

From Open Source Communications to Knowledge

Alun Preece^a, Colin Roberts^a, David Rogers^a,
Will Webberley^a, Martin Innes^a,
Dave Braines^b

^aCrime and Security Research Institute, Cardiff University, Cardiff, UK

^bEmerging Technology Services, IBM United Kingdom Ltd, Hursley Park, Winchester, UK

ABSTRACT

Rapid processing and exploitation of open source information — including social media sources — in order to shorten decision-making cycles, has emerged as an important issue in intelligence analysis in recent years. Through a series of case studies and natural experiments, focussed primarily upon policing and counter-terrorism scenarios, we have developed an approach to information foraging and framing to inform decision making, drawing upon open source intelligence — in particular Twitter, due to its real-time focus and frequent use as a carrier for links to other media. Our work uses a combination of natural language (NL) and controlled natural language (CNL) processing to support information collection from human sensors, linking and schematising of collected information, and the framing of situational pictures. We illustrate the approach through a series of vignettes, highlighting (1) how relatively lightweight and reusable knowledge models (schemas) can rapidly be developed to add context to collected social media data, (2) how information from open sources can be combined with reports from trusted observers, for corroboration or to identify conflicting information; and (3) how the approach supports users operating at or near the tactical edge, to rapidly task information collection and inform decision-making. The approach is supported by bespoke software tools for social media analytics and knowledge management.

Keywords: sensemaking; situation awareness; open source information; natural language; controlled natural language

1. INTRODUCTION

The widespread uptake of online social media has led to a transformation of the open source intelligence domain.¹ While ‘traditional’ media such as newsfeeds continue to play a role, they have become absorbed into the wider space of communicative channels dominated by the commercial social networks including Facebook, Instagram and Twitter. Twitter in particular, because of its emphasis on short, easily-consumable, and timely posts, has emerged as a key source of real-time ‘breaking’ information, often serving as a carrier of metadata to other channels, such as links to mainstream media news items.² The real-time public dissemination of information (and misinformation) via social media, particularly Twitter, is having significant impact on the practice of intelligence analysis as well as related areas of police investigation³ and management of disasters and large-scale events.^{2,4} Consequently, the ability to rapidly process and exploit open source information in the context of high-tempo decision cycles has grown in significance in recent years.

While technology plays a key element in open source intelligence exploitation, especially in the area of social sensing,⁵ a tendency to ‘oversell’ its capabilities has been observed^{6,7} leading to a view that ‘glass box’ rather than ‘black box’ approaches are required in order to gain robust situational understanding from open source media.^{8,9} For example, an analysis of the failure of Google’s ‘Flu Trends’ algorithm in early 2013 pointed to a lack of transparency in both algorithms and social media platforms as being key factors in undermining reliability and replicability of analysis results over time.⁷ This issue is even more acute when the results of open source data analyses needs to be interpreted rapidly by decision makers without recourse to personnel with a high degree of technical knowledge of the analytics techniques: this requirement ‘to reduce the cycle time and manpower requirements for analysis and use of large data sets’ was characterised by the US Department of Defense as Data-to-Decisions (D2D).¹⁰

Send correspondence to A.Preece, E-mail: PreeceAD@cardiff.ac.uk, Tel: +44 (0) 29 2087 4653

