Institute of Development Studies

Brief supporting Evidence Report 122

TRACKING RESEARCH AND POLICY CONVERSATIONS IN ONLINE SPACES

Policy Anticipation, Response and Evaluation

Alistair Scott and Tamlyn Munslow

March 2015

The IDS programme on Strengthening Evidence-based Policy works across seven key themes. Each theme works with partner institutions to co-construct policy-relevant knowledge and engage in policy-influencing processes. This material has been developed under the Policy Anticipation, Response and Evaluation theme.

The material has been funded by UK aid from the UK Government, however the views expressed do not necessarily reflect the UK Government's official policies.

AG Level 2 Output ID: 568

TRACKING RESEARCH AND POLICY CONVERSATIONS IN ONLINE SPACES

Alistair Scott and Tamlyn Munslow

March 2015

First published by the Institute of Development Studies in March 2015

IDS is a charitable company limited by guarantee and registered in England (No. 877338).

Why is this important?

Whether in print or online, research is published in the hope that people will read it and act on it. Ideally research would be used as evidence to inform projects and debates shaping decisions and actions that affect people's lives. To make this more likely, researchers, and the institutions in which they are based, often make great efforts to disseminate published findings among particular audiences. Increasingly, this involves distilling key messages from research findings and publishing these messages in the form of more concise, additional documents, whose purpose is to summarise, describe or merely signpost a key research and evidence report. These supporting documents are usually published online and they come in a variety of forms, including abstracts, summaries, briefings, blogs, email newsletters and increasingly as social media posts on Twitter, Facebook and other platforms. These documents may trigger online debates that help to further publicise the research findings. These debates also help to build on, refine, contextualise and reframe the research findings in ways that the original authors may not have predicted. In this way publication of a main research report, followed by supporting documents and posts, is only the beginning of an electronic dialogue in which the internet (and especially its social media component) is employed as a debating platform through which interested audiences voice their own reactions to the evidence. When successful, this electronic dialogue becomes an important step in the policy engagement process. The ensuing dialogue draws in more and more people, including, ultimately, those who work in the spaces where policy decisions are made or implemented.

Variously described as communications, marketing or uptake, this awareness-building can be challenging to achieve and even more challenging to monitor. So how can we determine the success of our efforts to communicate research findings? Can we systematically assess what influence, if any, research-based evidence has had in the world?

One feature of online dialogue might help to answer these questions: unlike off-line conversations, public online debates leave a digital trace. Our project therefore constitutes a first step towards exploring how these debates in online and social media might be tracked, mapped and analysed, and how this might benefit the process of communicating research and evidence.

Of course web statistics, such as numbers of downloads or page views, are already often used as proxy measures to indicate the extent to which users have engaged with published research. However, these indicators are static and focus on individual users or entities to monitor change. At best they can measure how many times a publication is read, and by whom. What we have done in our project goes further than this and involves tracking and mapping key messages contained within publications. This creates the possibility of visualising much more dynamic and revealing information about the way that research findings are taken up. If it is possible to devise a tool that monitors messages – rather than single users or entities – then our potential for learning may be vastly increased. Our goal specifically was to assess the feasibility of such a tool and to build and test a prototype.

Our approach

To test our approach we decided to use the Hunger and Nutrition Commitment Index (HANCI). This is an index developed at IDS, which ranks governments in terms of their political commitment to tackling hunger and undernutrition. A key feature of HANCI is to frame hunger and undernutrition as issues of political commitment, rather than merely problems of technical capacity. HANCI aims to change the way in which these issues are portrayed and discussed in the media and in policy circles, and it encourages debate about how the problems of hunger and nutrition should be tackled. We therefore chose HANCI because of its use of research-based messages in this particular way. The publication date

for the 2013 HANCI report was 18 November 2014. This meant that our searches could cover the weeks leading up to the launch of the report, during which period a considerable number of communications activities would take place.

To map how messages spread, one first needs to be able to find them. There are many different types of application for searching and tracking online documents and text, including search engines, citation metrics, alternative metrics, media monitoring applications and social media tracking tools. We reviewed these different approaches and came to the conclusion that without modification, no single application was capable of generating exactly the right kind of search data for our purposes. We therefore decided to experiment with Vocus and Meltwater, two media monitoring applications, together with Zapier, an application for extracting data from social media platforms. The reason for this combined approach was to use media monitoring to enable broad searches across a wide range of content, while using Zapier to facilitate more in-depth searches on one specific social media platform: Twitter. The latter was a way to exploit Twitter's tendency to generate relatively simple and concise content and usage data.

