extracted files, sort information, and visualize word clouds.

Chorus' cluster map algorithm draws on a generated word index which computes a strength value for all co-occurrences between terms in a given dataset (collection of tweets). The closer different words appear in the cluster map, the more frequently they co-occur in tweets. In this way, an entire dataset can be visualized as a connected collection of clusters of strongly associated words³. Practically, this results in topic-based analyses of Twitter data as a collection of distinct topics and sub-topics and the tracing of connections between them.

² An important source of information came from a postgraduate dissertation that specialized on the topic (Haley, 2013).

³ Words that are mentioned together in tweets more frequently are given a higher co-occurrence value to reflect the strength of their association. The visualization then uses these indices to draw a series of cluster maps, where strongly associated terms, tweets and intervals are plotted closer together.

This technique has been used, for example, to inform research work into sufferer experiences of living with cystic fibrosis (Brooker et al., 2014).

The findings of this analysis were discussed in two workshops with five and six participants respectively. Cluster mapping was discussed as one of the possible techniques that could be part a social media analysis toolkit for policy makers. Participants were asked to provide feedback on the findings while discussing their wide information requirements regarding the use of social media data. The exercise was not aiming to address issues directly related to the topics at hand but rather mirror the types of audiences and analysis methods that policy makers can tap into, therefore assisting them to understand the potential of such methods. Issues raised at the workshops were further discussed in the interviews with key informants.

3.2. Interviews with key informants

We held seven semi-structured interviews with selected participants from different levels of the civil service and policy areas mainly related to communications and regulations about the environment, farming and local growth. Participants were first asked about their role within the organization. The rationale for selection was the relevance of social media data to policy tasks despite their varying levels of familiarity with social media. Interview discussions evolved around the following main questions:

- *Crowd acquisition*: what type of input from external stakeholders does your role require? What kind of information flows support this input (e.g. consultations, surveys, other stakeholder engagement activities)? How could information from social media change the ways in which you understand the needs of external stakeholders and publics?
- *Crowd assimilation*: How do you think input from social media can support the work of your policy team and DEFRA in general? How do you understand the value of social media as sources of policy evidence in your current role? Which resources and capacities can facilitate better use of social media data within your role?

Interviews lasted one hour on average and were taped and transcribed following permission from participants. Data analysis took place inductively starting from the two main categories of crowd acquisition and assimilation (Braun & Clarke, 2006; Krippendorff, 2013). These main themes subsequently provided a useful guide to structure the presentation of findings in section 4. In most interviews, participants illustrated their narratives around social media data using examples from their own experiences (e.g. following a Twitter hashtag or reading reports from analytics tools) or drawing on the findings of the social media analysis as a probe (cluster maps). The latter were not a technique that many participants had come across but they provided a hands-on example of how large datasets from social media can be summarized, navigated and transformed to potential evidence through analytical insight. This allowed participants to focus their observations on both the practical/every day aspects of social media data and the strategic/policy implications linked to new capability development.

In the next section, we begin with a brief case background followed by a presentation of the cluster maps developed from @NFUTweets as an illustrative analysis of farming and agricultural conversations on Twitter. In sections 4.2 and 4.3, we join together findings from

the two parts of the study (workshops and interviews) based on the two aspects of a crowd capability: assimilation (sources of input and the public) and acquisition (forms of analysis).

4. Findings

DEFRA is one of the largest government departments in the UK with remit in policy and regulation related to environmental protection, food production and standards, agriculture, fisheries and rural communities. DEFRA’s objectives and priorities include the improvement of technical infrastructure in rural areas, increasing exports and competitiveness in the food chain, simplifying farming regulation and improving water quality (DEFRA, 2015). The department employs over 10,000 staff working across 36 agencies and public bodies in England with devolved administrations in Wales, Scotland and Northern Ireland.

DEFRA’s approach to policy making places emphasis on issue definition and situation understanding where engagement with the public is identified as a priority (shown in figure 1). This is necessary for the work of DEFRA due to the high complexity, economic impact and technical nature of environmental issues as well as the wide variety of stakeholders usually involved. In practice, evidence-based assessments include activities like policy consultations, social science research, public understanding studies (e.g. geographical mapping, experiments, surveys and focus groups) and stakeholder management activities involving professional associations, academic teams and other experts.

Agriculture is a significant part of the UK economy with an estimated number of 250K farmers (DEFRA Statistics, 2015). There has been a particular focus on simplifying farming regulations and helping “farmers move online”, which includes measures to improve broadband connectivity in rural areas and introduce new digital services for payments, reporting and monitoring (e.g. Online Sheep and Goat Inventory) (DEFRA, 2014). British farmers and agricultural businesses generally form a large, geographically dispersed community based on regional shared identities (Oreszczy, Lane, & Carr, 2010). As a result, networking tools that facilitate quick information sharing based on weak ties can be particularly relevant in the farming profession. Communities of British farmers on Twitter have formed around influential accounts and hashtags, the most popular of which is #AgriChatUK that has been used to organize national weekly Twitter conversation about British farming since April 2011. Influential accounts of British farming include media (e.g. @FarmersWeekly, @FarmersGuardian), the @FarmersOfTheUK managed by a different farmer every week and the @TheFarmingForum (national forum about farming).

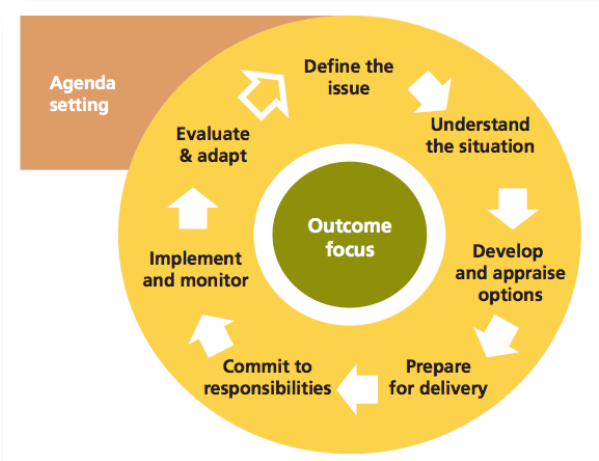


Figure 1:
The DEFRA Policy Cycle from Collier et al. (2010)

4.1. Illustrative analysis of farming and agricultural conversations on Twitter

The National Farmers Union is the main professional association for farmers in England and Wales with over 55,000 members and 300 branch offices. Trade associations and unions have an institutional role in collective representation and a natural interest in promoting views from the profession along with their own objectives. The central account @NFUTweets is followed by a wide variety of stakeholders many of which directly represent farmers and agricultural businesses (over 30K followers around the time of study).