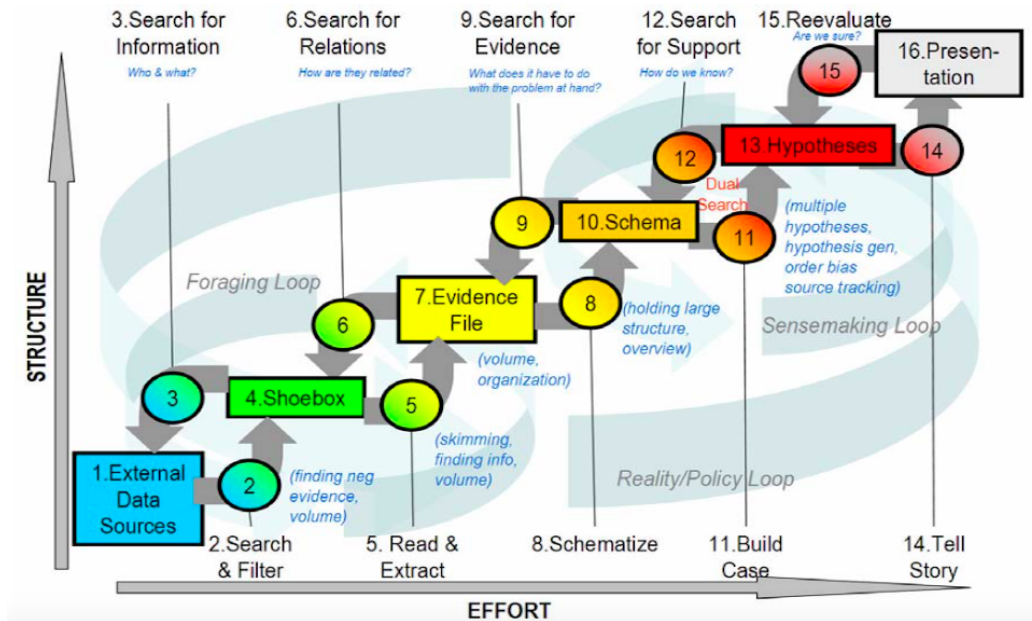


Figure 1. Pirolli and Card's model of the sensemaking loop for intelligence analysis¹²

This paper considers 'glass box' approaches for rapid exploitation of open source information in situational understanding, drawing on a number of case studies and experiments. Central to the approach is the use of relatively human-friendly information and knowledge representations to help frame and contextualise the results of analyses of social media data. A second key element of the approach is to combine information acquired from open sources with reports from trusted observers, to build a richer and corroborated picture of events. A third distinguishing aspect of the approach is to support users operating at or near the tactical edge¹¹ in two respects: tasking information collection and informing decision-making.

The paper is organised as follows: Section 2 sets the context of the proposed approaches in terms of background and related work; Section 3 describes the key elements of the approach; Section 4 presents a series of cases illustrating the application of the approach; finally, Section 5 concludes the paper with discussion and pointers to areas of future work.

2. BACKGROUND AND RELATED WORK

Intelligence analysis is a widely varied task domain in which Pirolli and Card's empirically-based work has been highly influential in structuring the processes involved.¹² They defined the sensemaking process as a set of interconnected loops, shown in Figure 1. The *foraging loop*, in which data is gathered from the external environment and assembled into a body of evidence, intersects with the *sensemaking loop*, in which schematised evidence is connected to hypotheses and cases are built to inform decision making. Progression of the process from bottom-left to top-right in the figure reflects increasing structuring and refinement of the information products created. The loops and 'backward' steps denote cycles of feedback and refinement throughout the model.

Open source information provides the sensemaking process with an important type of external data source at the bottom-left of the figure. Moreover, the vast amount of available data, especially in terms of social media, presents enormous challenges in enacting the foraging loop. In terms of the automatic processing of open source data, there has been considerable work done on event detection using social media streams, particularly Twitter.^{2, 4, 13, 14} However, the collection and processing of open source data to derive higher-level information products beyond the shoebox is fundamentally one of human-computer collaboration (HCC)¹⁵ and cannot be fully automated. Analysts are increasingly well versed in modern collaboration environments and social media, and systems are emerging that seek to combine the benefits of these approaches with existing software tools

and processes for structuring and supporting tactical intelligence analysis. A recent example of this seeks to enable analysts to identify the decision-relevant data scattered among databases and the mental models of other personnel by employing familiar social media-style collaboration techniques.¹⁶ There is some evidence to indicate that, not only is it useful to collaborate within the same analyst team, but when collaboration is extended to the crowd and mediated by an intelligent software agent, the outcome of the intelligence analysis can be greatly improved.¹⁷

In the context of current concerns regarding transparency in big data approaches,⁷ clearly a significant issue in the use of all open source intelligence concerns the potential for misinformation.¹⁸ There is work being done to mitigate these risks¹⁴ but it has also been observed that patterns of misinformation flow are often extremely valuable in terms of situational awareness: for example, rumouring is often a form of coordinated activity which needs to be countered.⁴

