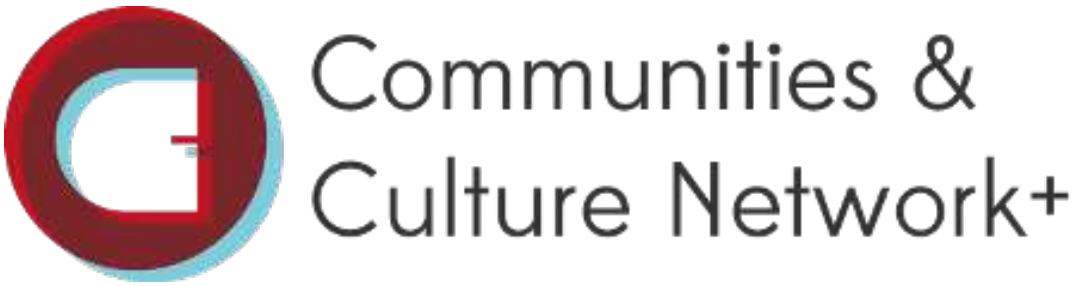
Digital Data Analysis, Public Engagement and the Social Life of Methods: Final Report

Helen Kennedy, Giles Moss, Chris Birchall, Stylianos Moshonas

Project grant number: EPSRCCCN2013-P1



Contents

1. Executive Summary

Digital Data Analysis, Public Engagement and the Social Life of Methods (DDA) explored how public sector organisations might use digital data analysis applications to know and engage their publics. The project experimented with a range of tools in collaboration with three public sector partner organisations (two councils and one museums group). The tools we used included: Social Mention, Topsy, TweetReach, Peer Index, Klout, Kred, DataSift, IssueCrawler, NodeXL, Gephi, Meltwater Buzz and Brandwatch.

DDA concluded that the use of such tools is much more complex than celebratory rhetoric about them suggests, because: they require expertise; it is not clear how they work; certain decisions have been made which affect how and where they look for data and this affects the data that results; and data itself is not always abundant.

There is a range of possible public sector uses of such tools, such as: measuring public engagement; identifying key 'influencers' and networks with which to engage; managing and analysing enquiries and feedback from the public about services; monitoring relevant publicity; identifying and analysing public conversation and opinion on local issues; identifying and

engaging the public in policy development. However, there are also a number of challenges and barriers to their use. These include: the fact that data about local publics is hard to access; the limited expertise of public sector employees in the use of such tools; lack of human resources; a lack of understanding and support in relation to these methods within organisations; the limited and black-boxed nature of the tools; the lack of relevant local data and the need to create public discussions if they do not already exist.

Because the consequences of using digital data analytics for public sector decision-making are serious (some groups or communities may be excluded from service provision, for example), we conclude that insights from digital data analysis should complement, not replace, established forms of public engagement (such as consultations, citizen panels, and petitions).

Building on the empirical research, the project concluded with speculation about ways in which analytics might be developed in order to serve the public good. We propose that analytics could become more public itself, in three main ways. First, analytics needs to be available to the public to use. Second, analytics should be open to public supervision. Third, analytics could be rethought as a more reflexive and participatory process, not just as a way for experts to track and know the public with ever-greater precision, but rather as offering the means for publics to come to know and represent themselves. However, this kind of democratisation of analytics raises political and normative questions about whether analytics an appropriate tool with which to overcome emergent digital divides.

2. Aims and Objectives

People's web and social media use generates a vast new source of data which, it is suggested, can be analysed for new insights into what people think and feel, how they behave, and about the nature of social networks and relationships. *Digital Data Analysis, Public Engagement and the Social Life of Methods* (DDA) aimed to interrogate such bold claims about what the analysis of digital, social media or 'big' data might tell us, by investigating the use of digital data analysis in practice. Working with three public sector organisations (two city councils and one museums group), our research experimented with different forms of digital data analysis in order to examine how they might help public organisations to know their publics better. DDA asked these questions:

- How can those without the economic means to pay for digital data and who want to use it for the public good access it?
- What is the potential application of digital data analysis (e.g., social network analysis, machine learning, natural language processing) for understanding and engaging publics?
- What are the broader (epistemological, normative) implications of the ways publics are being constituted and made visible through digital data methods?