Starting from a random sample of 1,000 followers of @NFUTweets, the aim of this exercise was to identify a group of candidate accounts of users that are more likely to post information relevant to farming. Using the daily tweeting frequency as the main criterion, we excluded accounts that tweeted infrequently and ones that tweeted abnormally and were likely to be automated or marketing accounts ⁴. We further refined and manually checked the list of 700 accounts before conducting a retrospective analysis of content produced by this sample over a period of time. This time frame was set to six months (starting from 1st August 2014), which yielded a dataset of 85,436 tweets; a reasonably large dataset that still remains within manageable limits for an exploratory analysis.

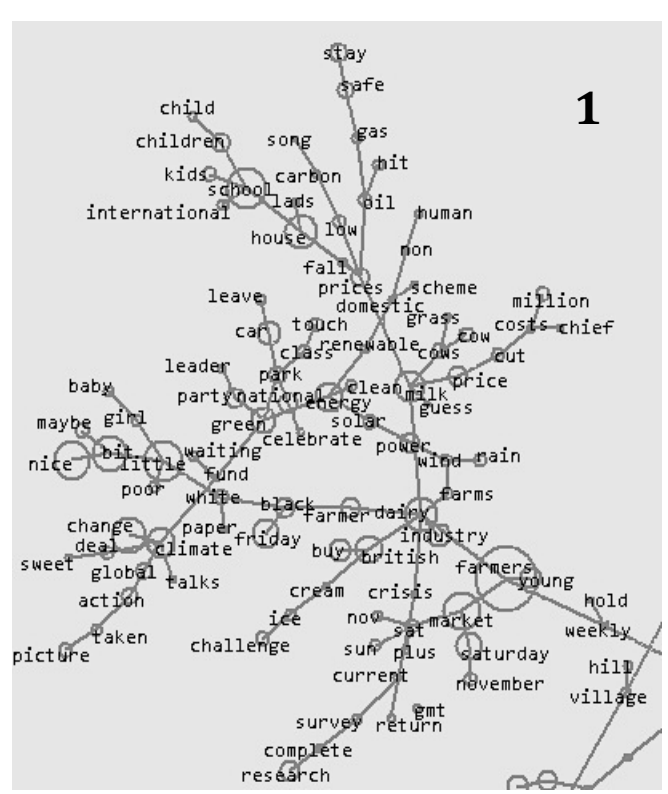
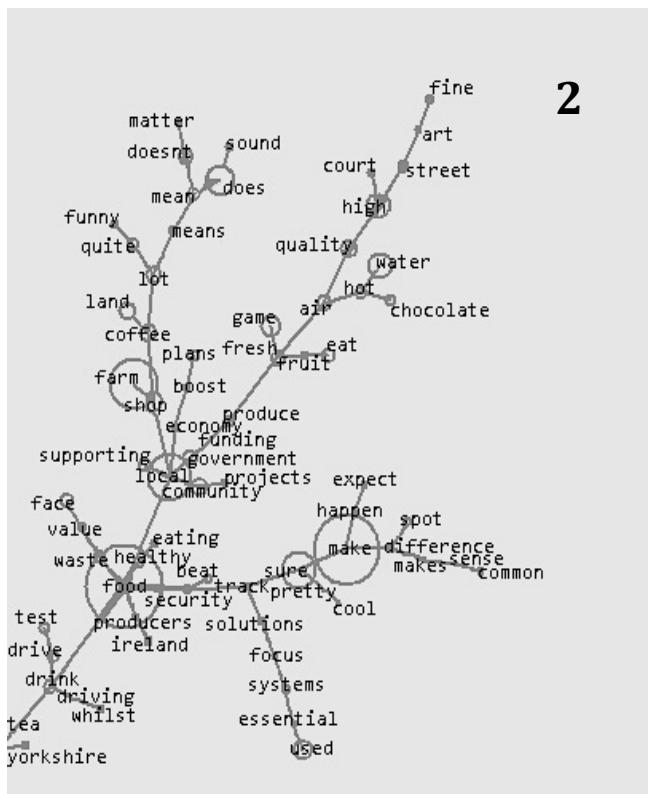
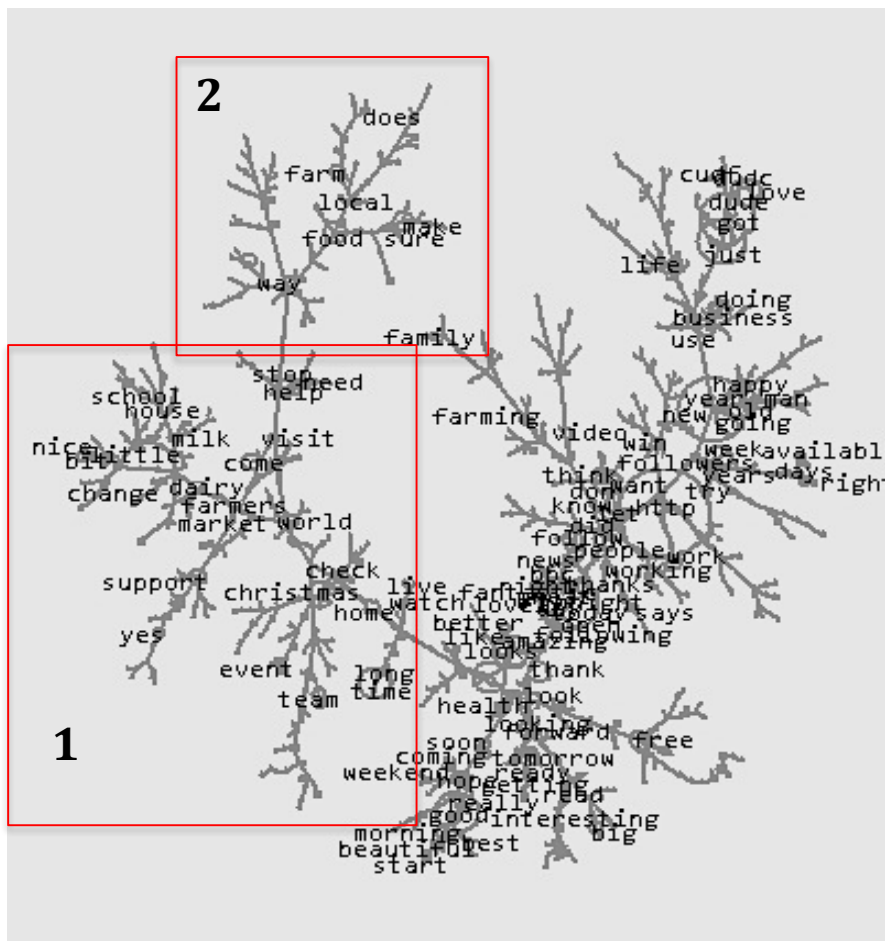
Figure 2 shows the resulting graphs used for this analysis as a probe to data acquisition. Visualizing the dataset of 85,436 tweets using cluster maps makes it clear that it is dominated by general talk such as you might expect to see in a random collection of everyday timelines (rightmost side of the overall term map in figure 2). For instance, there is a topical strand around the terms "happy", "new" and "year", as well as general conversational terms like "just", "thanks", "says", and so on. Given that the sampled 700 accounts only have in common a following relationship with @NFUTweets, it was expected that a large part of the visualized map would be generic everyday conversation. There are also however two separate, farming-relevant branches that make up the leftmost side of the overall cluster map (labeled as 1 and 2 in the figure). In fact, the collection of keywords used in each category serves to make a clear distinction between the two types of farming – dairy and arable farming – with each category relying on a different set of keywords in its expression.