Much prior work focusses on the use of text-based information in social media, but often there is considerable value in imagery data both attached to tweets on Twitter and on imagery-centric platforms such as Instagram and Vine. In this sense, social media is a source of both hard and soft data, leading to interesting challenges in data fusion.^{19,20} The application of image processing to extract key features from these sources — particularly common symbols and objects, as well as face recognition — offers considerable potential but remains a hard problem. In the near term, processing such data effectively in another aspect of sensemaking that requires HCC.

There is a trade-off in supporting the progressive development of information products from bottom-left to top-right in Pirolli and Card’s model, in that the lower-level kinds of product need to favour machine-processability — not least to make sense of large volumes of data or to extract ‘needles from haystacks’ — while the higher-level kinds of products need to favour human-processability — to allow users with little or no technical training to exploit the products in robust decision making. As noted above with the ‘Google Flu Trends’ example, a disconnect between the lower and higher-levels of product can result in broken data-to-decision chains. The remainder of this paper considers approaches to maintain connectivity between information products across the whole sensemaking process, while attempting to balance machine and human processability of those products, and operating in an open and transparent ‘glass box’ manner.

3. APPROACH

The approach takes natural language (NL) as the basis for maintaining connectivity between information products across the whole sensemaking process, given that NL comprises a significant amount of open source input to the process as well as the major part of output products, for example, in the form of cases and presentation briefings). Traditionally, computational support in the middle stages of the process relies on products being represented in a form that is readily machine-processable; open standards such as XML*, RDF† and JSON‡ are often used for this. However, another approach is to use a *controlled natural language* (CNL) as an alternative to these traditional representations. CNLs have been developed for a variety of purposes including easing human-to-human communication and facilitating human-machine communication.²¹ In the latter case, CNLs are typically designed to make machine-processability as straightforward as possible — for example, by minimising potential for ambiguity — while remaining relatively easy for humans to understand. Imagery can be linked with CNL in similar ways to representations such as XML, RDF and JSON, for example through embedded objects or hyperlinks.

Considering an instance of the sensemaking process where external data sources are predominantly text-based social media (possibly with linked imagery), the relative dominance of human vs machine processability is shown in Figure 2. The implication here is that NL dominates the lowest and highest levels of information product, while CNL is best employed in the middle of the process. Moreover, there needs to be machine support to process NL into CNL and vice versa. At the lower end, in foraging, there is a need to process NL input into CNL for higher-level machine processing, and also a need to process machine-generated CNL searches and information

*<https://www.w3.org/XML/>

†<https://www.w3.org/RDF/>

‡<http://www.ecma-international.org/ecma-262/6.0/>

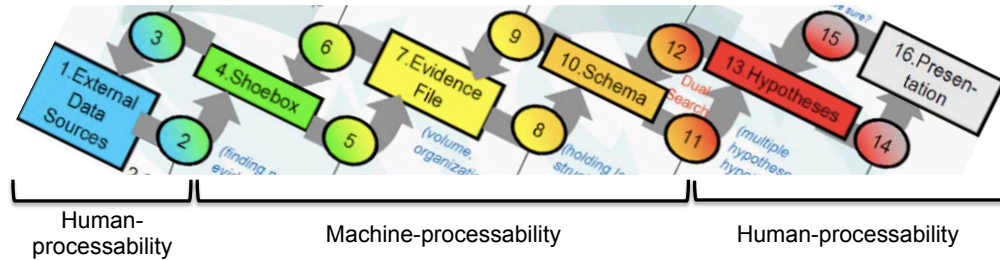


Figure 2. Relative dominance of human vs machine processability of information products in the sensemaking process

requests (step 3 of Pirolli and Card’s model) into NL in cases where humans are to be tasked to perform such searches and requests (for example, by instructing patrols or issuing crowdsourcing calls).