Vocus and Meltwater were used to search a wide range of online sources, including websites (particularly news and media sites), blogs, public e-discussion fora and social media (including Twitter, Facebook, LinkedIn, YouTube, Instagram, Mendeley and others). Zapier on the other hand was used to generate search results containing a range of Twitter-specific data, including content, the author of each tweet, the author's profile (a short self-description provided by each Twitter user) and the number of Twitter followers for each author. Crucially these data also contain variables indicating whether each recorded tweet was original or a retweet of someone else's tweet, or even a so-called modified tweet (that is a retweet in which the content has been edited). In all cases the date of each search result was also recorded, so that all the searches could be plotted on a timeline, indicating the chronological sequence in which each item in the search results originally appeared online. We also decided to focus our searches on the period between September and November 2014.

We concentrated on developing a bespoke approach to searching online and social media spaces and to visualising the resulting data in formats that are conducive to further analysis. Our plan consisted of two main steps:

1. **Searches:** using search algorithms to identify online and social media 'objects' (blogs, Facebook posts, tweets, news articles and so on) that refer to particular concepts expressed within published outputs
2. **Visualisations:** developing novel approaches to displaying and mapping the search results over time in order to contribute to a deeper understanding of the behaviour of different actors within the research uptake process.

To run the searches in Vocus and Meltwater we used a carefully designed and tested Boolean search string.[1] In the case of Zapier, however, we were searching data specific to Twitter and we therefore achieved the best results by searching on two Twitter hashtags: #HANCIndex and #nutritionlive.

Vocus, Meltwater and Zapier are all able to export search results as spreadsheet files, but for the visualisations we needed some kind of rapid prototyping tool, capable of manipulating and displaying the exported data in a variety of graphical forms. Vocus and Meltwater produce their own visualisations, which are impressive and useful, particularly for visualising brand awareness or market penetration of products. However, neither application produces precisely the kind of chronological visual presentations we were hoping to develop. Zapier,

[1] A Boolean search is a technique for improving the relevance of search results by using a combination of keywords with operators such as AND, NOT and OR.

moreover, has no functionality at all for generating graphical displays. We therefore decided to use Tableau, a leading commercial data visualisation application, to design bespoke visualisations for this study. The resulting visualisations were basically Cartesian graphs, where the horizontal axis showed time and the vertical axis showed the author (individual or organisation) responsible for each search result. The results themselves were then displayed as points on the graph. In addition, clicking on each data point opened a pop-up window displaying a variety of data specific to that particular result and its author.