The first of these branches (label 1) concerns dairy farming and displays a messy array of multiple different themes and topics. Most notably, there is a selection of key terms towards the center of the cluster: "farmer", "dairy", "cream", "milk", "British", and so on. These specific terms to dairy farming are situated closely to indirectly related topics such as fossil fuels, sustainability and climate change (with terms like "solar", "power", "clean", "energy", "green" and "party" for instance). Clearly for these users, dairy farming and renewable energy have some mutually connected relevance as topics, which was an interesting insight to explore further in discussions with policy makers.

The second key farming branch (label 2) concentrates on arable farming (i.e. terms like "land" and "food" as they appear alongside farming marketing terms like "shop" and "producers"), and shows a different type of conversation than the one expressed around dairy farming. One

⁴ At the first step, we retained in our dataset only accounts whose average tweets-per-day count was within one standard deviation from mean of the initial sample of 1,000 followers.

Figure 2: clusters maps of 85,436 tweets sourced from a random sample of 700 followers of @NFUTweets. Visualized with the help of Chorus Analytics (2017).



particularly strongly associated sub-topic is visible: terms like "economy", "government", "support", "local" and "community" demonstrate that the issue of government funding and local community are particularly interconnected with arable farming (as opposed to dairy farming for instance, which is surprising given national media coverage on falling milk prices). This sub-topic contains a more specific reaction of farmers' perceptions of how government funding is directed, how local communities are perceived to feature in agricultural policy, and different impacts between arable and dairy farming. Figure 3 visualizes the 630 tweets included in this sub-topic in the form of a word cloud.

There are also comparisons to be made across the two branches – dairy and arable farming – in terms of their broader conversational patterns. For instance, dairy farming appears as a 'rhizomatic' and messy aggregation of a diverse array of issues – this is visible in the messy sprouting structure of the dairy-farming cluster. In contrast, arable farming appears as a less voluminous conversation, presented with a few linear branches extending out from the central term "food". This suggests that the conversation is characterized as having fewer overall issues connected to it; either arable land farming is less discussed on Twitter (by the selected users) than dairy farming, or there are fewer sub-topics of arable farming that diversify the discussion. Exploring the differences between the tweeting practices around the topics of dairy and arable farming in this way may help to further characterize either conversation and thus help policy makers make better use of the data. These distinct topics and patterns of conversation can provide useful insight into everyday farming issues.



Figure 3:
Indicative word cloud of 630 tweets filtered from the cluster maps

4.2. Crowd acquisition: audiences, sources and frequency of monitoring

As many of DEFRA’s traditional stakeholders have developed a digital networking presence, the potential relevance of social media sources per se was rather uncontested by interview and workshops participants. To use an interviewee’s words: “*The government opening up more is positive. It can allow government to be seen less as about men hidden away in ivory towers and more about engaging directly with people*”. As a result, the identification of so many popular Twitter accounts in British farming was much encouraging for policy makers.

However, there were diverse opinions about the value of different sources and forms of social media input. Three main themes were identified in relation to crowd acquisition.

4.2.1. Sources and frequency of monitoring

It was mainly through accumulated experience that policy makers were developing their assessments of the value of social media within a dynamic digital landscape. Useful sources identified during interviews included a range of blogs, LinkedIn groups and Twitter hashtags or lists. Content sharing websites and comments below popular news articles were generally considered as less important.

The organization's experience with monitoring tools and sourcing of social content came from two different types of activities: trend-based and event-based. Monitoring popular trends and collecting updates for national campaigns or high-profile conversations was done centrally within the organization. For example, content was captured to oversee reactions against DEFRA's decision to implement a 'badger cull' initiative in 2012. Monitoring tools either managed internally or with the support of media companies would filter a large incoming flow of content and then produce ad hoc or periodic reports of trending content and influential contributors. Assessments of crowd acquisition were mostly difficult to make in trend-based summary reports.

While this type of central monitoring mostly extended traditional media briefings, a few of the department's agencies had adopted monitoring software for event-based needs. The most important context here were emergency events where DEFRA and its affiliated agencies need to facilitate timely communication of risks to the public. For example, the Environment Agency is the body responsible for handling emergencies related to natural disasters. Its main Twitter account had over 250K followers at the time of the study and was part of Twitter Alerts, the network's official warning system. Monitoring related to emergencies like floods provided a clearly defined set of keywords and timeframes for sourcing and interpreting content. This was achieved through a combination of flood-related hashtags, direct mentions from the public and scanning content from open sources. An interviewee involved in such activities noted: *"social media has been a revolution. It has helped incredibly with the Environment Agency incident management role and adds another dimension to a communicator's role. For regulation there is probably still a way to go"*.