3. Key Findings

3.1. How can those without the economic means to pay for digital data and who want to use it for the public good access it?

To address this question, we experimented with a range of digital data analysis tools. Prior to DDA, we had undertaken a scoping study exploring whether digital data methods were used by public sector organisations, including those we worked with on DDA, and we found that some

tools were used, but not systematically, and that organisations were keen to do more, but resources were limited. Hence our initial focus was on some fairly advanced, freely available tools:

- DataSift (for data gathering)
- IssueCrawler (for issue mapping)
- NodeXL (social network analysis)
- Gephi (data visualisation).

However, partly because these tools were complex to use, partly because during the course of the research we discovered that less digital data analysis was undertaken than we initially assumed, and partly because one of our partners requested it, we introduced two commercial tools to the study. These were:

- Meltwater Buzz and Brandwatch (commercial tools which perform a range of social media analyses)

In a workshop at the end of the project, we also introduced partners to the following freely available, easy-to-use online tools: Social Mention, Topsy, TweetReach, Peer Index, Klout, Kred.

We found that the assumption in the claims mentioned above that it is possible to pick up these tools, run with them, and easily find out about people's feelings, opinions and actions is problematic. Our research suggested that digital data analytics in practice does not match the rhetoric, for four reasons: 1) digital data analytics is 'socially shaped'; 2) human expertise is needed to use the tools; 3) data is sometimes not abundant; 4) tools are black-boxed. These points are discussed in more detail below. We conclude that enabling resource-limited groups to do digital data analysis is more complex than simply providing them with access to tools, whether these are free (like NodeXL, Gephi, Social Mention and others), cheap (like DataSift) or more costly (like Meltwater Buzz and Brandwatch).

3.1.1. Digital data analytics is 'socially shaped'

Digital data analysis is shaped in three key ways. First, it is shaped when analytics software is designed and developed. Different tools are designed to source and analyse data in different ways, and these choices shape the resulting data. DataSift, for example, does not search the kinds of local forums where our partners might find relevant data, such as city-based forums or regional newspapers' comments sections. Second, the choice of tools used to carry out analytics work shapes what analysis can be carried out and the data that results. A number of considerations influenced our choice of tools on this project (cost, the kinds of analysis they do), and we therefore sought tools which would allow them to advance their existing practice. In this respect, the interpretations of the research team shaped the resulting data.

The third moment of social shaping is when the tools are used. Software, methods and data require interpretive work in their implementation, as they are used by actors for different purposes in varying organisational contexts. During our research, the account director from one of the commercial companies who liaised with us during our trial directly configured the company's tool to search a particular local forum which we identified as relevant for content related to our searches. This resulted in a more complete result set than would otherwise have been possible. This helpful customer service increased our reach to one of our partner's target user groups, but it also highlights that some groups are excluded from results sets because the platforms they use to express their views are sometimes not searched by analytics tools.

3.1.2. Human expertise is needed to use the tools

Various types of expertise are needed to use digital data analysis tools. For example, expertise is required to generate or 'collect' data. Expertise in keyword search is crucial in this process, as we discovered when the keywords our partners provided in relation to search topics generated few results. DataSift, NodeXL and Gephi require particular technical knowledge (of the fields of data held in records returned by the APIs of platforms like Twitter and Facebook, and of social network analysis, to give just two examples).

The commercial tools that we trialled were also experienced as difficult to use by our participants, perhaps somewhat surprisingly given efforts made to produce usable graphical interfaces. However, they still require certain kinds of expertise, for example in writing Boolean searches. As a result, a number of participants preferred the freely available, easy-to-use online tools that we presented to them during our end-of-project workshop (Social Mention, Topsy, TweetReach, Peer Index, Klout, Kred). These provide very superficial analysis but, as one participant put it, 'they're free, they're very simple to use and it gives you something rather than nothing'.