Visualisations

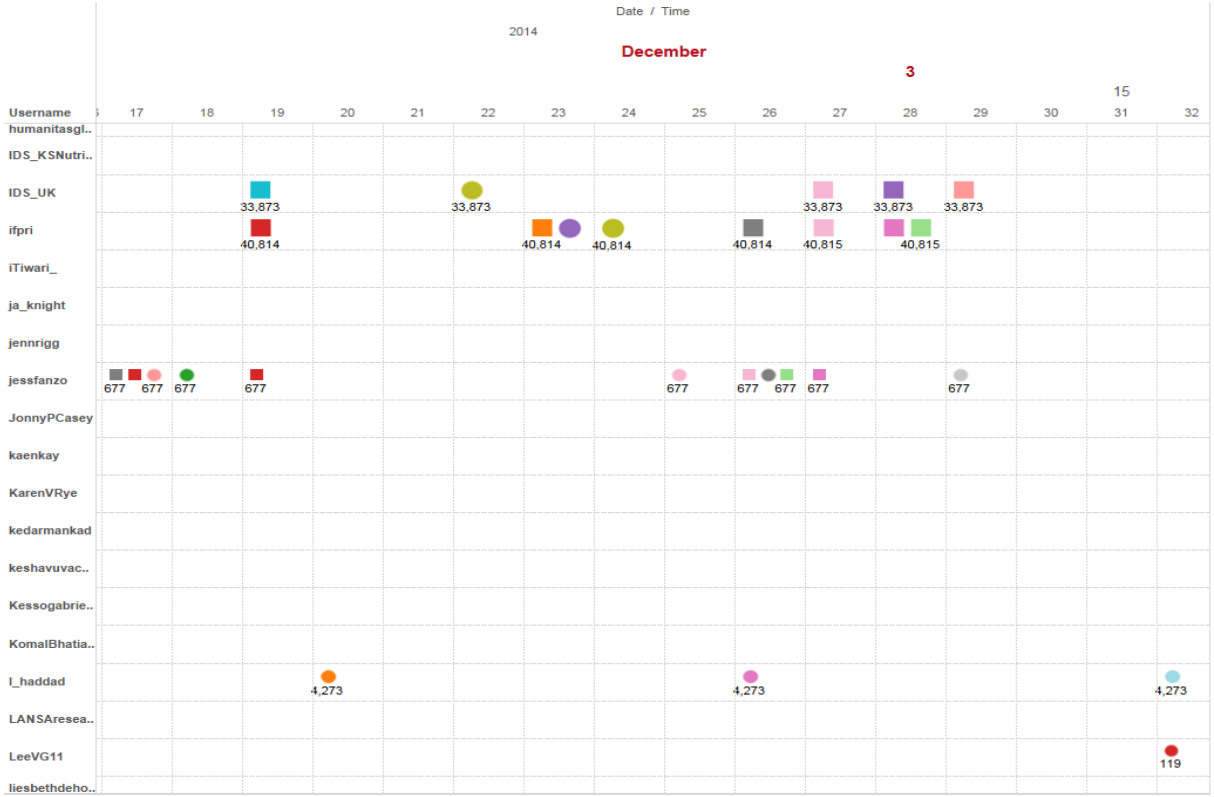
The Vocus and Meltwater searches produced a rich array of data sources, as was to be expected, given the many types of online and social media which the applications can search. However, they were not particularly useful for tracking messages. Both Vocus and Meltwater make a distinction in the way they handle online news media results and social media results. Consequently it can be difficult to create useful visualisations containing both types of data on the same graph. There was also a general lack of useful qualitative information about each data source. This meant for example that in most cases it was not possible to visualise much information about the professional interests of an individual author. A lack of useful quantitative data was also apparent in many cases. The absence of data about website traffic or social media followers, for example, meant that it was not always possible to distinguish the relative popularity – and hence potential influence – of different data sources. Indeed these visualisations were not markedly different from some of the standard visualisations which Vocus and Meltwater are able to generate by themselves.

The Zapier visualisations were much more focused because they used only Twitter data. Because these data included tweets and retweets it was possible to trace the path taken by particular messages over time, as shown in Figures 0.1 and 0.2. In Figure 0.1, tweets are represented by circular data points while retweets are shown by squares; each thread of tweets and retweets is represented by a different colour. Selecting a single colour allows the user to visualise a single thread of linked messages as illustrated in Figure 0.2. The visualisations constructed in this way came much closer to the aims of our study because we were able to display chronological pathways of data where individual messages were picked up and re-sent by subsequent authors.

The Zapier visualisation also used the number of Twitter followers as a proxy indicator for the popularity (and thus in crude terms, the potential influence) of each author. A label showing the number of followers was displayed next to each data point. This number was also used to vary the size of each data point. Small data points indicated relatively few followers, while large data points indicated many, so this provided an immediate visual indication of the relative influence of each contributor.