It was generally understood that crowd acquisition outside the context of trends or events would have important resource, methodological and technical implications. Network-based analyses that overview snapshots of general conversations – like the analysis of the @NFUTweets followers – were seen as potentially useful in terms of broadening the sources and scope of crowd acquisition.

4.2.2. Social media audiences

Aligned with DEFRA's commitment to evidence-based policy, interview and workshop participants had a good understanding of traditional stakeholders, important influencers and the value of different public engagement activities. Nevertheless, assessments of the social media audience proved to be quite challenging when it came to finding the right "crowd" for a

policy issue. Policy makers usually had to make assumptions about how specific groups of professionals have a presence on social media, how they connect to each other, how they create content and whom they represent. When it came to personal use of social media, at least two interviewees expressed a personal rule of thumb as *'you need to know who to follow, and have your stakeholders in mind; is what they're saying useful?'*

Answers to these questions could be more straightforward for social networking groups (e.g. on LinkedIn) but less obvious on channels that support open information flows and dynamic conversations like Twitter, Facebook and blogs. Similar to emergency events, crowd acquisition outside specific conversations organized around hashtags was an issue policy makers found much harder to grasp. Twitter lists or keyword searches could act as a first step of filtering but still resulted in large amounts of unstructured content that did not include information about the audience. Limitations of audience understanding had an effect on the value of social media data as evidence, which was an important point of discussion in the assimilation phase. Despite these limitations, study participants were confident to identify the positive aspect of social monitoring to source opinions from "real" people or groups of the public that extend beyond stakeholders who make regular contributions to policy consultations. For example, interviewees with experiences in assessing input from the Red Tape Challenge, the UK government's crowdsourcing system, highlighted the importance of reaching more "real" people.

4.3. Crowd assimilation: forms of analysis and evidence assessments

4.3.1. Analysis and visualization methods

There was wide agreement that for input to be considered as useful in any form it had to be relevant and appropriately summarized. Social monitoring tools already used would normally produce daily or weekly reports to be circulated within the organization. Reports would usually be summarized based on popular content but the relevance was not always clear. Interview participants generally understood that open information channels like Twitter host a large amount of frequent, immediate and potentially relevant content, however the high "noise to signal" ratio was seen to compromise the potential value of insights gained. The direct applicability of sourced content was emphasized by an interviewee who does not use Twitter as: *"we already receive a lot of useless information through other channels e.g. email, so why seek to get more?"*. This was identified as a major challenge in all discussions related to crowd assimilation.

The cluster maps shown in figure 2 were developed as an example of summarizing large datasets and filtering content to explore participants' views on assimilation. Workshop participants were positive about the potential of advanced methods like this to quickly generate insights from large datasets that cannot be otherwise summarized (for example, they expressed interest to use cluster maps of more carefully constructed lists of users depending on engagement needs). It also served as an example of an analytical technique that was available in real time but could extend retrospectively into the past (up to 3,200 tweets per user based on Twitter's current restrictions on publicly available data). Other digital methods like sentiment or network analysis were used as alternative techniques. Reactions from policy

makers were mixed about their value, especially sentiment analysis with which many participants were already familiar. In a view alternative view to summarizing information, an interviewee outlined another challenge in crowd assimilation: “[social media] has maybe created a world in which information tends to be much more summarised, which means for example you can’t communicate the pros and cons of a complex issue. The danger is that everything becomes soundbites that aren’t helpful to a meaningful debate”.

4.3.2. Evidence limitations

The limitations of social media data were an issue that stimulated intense discussions in our study. For policy makers with training in social science and economics research, exploratory analytics from large datasets of unstructured content could not be used as “evidence” the same way as traditional methods. These participants were particularly concerned about three limitations of crowd assimilation: passive rather than active sampling of participants, sourcing content driven by keyword searches and partial or total lack of audience awareness.

This was not simply a matter of lacking demographic information about users but also the difficulties in resolving issues of sampling and representation. Social media users include a variety of domain experts as well as many users who are not experts but have a primary stake in policy topics. All these diverse users post content at different frequencies and with different intentions about whom they are talking to. As a result, monitoring content around keywords only captures the perspective of those users who decide to make a contribution within a specific timeframe, which inevitably leads to a “self-inclusion” perspective. Self-inclusion was evident when considering a snapshot of the state of a particular community like the cluster maps shown in figure 2. This was a more important limitation than sample sizes. For example, our study identified an estimated network of at least 10K Twitter users from the UK that tweet about issues relevant to farming. Representation issues were not only related to the fact that a potential audience of 10K Twitter users is a small proportion of an estimated total of 250K farmers and agricultural businesses in the UK (DEFRA Statistics, 2015). The issue was that, apart from a general awareness of their professional identity, there was no systematic information in tweets or account metadata about who these users are and what motivates them to contribute to specific discussions.

4.3.3. Evidence and the policy cycle

In theory, earlier stages of the policy making lifecycle might seem more suitable for incorporating social media input given its exploratory nature (see “issue definition” and “situation understanding” in figure 1). Our study however indicated high interest in social media content for commitment, implementation and evaluation activities as well (reactions to government policies or consultations in progress). This was particularly evident in the search for specific keywords within our large datasets that covered several months of tweeting and were related to issues where policy had already been implemented. The cluster maps identified reactions to important government policies but also other regional issues together with talk that was related to the everyday life of farming. The issue of evidence in the policy cycle remained an open one with an interviewee noting: “it is quite a difficult balance to

strike between getting meaningful insight through responses to relatively well formed ideas, and engaging early to allow people to influence how policy develops”.

Another interesting finding from the workshops was the mismatch between policy makers’ assumptions about what farmers would be talking about and what was actually said in the tweets. On some occasions, farming-related discussions on Twitter focused on high-profile topics that were expected to be an issue of concern within the profession (e.g. the price of milk, competitiveness of British farming or renewable energy). At the same time, participants were surprised that other issues expected to generate at least some reactions from social media users were almost absent from our datasets (e.g. DEFRA’s new national pollinator strategy or bovine tuberculosis policy). Although these results might be related to sampling limitations, they do point to crowd assimilation exercises containing content about issues that might not otherwise be evident to DEFRA’s policy teams.