Like the traditional representations, a CNL provides a means to define information *models* as well as structured *instance data*. Models can serve as *schemas* for instance data or *ontologies* intended to describe some domain of knowledge (but not necessarily to be instantiated with instance data). Model and instance elements collectively form a knowledge base. The CNL used in this work is ITA Controlled English (CE)²² which offers approximately the same expressivity in terms of information modelling as the Web Ontology Language (OWL)²³ and is available in both Java (full[§]) and JavaScript (lightweight[¶]) open source implementations.

As an example, consider the tweet shown in Figure 3. The key entities mentioned here are an instance of an organisation (Blue Group), a location (Central Square) and a kind of event (protest). A CE model of this domain would include the concepts `organisation`, `location`, `event`, and `protest`, together with the information that a `protest` is a kind of `event`. Instances would include `Blue Group` and `Central Square`. In principle these could be discovered through named entity recognition in natural language processing (NLP)^{||} though in this case both entities would probably already be known *a priori*.



Figure 3. An example tweet reporting an event

Simply collecting tweets such as this one into a foraging shoebox would be unmanageable except in very small numbers. CE models offer various ways to help contextualise such social media fragments. One way is to perform NLP to associate the input media with related knowledge base elements, in this case by detecting that the tweet refers to `Blue Group` (an `organisation`), `Central Square` (a `location`) and a `protest` (an `event`). This information, together with metadata about the tweet — the poster (together with any knowledge of their reliability), the time, GPS coordinates if the tweet is geotagged, and so on — provides a substantial amount of context allowing the tweet to form part of a larger situational picture (for example, activities of the Blue Group, currently-known protests, or disruption in the Central Square area). Part of assembling that picture would likely also involve seeking corroboration unless the tweeter was a trusted source.

Going further, NLP could generate CNL statements as an attempt to interpret the content of the tweet where possible. In the example this would result in something like the following (where additional information such as the time and poster details are omitted for brevity):

[§]<https://github.com/ce-store/>

[¶]<http://cenode.io>

^{||}For example, using <http://nlp.stanford.edu/software/CRF-NER.shtml>

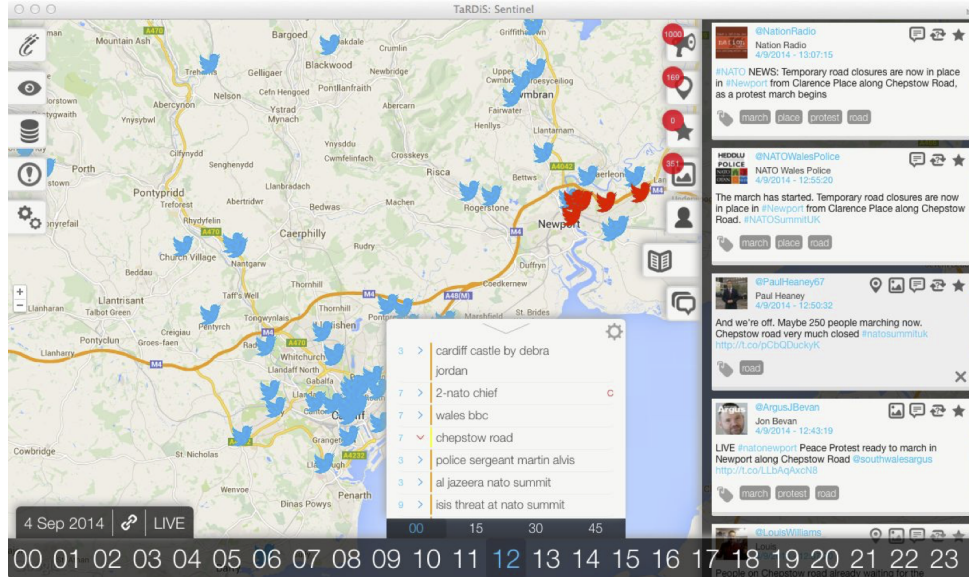


Figure 4. A screenshot of the Sentinel social media analytics tool⁴

there is a protest named ‘Central Square protest’ that involves the group ‘Blue Group’ and is located at the place ‘Central Square’ and ...