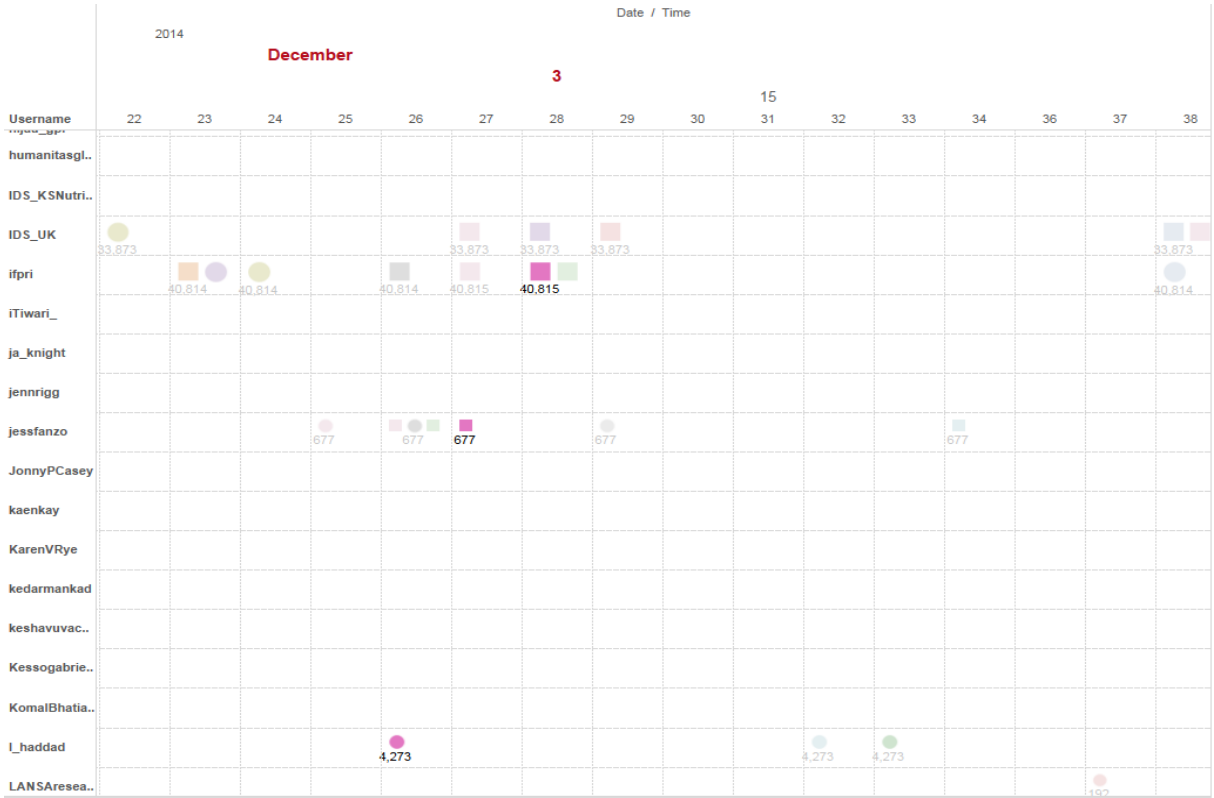
The user profile for each author is also known and it can be displayed in the pop-up window for each data point, together with other useful data as shown in Figure 0.3. This was a particularly useful feature for our search on #nutritionlive, because this was a visualisation of a Global Nutrition Report e-discussion, which took place on 3 December 2014. Being able to quickly see the interests, institutional affiliations and job titles of the contributors provided valuable information about the reach and effectiveness of the e-discussion event.

Figure 0.1 Partial view of Zapier #nutritionlive visualisation field



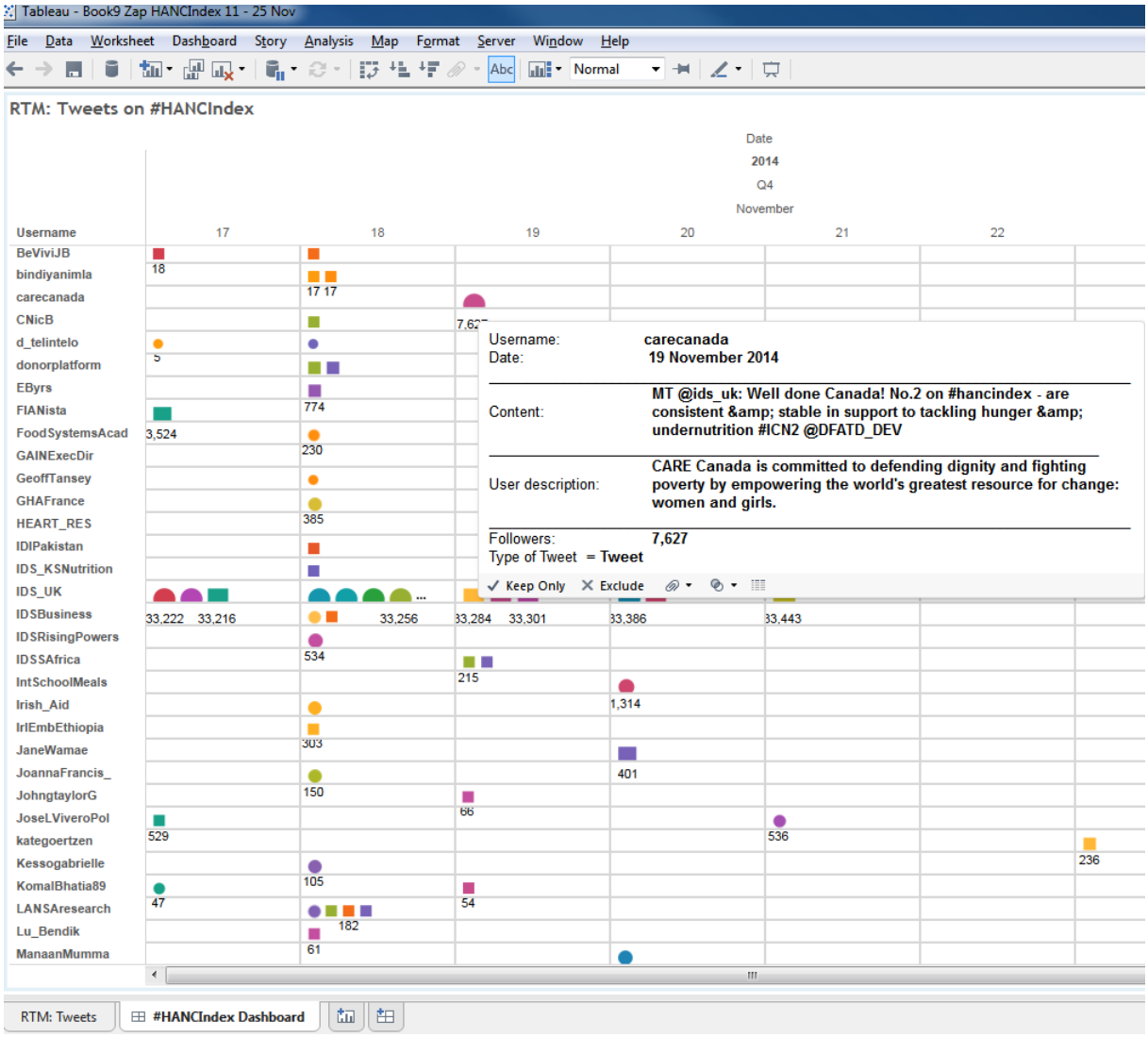
Source: Authors' own.