5. Discussion and implications

This study explored how policy makers perceive the value of social media as policy evidence, and which capabilities support the needs of policy makers. In response to the first research question, we discuss policy makers’ assessments of the value of social media data compared with traditional sources of evidence. We then discuss the new capabilities that can support more systematic use of social media data in evidence-based policy. Finally, we discuss the study’s contribution and use of crowd capabilities as a theoretical framework.

5.1. Social media data as policy evidence

Evidence-based policy includes pragmatic approaches that can help policy makers understand “what works” and support institutional learning in practice (Cartwright & Hardie, 2012; Davies et al., 2001; Sanderson, 2002). Some work to date has been invested in considering how social media integrate in existing policy processes and workflows (Janssen & Helbig, 2016; e.g. Mergel & Bretschneider, 2013; Zavattaro & Sementelli, 2014). Our analysis – particularly from the interview data – point towards how social media might can reshape and even redefine the work of policy making. To summarize the main points of this discussion, table 1 presents a systematic comparison of social media data compared to traditional sources in evidence-based policy.

Policy makers made predominantly idiosyncratic assessments in their understanding of the value of social media data (Ingold & Monaghan, 2016). All interviewees – even those whose professional experiences with social media had been limited – agreed that social media has potential value and relevance. Yet it was also clear that those who were most engaged with social media as part of their work had done so through their own explorations (whether in the contexts of their work, their personal lives or both). Their reporting of the process of coming to understand the relevance of social media data was based on these decidedly subjective engagements, yet interviewees seemingly struggled to connect these insights to data acquisition in systematic ways. Rather, people voiced more quantitatively-oriented concerns around the status of social media data as “hard” evidence, which would more easily fit in with

current policy processes. Given the persisting attractiveness of social media data as an open and plentiful information source, one way to handle this disconnect would be to explore the possibility of adapted forms of policy building that are constructed on a more qualitative basis. This, we argue, might conceivably make better use of the available data, though an exploration on the practicalities and implications of such a transformative project extends beyond the present paper.

Table 1: Traditional sources vs. social media data in evidence-based policy

| | Traditional sources | Social media data |
|--------------------------------|---|---|
| Audience acquisition | <ul style="list-style-type: none"> – Established methods of data collection like surveys, focus groups, interviews etc. – Clear sampling and selection of participants – Audience information directly sought from participants or already known – Participants consent and contribute actively – Input can be private or public depending on the method | <ul style="list-style-type: none"> – Keyword and network-based approaches to sourcing content – Sampling driven by participants’ own engagement (self-inclusion) – Audience constructed according to each platform’s available metadata about participants – Participants contribute passively with implicit consent – Input mostly sourced from public spaces |
| Scalability of input | <ul style="list-style-type: none"> – Cost and time limitations in collecting feedback – Sourcing more input is usually resource-intensive | <ul style="list-style-type: none"> – Input available immediately – Input scalable to a large extent irrespective of the number of contributions |
| Relational capabilities | <ul style="list-style-type: none"> – Stable and long term relationships with stakeholders – Traditional influencers known and their input is anticipated | <ul style="list-style-type: none"> – Relationships more dynamically constructed within digital networks – Influencers might be more trend and topic-dependent |
| Assimilation methods | <ul style="list-style-type: none"> – “Hard” evidence through statistical relationships, econometric analyses and social science methods | <ul style="list-style-type: none"> – “Soft evidence” though engagement analytics, text mining and other textual and network-based analysis techniques |

To some extent, our observations show that these transformations are taking place in the context of trend or event-based sourcing where crowd acquisition is not equally dependent on in-depth audience understanding (e.g. when emergencies or popular campaigns take place). When in-depth audience understanding was necessary, policy makers were unable to make

solid assumptions about whom they are “listening to” due to the representation and self-inclusion issues of participants in large datasets that came from open platforms like Twitter.

Yet, when it comes to assimilation, there are clear benefits to this this new type of “soft” evidence produced via engagement analytics, text mining (e.g. cluster maps) and other textual and network-based analysis techniques. First, social monitoring can lead to the discovery of conversations that might otherwise be outside spaces of visibility. Second, content from social media that is collected passively allows the acquisition of a much larger and spontaneous pool of reactions. It was recognized that the scale and cost-effectiveness of such analyses could not be achieved using traditional means – results were available almost at real time and required far less resources than any other method of collecting evidence. These propositions are significant given regulatory priorities to develop less direct and more cost-effective forms of regulation – at least as signaled by the UK Government (2015).

5.2. New capabilities for policy makers

Our analysis points to the fact that the acquisition of large amounts of content from social media does not always equate to the construction of a “crowd” that can provide useful feedback in policy development. Although this is a generally established conclusion of big data research (e.g. Boyd & Crawford, 2012; Schintler & Kulkarni, 2014), our study provides a specific case of how these limitations were interpreted by policy makers. From an assimilation perspective, crowdsourced contributions can provide useful insights when they are analyzed flexibly and appropriately summarized. Table 1 clearly illustrates that, compared to more established methods, the technical complexities of social media analysis require new types of ad hoc data science capabilities (e.g. tools and skills for merging/manipulating large datasets). For example, digital methods like hashtag analysis are generally established in social science research (e.g. Bruns, 2012; Highfield et al., 2013) but have not been applied in commercial or policy applications that tend to focus on summary statistics and virality measures.

On a broader scale, the needs of policy makers from both the acquisition and the assimilation of social media content point to the requirement for developing tools with more advanced network-feedback features. These features emerge from the needs to understand how information networks evolve when different groups of the public engage in discussions and reflect the state of particular communities (see audience acquisition in table 1). Such features can complement existing keyword-based searches that show how social media users interact around issues related to the work of government (but not focusing on how they “like” or “share” content).

To elaborate on this important information requirement, we used cluster maps to visualize large exploratory Twitter datasets and discover how conversations evolve. This method allowed narrowing down conversations to specific subsets of tweets that contain keywords of interest or were posted at specific times. Policy makers identified the necessity of making more systematic use of account metadata and even attempting to link findings from social media to a statistically supported segmentation of a target population (e.g. Twitter users that commented on milk prices or climate change). Other tools that aggregate content from digital

sources have developed context-specific solutions so that audience information can make sense in more focused types of applications (e.g. the Altmetrics service used to track reactions to academic articles). Social media analysis tools, whether commercial solutions or research tools like Chorus Analytics, might not be able to adapt to that level of accuracy unless social media platforms make information about their users more available in some summary form within account metadata (e.g. location for all tweets/users, demographic information etc.). These “engagement analytics” are gradually being released for platforms like Twitter and Facebook.