This richer interpretation could allow further automatic inferences to be drawn by machine agents, such as whether or not this protest is already known to the knowledge base, whether the Blue Group is known to be active in the area, and so on.

Figure 4 shows a screenshot of a social media analytics tool called Sentinel⁴ into which a CE-based knowledge base and ontology has been integrated for the purposes of field-testing the approach. Sentinel performs event detection using the FlexiTerm method²⁴ to identify recurring ‘trending’ phrases in 15-minute periods (shown in the bottom-centre panel) and links the detected phrases to the relevant tweets (shown on the right). Where the tweet refers to terms in Sentinel’s CE ontology these are shown as ‘tags’ under each tweet. Tweets can also be navigated by selecting a tag (for example, **protest**).

Integrated into Sentinel is a conversational software agent called Moira (Mobile Information Reporting App)²⁵ that supports NL dialogues on the contents of Sentinel’s CE knowledge base (the contents of which is partially foraged from social media) as illustrated in Figure 5. When active, the Moira app occupies the right hand panel of Sentinel shown containing tweets in Figure 4. Figures 4 and 5 show the desktop configurations of the Moira and Sentinel apps; both can also be run in a mobile device configuration for users in the field.

4. FIELD EXPERIMENT & VIGNETTES

Application of the approach is illustrated through a series of vignettes based around a field experiment, the configuration of which is shown in Figure 6. The purpose of the experiment was to monitor a large-scale event through a combination of real-time social media analysis and fieldwork, while also testing the information architecture and tools described in the previous section. Variants of this set-up have also been used in a variety of studies in relation to policing and counter-terrorism.^{3, 4, 8, 27} After describing the set up, the remainder of this section will examine vignettes showing how the information architecture assists in information foraging and framing to inform decision making, drawing upon open source intelligence.

To demonstrate the capability to support real-time analysis of a situation involving analysts at multiple locations, two sites were set up for the experiment: a primary site and a secondary site. CE knowledge bases

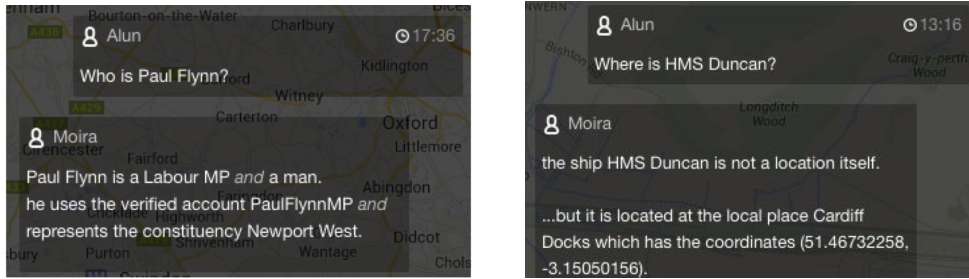


Figure 5. An example of question answering from forged data using the Moira app in Sentinel²⁶

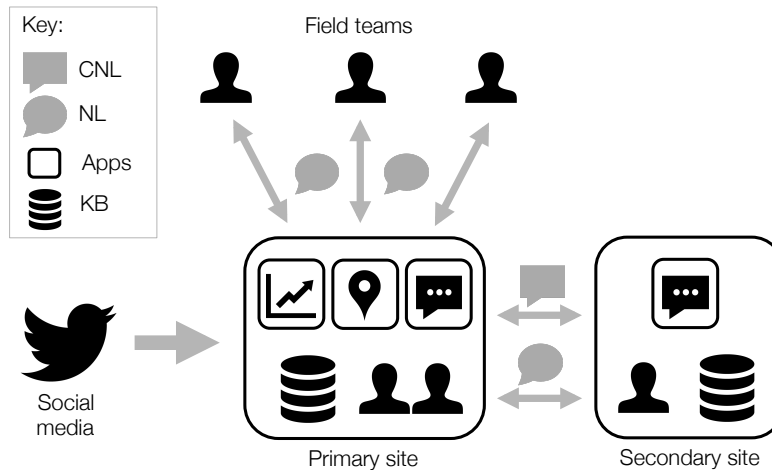


Figure 6. Field experiment set-up

were maintained at both sites, though these were not identical, with each incorporating local as well as shared information. Different analytical applications were also used at each site. For example, Sentinel was used only in the primary site (and by field teams) though versions of the Moira agent were used in both sites as well as by the field teams. Commercial and freely-available 3rd party applications such as Tweetdeck were also used at both sites. Information and knowledge was communicated between both sites in NL and CNL. Audio and video communication via Skype was also used.