Figure 0.2 Zapier visualisation of a single Twitter thread (pink)



Source: Authors' own.

Figure 0.3 Partial view of Zapier #HANCIndex visualisation field showing pop-up window



Source: Authors' own.

Key insights and suggestions for further development

Though rudimentary, these visualisations demonstrated the feasibility and utility of developing systems for monitoring and tracking research messages in online spaces. They also suggested that it may be more productive, initially at least, to focus these efforts on specific social media platforms, such as Twitter, rather than attempting more eclectic approaches.

The prototype shown here could easily be developed into a fully functioning application and a number of simple improvements could be incorporated along the way. For example, the ability to visualise individual message pathways could be enhanced by programming the software to draw a connecting line between all data points sharing the same message thread. It would also be relatively easy to incorporate Google Analytics and webstats data about specific publications into the same visualisation. This would permit one to observe the extent to which downloads of key research publications correlate with particular bursts of related social media activity.

More powerful functionality could be developed, too. For example, the fact that Twitter uses short, simple messages, coupled with a number of structural conventions such as hashtags, retweets and modified tweets opens up the possibility of tracking how messages are changing over time. Twitter records data that link an original tweet to any subsequent modified reincarnations, so in principle at least it should be possible to visualise evolving threads of ideas.

Twitter is a relatively crude communication tool. A limit of 140 characters per tweet does not allow much space for the development of an argument. But Twitter users frequently embed links to longer documents in their tweets. Identifying Twitter threads and then examining the documents they link to could be one way to enrich the exploration of these social media pathways. Of course Twitter activity does not take place in isolation. Having plotted an ‘ideas pathway’ across a Twitter space, our visualisations show how easy it is to display related activities in other online spaces during the same time period.

The 160-character limit of the Twitter user profile, and the consequent tendency for Twitter users to use hashtags and keywords in their profiles means that it can be searched and analysed using relatively simple pattern-matching search techniques. Using these data would mean that message threads could also easily incorporate data about who the Twitter users are, their countries of origin, their interests, professions and institutional homes. Thus it would be possible to generate visualisations of groups of authors who share similar specialisms, interests or professional roles, and overlay these on the message pathways.

Having developed such a monitoring system for Twitter, possibly the easiest of the social media platforms to apply this approach to, it should be possible to develop similar applications for other social media platforms, such as Facebook. The greater the number of platforms being monitored, the richer will be the resulting visualisations.

To sum up, our study suggests that it is both useful and practical to develop a software application to monitor social media and public online spaces in order to assist in the analysis of the influence of research evidence on real-world decisions. Such an application has the potential to help analyse how ideas are framed and reframed through public debate.