As a further policy implication, the findings point to the need to consider how information flows between governments and the public that used to be private are now becoming more ubiquitous in open digital spaces. As part of developing new capabilities, there needs to be consideration of how the sourcing of contributions from social media can be communicated to the public. Bekkers et al. (2013) place this in the context of avoiding surveillance but it can also be seen as an opportunity to reiterate commitment to listening and developing new forms of engagement that can support inclusivity in policy making. Groups of the public like farmers can be encouraged to organize and connect on social media so that their contributions can be sourced. This can be a task supported by intermediary organizations like professional associations, trade unions or other representation bodies that can facilitate professional networking and raise the profile of input from their own audiences. We identified promising signs of this happening in our analysis of farming conversations on Twitter as regional and national groups were coming together motivated by organizations and community entrepreneurs.

5.3. Theoretical implications

The intended contribution of this paper lies in understanding social media data in evidence-based policy as summarized in table 1. The concept of crowd capabilities – adapted from Prpić et al. (2015) – was used to structure the examination and consider issues around sourcing content from unknown audiences (acquisition) and the extent to which this content can be transformed and used appropriately (assimilation). As a theoretical concept, crowd capabilities draw attention to the configuration of resources needed to source and incorporate data from digital audiences. Acquisition and assimilation work as separate but interrelated processes – together they make evident in our analysis that social monitoring was much more than a technical problem of scraping, mining and aggregating data from as many sources as possible. Instead, it is a strategic challenge to be considered as part of capabilities development (e.g. data science skills, new tools and methods of analysis, adaptations of policy making processes). The crowd capabilities analysis establishes the link to the organizational literature by extending previous work on absorptive capacity and processes of value creation in organizations (Culnan et al., 2010; Ooms et al., 2015; Roberts et al., 2012). Furthermore, the analysis places previous work on social media capabilities in a public sector context where there are different challenges to impression management and corporate communications (Benthaus et al., 2016; Schniederjans et al., 2013).

Crowd capabilities is a starting point of analytical models related to crowdsourcing applications that can be further developed and elaborated on in a public sector context (e.g.

Chiu, Liang, & Turban, 2014; Estelles-Arolas & Gonzalez-Ladron-de-Guevara, 2012). Within this, crowd capabilities draw on acquisition and assimilation, which are the first two components identified by Zahra and George (2002) – the others are transformation (reshape existing processes to fit new knowledge) and exploitation (creating new competencies from absorbed knowledge). The transition from assimilation to transformation and eventually exploitation could be considered in more depth; for example, the relevance of social media data in developing communication and engagement capabilities. The relationship between social media content and audiences has already been seen as an iterative process of monitoring and engaging (Mergel, 2013; Panagiotopoulos et al., 2015). Our study reiterates this conclusion and shows how social media analysis is fundamental to understanding current discourses within digital communities.

6. Conclusion

DEFRA's broad remit and commitment to evidence-based policy provided a stimulating case to examine the value of social media data. Organized in the form of an exploratory case study, the research examined the potential of and capabilities needed to integrate social media data into evidence-based policy. The findings should be taken into account with considerations to their contextual limitations. DEFRA's policy cycle and commitment to evidence-based policy might not always be the most fruitful ground for experimental approaches to new data sources. Furthermore, the identification of certain themes about social media sources by study participants cannot be seen as a complete overview of perceptions within this large government department or across the UK government. To some extent, the inevitable selection of participants with willingness to engage with this research hints to a more positive version of the findings or might exclude those less familiar with the topic at the time of the study.

We further need to consider that environmental and agricultural policies tend to attract a large number of views from diverse publics. In other policy topics, there might not be that much potential insight to source or the audience is more uniform, hence making assumptions about the value of social media data more straightforward. Respectively, in other policy topics, members of the public might have motivations to put forward a high volume of contributions on social media, especially for polarizing issues. Future work could look further into the conditions under which social media provide a useful source of input with respect to privacy, ethical and risk management considerations.

Finally, research on crowd capabilities could be extended to explore how content creation and monitoring processes within government organizations are interrelated and how they contribute towards better policy delivery. Alternatively, it would be interesting to ask social media communities directly about: (1) how valuable they think channels like Twitter are for discussing issues within their professional interests and (2) the extent to which they would be willing to consider their contributions as part of policy making input.

Acknowledgments

The authors gratefully acknowledge interview and workshops participants from the Department for Environment, Food & Rural Affairs (DEFRA). In particular, our gratitude goes to the members of the Legislative Strategy, Capability and Better Regulation Team for facilitating the research and sharing their experiences on multiple occasions.

We further gratefully acknowledge members of the project team: Chris Taylor for his involvement in data collection and Tim Cribbin and Julie Barnett for their input as part of Chorus Analytics. Finally, we are thankful to Simon Haley and his colleagues from AgriChatUK for their input in the project and exceptional work as community leaders. The research was supported via the NEMODE network, Research Councils UK (New Economic Models in the Digital Economy, EP/K003542/1). An earlier version of this paper was presented at the 76th Annual Meeting of the Academy of Management, 5-9 August 2016 in Anaheim, California, USA (Panagiotopoulos et al., 2016).

References

- Barnett, J., Burningham, K., Walker, G., & Cass, N. (2012). Imagined publics and engagement around renewable energy technologies in the UK. *Public Understanding of Science*, 21(1), 36–50.
- Bekkers, V., Edwards, A., & de Kool, D. (2013). Social media monitoring: Responsive governance in the shadow of surveillance? *Government Information Quarterly*, 30(4), 335–342.
- Benthaus, J., Risius, M., & Beck, R. (2016). Social media management strategies for organizational impression management and their effect on public perception. *The Journal of Strategic Information Systems*, 25(2), 127–139.
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679.
- Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- British government. (1999). Modernising Government white paper. Retrieved January 10, 2017, from <http://www.archive.official-documents.co.uk/document/cm43/4310/4310.htm>
- Brooker, P. D., Barnett, J., Cribbin, T., Lang, A., & Martin, J. (2014). User-driven data capture: locating and analysing twitter conversation about cystic fibrosis without keywords. In *SAGE Research Methods Cases*. London, United Kingdom
- Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection - harnessing the web for public health surveillance. *New England Journal of Medicine*, 360(21), 2153–2157.
- Bruns, A. (2012). How Long is a Tweet? Mapping Dynamic Conversation Networks on Twitter using Gawk and Gephi. *Information, Communication & Society*, 15(9), 1323–1351.
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: a practical guide to doing it better*. Oxford: Oxford University Press.
- Castelló, I., Morsing, M., & Schultz, F. (2013). Communicative Dynamics and the Polyphony of Corporate Social Responsibility in the Network Society. *Journal of Business Ethics*, 118(4), 683–694.
- Charalabidis, Y., Loukis, E. N., Androutsopoulou, A., Karkaletsis, V., & Triantafillou, A. (2014).