Due to its properties as an effective means of disseminating real-time and situational information, as well as carrying links to other media (see Section 1), social media data collection focussed on Twitter. Social media data collection and processing was based at the primary site, with Sentinel being the primary analytics tool used. Information acquired via Twitter and added to the knowledge base was then shared automatically with the secondary site.

A key feature of the experiment set-up was the ability to task field teams to visit specific locations and gather on-the-spot reports, to supplement or verify data collected from social media. Field team members were equipped with mobile devices and communicated with the primary site, chiefly via text messages and, where appropriate, attached imagery. Generally, to allow field team members to focus on observation tasks and eliminate any overheads in terms of technology usage, they were allowed to use any tools with which they were comfortable; however, during the experiment the entire team was encouraged to use the Slack collaboration platform** as a common communication environment. Slack supports real-time messaging, organised into a

**<http://slack.com>

collection of thematic ‘channels’, including posting of documents and other media, including items foraged from social media, and provides a persistent timeline-based record of discussions after the event.

4.1 Vignette 1: Rapid Knowledge-Based Framing of Foraged Data

A key aspect of the experiment focussed on detection and characterisation of public protest activity in connection with a large-scale event. In particular, the goal was to explore how relatively lightweight and reusable CE knowledge models could rapidly be developed to add context to collected social media data. As described in the previous section, the Sentinel tool was able to detect sets of tweets referring to protest activity by first identifying a particular common linguistic expression via the FlexiTerm method, and then identifying references to ontology terms such as `protest`, `demonstration` and `march` in those tweets (this can be seen on the right of Figure 4).

In this way, by automatically detecting patterns in the Twitter stream corresponding to events, and associating these with ontological terms denoting types of events of interest, Sentinel is able to rapidly enact the foraging loop by assembling relevant social media fragments into a shoebox and, by linking them to the machine-processable ontology, semantically organise the fragments as evidential support for particular events occurring. Placing the information in the context of the ontology of events is the lowest level of ‘schematizing’ the data, i.e. the start of the sensemaking part of the loop (step 8 in Figure 1).

However, once the data has been contextualised in the machine-processable part of the information space, further inferential support becomes possible. For example, by cross-referencing newly-detected `protest` events with information on the time and location of planned protests (which is typically made available in advance via open sources to minimise disruption to the public) it became possible to automatically characterise these events via CE rules as ‘expected’ and ‘unexpected’, with the latter being the more interesting category.

Further interpretation of the detected events is possible by inference from the involved parties. For example, where a known group or prominent individual is involved in an event (as illustrated by the fictitious example of the ‘Blue Group’ in the previous section) generally the group will have a known stance on events (for example, being against some other organisation or political position) and detection of the involvement of that group or individual’s stance also characterises what the protest may be about, by a chain of inference. This provides further semantic enrichment of the raw situational data, pushing out from the foraging loop into the lower stages of the sensemaking loop.

4.2 Vignette 2: Conversational Information Fusion

As introduced at the start of this section, a key feature of the experiment was to combine data foraged from social media with reports from trusted observers. The typical mode of operation here was to present detected and framed events to team members in the two sites, identifying where there was a need for additional information or corroboration. The CE model serves to determine what is important about concepts, establishing a requirement to gather further information. For example, there may be missing or conflicting information on key features such as the `size` of a protest, or the `actors involved`.

These needs for further information are represented by the feedback steps in the sensemaking loop. For example, incompleteness in the shoebox leads to searches for further information — step 3 in Figure 1 — while a lack of information on the stance of a protest may be revealed at a higher level in the process, such as steps 10 (‘schematising’) or 13 (‘hypothesising’). The key point is that the missing, conflicting, or uncorroborated information can be detected with machine support once the data has been framed in the context of the knowledge base.