- Passive crowdsourcing in government using social media. *Transforming Government: People, Process and Policy*, 8(2), 283–308.
- Chiu, C.-M., Liang, T.-P., & Turban, E. (2014). What can crowdsourcing do for decision support? *Decision Support Systems*, 65, 40–49.
- Chorus (2017). Chorus analytics data harvesting and visual analytics suite. Retrieved November 5, 2017, from <http://chorusanalytics.co.uk/>
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), 128–152.
- Collier, A., Cotterill, A., Everett, T., Muckle, R., Pike, T., & Vanstone, E. (2010). *Understanding and influencing behaviours: a review of social research, economics and policy making in DEFRA*.
- Cribbin, T. 2009. Visualising the structure of document search results: A comparison of graph theoretic approaches. *Information Visualization*, 9(2): 83–97.
- Criado, J. I., Sandoval-Almazan, R., & Gil-Garcia, J. R. (2013). Government innovation through social media. *Government Information Quarterly*, 30(4), 319–326.
- Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value. *MIS Quarterly Executive*, 9(4), 243–259.
- Davies, H., Nutley, S., & Smith, P. (2001). *What Works?: Evidence-Based Policy and Practice in Public Services*. Bristol: Policy Press.
- DEFRA. (2014). *Independent Farming Regulation Task Force Implementation Group*. Retrieved February 22, 2015, from <https://www.gov.uk/government/publications/farming-regulation-task-force-implementation-group-final-assessment-of-progress>.
- DEFRA. (2015). DEFRA Business Plan. Retrieved February 22, 2015, from <http://transparency.number10.gov.uk/business-plan/10>
- DEFRA Statistics. (2015). Agriculture in the United Kingdom. Retrieved February 22, 2015, from <https://www.gov.uk/government/collections/agriculture-in-the-united-kingdom>
- Dong, J. Q., & Wu, W. (2015). Business value of social media technologies: Evidence from online user innovation communities. *The Journal of Strategic Information Systems*, 24(2), 113–127.
- Estelles-Arolas, E., & Gonzalez-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200.
- Evans, H., Ginnis, S., & Bartlett, J. (2015). *#SocialEthics a guide to embedding ethics in social media research*, Ipsos MORI.
- Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, 57(6), 74–81.
- Ferro, E., Loukis, E. N., Charalabidis, Y., & Osella, M. (2013). Policy making 2.0: From theory to practice. *Government Information Quarterly*, 30(4), 359–368.
- Haley, S. (2013). *The value of social media for the rural industry (MSc dissertation)*. Harper-Adams University.
- Head, B. W. (2008). Three Lenses of Evidence-Based Policy. *Australian Journal of Public Administration*, 67(1), 1–11.
- Highfield, T., Harrington, S., & Bruns, A. (2013). Twitter as a Technology for Audiencing and Fandom. *Information, Communication & Society*, 16(3), 315–339.
- Hoffman, D. L., & Fodor, M. (2010). Can You Measure the ROI of Your Social Media Marketing? *Sloan Management Review*, 52(1), 41–49.

- Howlett, M. (2009). Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration*, 52(2), 153–175.
- Ingold, J., & Monaghan, M. (2016). Evidence translation: an exploration of policy makers' use of evidence. *Policy & Politics*, 44(2), 171–190.
- Janssen, M., & Helbig, N. (2016). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*, (advanced online publication).
- Kasabov, E. (2008). The challenge of devising public policy for high-tech, science-based, and knowledge-based communities: evidence from a life science and biotechnology community. *Environment and Planning C: Government and Policy*, 26(1), 210–228.
- Katona, Z., & Sarvary, M. (2014). B2B Social Media - "It's Communication, not Marketing." *California Management Review*, 56(3), 142–156.
- Kidd, J. (2011). Enacting Engagement Online: Framing Social Media Use for the Museum. *Information Technology & People*, 24(1), 64–77.
- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social Media? Get Serious! Understanding the Functional Building Blocks of Social Media. *Business Horizons*, 54(3), 241–251.
- Kostkova, P., Szomszor, M., & St. Louis, C. (2014). #swineflu. *ACM Transactions on Management Information Systems*, 5(2), 1–25.
- Krippendorff, K. (2013). *Content Analysis: an Introduction to Its Methodology* (3rd ed.). California, USA: SAGE Publications.
- Kum, H.-C., Joy Stewart, C., Rose, R. A., & Duncan, D. F. (2015). Using big data for evidence based governance in child welfare. *Children and Youth Services Review*, 58, 127–136.
- Lodge, M., & Wegrich, K. (2014). Crowdsourcing and regulatory reviews: A new way of challenging red tape in British government? *Regulation & Governance*, 9(1), 30–46.
- Marston, G., & Watts, R. (2003). Tampering with the evidence: a critical appraisal of evidence-based policy-making. *The Drawing Board: An Australian Review of Public Affairs*, 3(3), 143–163.
- Mergel, I. (2012). The social media innovation challenge in the public sector. *Information Polity*, 17(3-4), 281–292.
- Mergel, I. (2013). A framework for interpreting social media interactions in the public sector. *Government Information Quarterly*, 30(4), 327–334.
- Mergel, I., & Bretschneider, S. I. (2013). A Three-Stage Adoption Process for Social Media Use in Government. *Public Administration Review*, 73(3), 390–400.
- Misuraca, G., Codagnone, C., & Rossel, P. (2013). From Practice to Theory and back to Practice: Reflexivity in Measurement and Evaluation for Evidence-based Policy Making in the Information Society. *Government Information Quarterly*, 30, S68–S82.
- Murthy, D., Gross, A., & Pensavalle, A. (2016). Urban social media demographics: An exploration of Twitter use in major American cities. *Journal of Computer-Mediated Communication*, 21(1), 33-49
- Newman, J., Cherney, A., & Head, B. W. (2017). Policy capacity and evidence-based policy in the public service. *Public Management Review*, 19(2), 157–174.
- Obama, B. (2009). Transparency and Open Government: Memorandum for the Heads of Executive Departments and Agencies.
- Ooms, W., Bell, J., & Kok, R. A. W. (2015). Use of Social Media in Inbound Open Innovation: Building Capabilities for Absorptive Capacity. *Creativity and Innovation Management*, 24(1), 136–150.