4.3 Vignette 3: Tasking and Deciding at the Edge

The requirements to collect further information described in the previous vignette may be served by additional searches on social media. However, in many cases this will not be feasible (because the information has not been posted or is not findable) or appropriate (because a trusted source is not available to complete the information picture, resolve a conflict, or corroborate an important finding). In such cases, the approach supports the rapid tasking of information collection via field teams: field teams can be tasked to provide input on such missing or conflicting features using the Moira app or their preferred communication software tools.

A common example of this involved cases where social media data indicated a potential future event, such as a call to gather protesters at a particular time and place. By tracking field teams to observe the location at that time, reliable information on the real-world response to the social media call could be obtained, providing valuable information on the strength of a particular group or stance. In some cases, it was observed that no protesters responded to the call — something that was impossible to detect by social media monitoring alone.

Deploying tools such as Sentinel and Moira as mobile apps closes the loop completely, in that it allows decision-makers who need to understand the unfolding situation to operate in the field, able to access the foraged and semantically-framed events via Sentinel, and conduct searches for further information via Moira, as well as feeding their own observations and hypotheses into the knowledge base from a position of direct observation.

5. DISCUSSION & CONCLUSION

The experimental deployment of the approach so far indicates a good degree of success in integrating machine and human-processing of data, including social media and field observations, spanning the sensemaking process including elements of the foraging and sensemaking loops. Use of CNL-based models help to create a ‘glass box’ approach where the results of a machine-driven analysis are understandable and interpretable by human analysts and field teams tasked to collect further information. The approach thus supports both directions of the process as characterised in Figure 1, the forward construction of progressively higher-order information products, and the backward search and refinement steps.

The experimental work highlighted several challenges, however. One key finding was that, in general, the field teams preferred to work with commercial and freely-available tools rather than custom tools such as Sentinel and Moira. Part of this was due to their relative familiarity with those existing tools, and part of it was due to the relative maturity and robustness of the tools (Sentinel and Moira being research-grade rather than production software). In some cases, field teams even relied on traditional pen-and-paper note-taking where it was undesirable to use digital tools (for example, when subjects were uncomfortable with the team members using digital recording devices). Going forward, more work is required to consider the best ways to integrate the CNL-based approaches into users’ preferred ‘ecosystem’ of tools and methods.

A second key finding was that imagery plays a vital role in situational understanding, alongside text. One highly-useful feature of the Sentinel and Moira tools were their ability to generate collages of images collected from social media, in the context of particular searches or detected events, allowing a user to very rapidly scan the images and make sense of their content. Automated image processing would seem to offer additional support in this activity, and will be explored in future work.

Acknowledgements

This research was sponsored by the US Army Research Laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the US Government, the UK Ministry of Defence or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Development of the Sentinel platform was funded by the European Commission under the project ‘Tackling Radicalisation in Dispersed Societies (TaRDIS)’, the ESRC via the project ‘After Woolwich: Social Reactions on Social Media’ (ES/L008181/1), and the UK College of Policing via the Open Source Communications Analytics Research (OSCAR) Development Centre (<http://upsi.org.uk/oscar/>). Cardiff University and IBM UK provided funding for the pilot study examining community impacts of the NATO Summit.

REFERENCES

1. A. Olcott, *Open Source Intelligence in a Networked World*, Bloomsbury, 2012.
2. F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, “Semantics + filtering + search = Twitcident. exploring information in social web streams,” in *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, pp. 285–294, 2012.