- Oreszczyn, S., Lane, A., & Carr, S. (2010). The role of networks of practice and webs of influencers on farmers' engagement with and learning about agricultural innovations. *Journal of Rural Studies*, 26(4), 404–417.
- Panagiotopoulos, P., Bowen, F. E., & Brooker, P. (2016). Social Monitoring in Government: Developing “Crowd” Capabilities in Evidence-Based Policy. In *Academy of Management Proceedings* (Vol. 2016, No. 1, p. 16629). Academy of Management
- Panagiotopoulos, P., Shan, L. C., Barnett, J., Regan, Á., & McConnon, Á. (2015). A framework of social media engagement: Case studies with food and consumer organisations in the UK and Ireland. *International Journal of Information Management*, 35(4), 394–402.
- Pang, M.-S., Lee, G., & DeLone, W. H. (2014). In public sector organisations: a public-value management perspective. *Journal of Information Technology*, 29(3), 187–205.
- Parsons, W. (2002). From Muddling Through to Muddling Up - Evidence Based Policy Making and the Modernisation of British Government. *Public Policy and Administration*, 17(3), 43–60.
- Picazo-Vela, S., Gutiérrez-Martínez, I., & Luna-Reyes, L. F. (2012). Understanding risks, benefits, and strategic alternatives of social media applications in the public sector. *Government Information Quarterly*, 29(4), 504–511.
- Prpic, J., & Shukla, P. (2014). The Contours of Crowd Capability. In *2014 47th Hawaii International Conference on System Sciences* (pp. 3461–3470). IEEE.
- Prpić, J., Shukla, P. P., Kietzmann, J. H., & McCarthy, I. P. (2015). How to work a crowd: Developing crowd capital through crowdsourcing. *Business Horizons*, 58(1), 77–85.
- Prpić, J., Taeihagh, A., & Melton, J. (2015). The Fundamentals of Policy Crowdsourcing. *Policy & Internet*, 7(3), 340–361.
- Roberts, N., Galluch, P., Dinger, M., & Grover, V. (2012). Absorptive Capacity and Information Systems Research: Review, Synthesis, and Directions for Future Research. *MIS Quarterly*, 36(2), 625–648.
- Sanderson, I. (2002). Evaluation, Policy Learning and Evidence-Based Policy Making. *Public Administration*, 80(1), 1–22.
- Schintler, L. A., & Kulkarni, R. (2014). Big Data for Policy Analysis: The Good, The Bad, and The Ugly. *Review of Policy Research*, 31(4), 343–348.
- Schlagwein, D., & Hu, M. (2017). How and why organisations use social media: five use types and their relation to absorptive capacity. *Journal of Information Technology*, 32(2), 194–209.
- Schniederjans, D., Cao, E. S., & Schniederjans, M. (2013). Enhancing financial performance with social media: An impression management perspective. *Decision Support Systems*, 55(4), 911–918.
- Seawright, J., & Gerring, J. (2008). Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Political Research Quarterly*, 61(2), 294–308.
- Sutton, J. N. (2009). Social Media Monitoring and the Democratic National Convention: New Tasks and Emergent Processes. *Journal of Homeland Security and Emergency Management*, 6(1).
- UK Government. (2015). Regulation Reform. Retrieved April 1, 2015, from <https://www.gov.uk/government/topics/regulation-reform>
- Urquhart, C., & Vaast, E. Building social media theory from case studies: A new frontier for IS research, ICIS 2012 Proceedings (2012).
- Walker, G., Cass, N., Burningham, K., & Barnett, J. (2010). Renewable energy and sociotechnical change: imagined subjectivities of “the public” and their implications. *Environment and Planning A*, 42(4), 931–947.

- Wastell, D. G. (2006). Information systems and evidence-based policy in multi-agency networks: The micro-politics of situated innovation. *The Journal of Strategic Information Systems*, 15(3), 197–217.
- Westerman, S., & Cribbin, T. 2000. Mapping semantic information in virtual space: dimensions, variance and individual differences. *International Journal of Human-Computer Studies*, 53(5), 765–787
- Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization*, 26(1-2), 1–12.
- Williams, M. L., Edwards, A., Housley, W., Burnap, P., Rana, O., Avis, N. et al. (2013). Policing cyber-neighbourhoods: tension monitoring and social media networks. *Policing and Society*, 23(4), 461-481.
- Yin, R. K. (2014). *Case Study Research: Design and Methods* (5th ed.). Thousand Oaks: SAGE Publications.
- Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review*, 27(2), 185–203.
- Zavattaro, S. M., & Sementelli, A. J. (2014). A critical examination of social media adoption in government: Introducing omnipresence. *Government Information Quarterly*, 31(2), 257–264.

Dr Panos Panagiotopoulos is a Senior Lecturer in Information Management (Associate Professor) at Queen Mary University of London. His research examines the impact of information technology on political and administrative processes, with focus on capabilities, public engagement, social media and digital research methods.

Professor Frances Bowen is the Dean of the School of Business and Management and Professor of Innovation Studies at Queen Mary University of London. She specializes in corporate strategy, environmental governance and information-based approaches to regulation. Prof Bowen has extensive experience advising companies and regulatory bodies in the UK and Canada in a wide range of topics.

Dr Phillip Brooker is a Lecturer in Sociology at the University of Liverpool. Part of Phillip's work is motivated by an interest in developing innovative methodological approaches for the emerging field of social media analytics. Currently, Phillip is exploring the potential for computer programming to feature in core social science research methods training (particularly for research involving digital data).