3. M. Innes, C. Roberts, and D. Rogers, "Critical timing," *Police Professional* **January**, pp. 17–18, 2014.
4. C. Roberts, M. Innes, A. Preece, and I. Spasić, "Soft facts and spontaneous community mobilisation: the role of rumour after major crime events," in *Data for Good: How big and open data can be used for the common good*, P. Baeck, ed., pp. 37–43, Nesta, 2015.
5. D. Wang, T. Abdelzaher, and L. Kaplan, *Social Sensing: Building Reliable Systems on Unreliable Data*, Morgan Kaufmann, 2015.
6. D. Gayo-Avello, "A meta-analysis of state-of-the-art electoral prediction from twitter data," *Social Science Computer Review* **31**(6), pp. 649–679, 2013.
7. D. Lazer, R. K. G. King, and A. Vespignani, "The parable of Google Flu: Traps in big data analysis," *Science* **343**, pp. 1203–1205, 2014.
8. M. Innes, C. Roberts, A. Preece, and D. Rogers, "Of instruments and data: Social media uses, abuses and analysis," in *Handbook of Online Research Methods*, N. Fielding, ed., Sage, 2 ed., 2016.
9. F. Pasquale, *The Black Box Society*, Harvard University Press, 2015.
10. B. Broome, "Data-to-decisions: a transdisciplinary approach to decision support efforts at ARL," in *Proc Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR III (SPIE Vol 8389)*, SPIE, 2012.
11. D. S. Alberts and R. E. Hayes, *Power to the Edge: Command and Control in the Information Age*, CCRP, 2003.
12. P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," in *Proceedings of International Conference on Intelligence Analysis*, 2005.
13. L. M. Aiello, G. Petkos, C. Martin, D. Corney, S. Papadopoulos, R. Skraba, A. Gker, I. Kompatsiaris, and A. Jaimes, "Sensing trending topics in twitter," *IEEE Transactions on Multimedia* **15**(6), pp. 1268–1282, 2013.
14. D. Wang, M. T. Amin, S. Li, T. Abdelzaher, L. Kaplan, S. Gu, C. Pan, H. Liu, C. C. Aggarwal, R. Ganti, X. Wang, P. Mohapatra, B. Szymanski, and H. Le, "Using humans as sensors: An estimation-theoretic perspective," in *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, IPSN '14*, pp. 35–46, 2014.
15. L. Terveen, "Overview of human-computer collaboration," *Knowledge-Based Systems* **8**(2), pp. 67–81, 1995.
16. A. Wollocko, M. Farry, and R. Stark, "Supporting tactical intelligence using collaborative environments and social networking," in *Proc Next-Generation Analyst (SPIE Vol 8758)*, SPIE, 2013.
17. R. Brantingham and A. Hossain, "Crowded: a crowd-sourced perspective of events as they happen," in *Proc Next-Generation Analyst (SPIE Vol 8758)*, SPIE, 2013.
18. F. Jin, W. Wang, L. Zhao, E. Dougherty, Y. Cao, C.-T. Lu, and N. Ramakrishnan, "Misinformation propagation in the age of twitter," *IEEE Computer* **47**(12), pp. 90–94, 2014.
19. D. L. Hall and J. M. Jordan, *Human-Centered Information Fusion*, Artech House, 2010.
20. J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, and F. White, "Revisiting the JDL data fusion model II," in *Proc Seventh International Conference on Information Fusion (FUSION 2004)*, pp. 1218–1230, 2004.
21. T. Kuhn, "A survey and classification of controlled natural languages," *Computational Linguistics* **40**, pp. 121–170, 2014.
22. D. Mott, "Summary of ITA Controlled English," 2010.
23. W3C, "OWL 2 Web Ontology Language document overview (second edition)." World Wide Web Consortium, Dec. 2012.
24. I. Spasić, M. Greenwood, A. Preece, N. Francis, and G. Elwyn, "FlexiTerm: a flexible term recognition method," *Biomedical Semantics* **4**(27), 2013.
25. A. Preece, C. Gwilliams, C. Parizas, D. Pizzocaro, J. Z. Bakdash, and D. Braines, "Conversational sensing," in *Proc Next-Generation Analyst II (SPIE Vol 9122)*, SPIE, 2014.
26. A. Preece, W. Webberley, and D. Braines, "Conversational sensemaking," in *Proc Next-Generation Analyst III (SPIE Vol 9499)*, SPIE, 2015.
27. A. Preece, W. Webberley, and D. Braines, "Tasking the tweeters: Obtaining actionable information from human sensors," in *Proc Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VI (SPIE Vol 9464)*, SPIE, 2015.