3.1.3. Data is sometimes not abundant

Despite widespread claims about the variety, velocity and volume of data on social media platforms, relevant data was not easy to find. When using the tools listed above, platforms that we had identified (through manual searches) as used for conversation about our chosen topics did not feature in results, a consequence of which was limited relevant data. Data shortage was sometimes because of lack of expertise in keyword selection, discussed above, but even when efforts were made to improve keywords (such as adding colloquialisms and common slang, or using the Flickr search API to find better terms), results improved only slightly and several topics still returned few or no results.

Being local organisations, our partners were interested in finding local conversations and local 'influencers' with whom to engage. However, very little social data actually contains accurate location information. Seeking geographical data can diminish an already small pool of data and exclude relevant contributors (such as local people writing comments on newspaper websites, forums, blogs and in most cases Facebook and Twitter too) who were not sharing their geographical location in any way.

3.1.4. Digital data analysis tools are black-boxed

It is often unclear how digital data analysis tools work. DataSift tells the user which platforms are to be searched and why, but the user has to look carefully through DataSift's documentation to find this information. Other tools do not provide this information. Topsy provides a brief white paper about sentiment analysis but not about its other metrics, and Social Mention states on its site that its functionality is not for public viewing – such information can be commercially sensitive. NodeXL, Gephi and IssueCrawler are transparent in that they are designed on the basis of academic research that is publicly available (IssueCrawler shares papers about its co-location analysis model, for example), but the software is designed by academics for academics, rather than the public, and as such is hard for non-experts to comprehend.

There are other ways in which the functionality of analytics tools is not transparent. For example, Meltwater Buzz and Brandwatch treat online newspaper stories and their associated comments fields differently. One of the tools counts each single comment on such a story as a data item, but not the newspaper article itself, whereas the other does the opposite, counting the whole page, article and comments, as one item. This affects how they count data, but none of this is transparent to the tool user. They use varied mechanisms for generating data sets, utilising

different aggregator services, yet none of these variations is evident at first glance. This highlights not only the black-boxing of analytics tools, but also the distinct ways in which data is generated and the distinct data that is generated.

Visualisations of data can also be black-boxed, or created using undocumented methods. An analytics tool could ask Twitter for a lot of data and then present findings in a visualisation as if a high proportion of results came 'naturally' from Twitter without explaining the process by which the data was acquired. Visualisations are often created by add-on libraries or tools, created externally and purchased by the tool developer, yet how such components work is not always known by the tools purchasing and integrating them. Tools access different data sources, have different access deals with platforms and present different end results in different ways. Most of these differences are hidden from view, or black-boxed, yet their impact on the ways in which data is generated, presented, consumed and used is significant.

For these reasons, we conclude that whilst free, cheap or apparently 'easy' tools are available, and they tell us something about our audiences, publics and communities, the factors outlined above point to the complexities and limitations of adopting such methods.

3.2. What is the potential application of digital data analysis for understanding and engaging publics?

3.2.1. There is a range of possible uses

DDA found that there is a range of ways digital data analysis could be used by public sector organisations to understand and engage their local publics. This includes:

- *Measuring public engagement.* Digital data analysis can be used as a way to track and measure the effectiveness of campaigns, initiatives, and services and in particular the public's engagement with them. The data that is generated can then be cited as evidence of the organisation's public impact and reach to senior managers (and, where relevant, to external bodies) and can be reviewed in order to make improvements to the design of campaigns, initiatives, and services in future.
- *Identifying key 'influencers' and networks with which to engage.* Digital data analysis can be used to identify key networks and 'influencers' (such as influential local tweeters or bloggers) of which the organisation was not previously aware. The enriched understanding of social media presence and local networks provided by digital data analysis can help organisations to spread their messages more effectively and widely, including to groups that organisations may not be able to reach via conventional channels, and to support better public engagement.
- *Managing and analysing enquiries and feedback from the public about services.* Digital data analysis can be used to help organisations manage enquiries and feedback about their services in a more effective and timely manner, while the aggregation of such data can help to detect trends, both positive and negative. As public sector organisations increasingly move their customer services towards web-based platforms (a process known as 'channel shift'), partly in order to bring down costs, the volume of feedback and enquiries coming through the web and social media is likely to rise. Digital data analytics could therefore play an increasingly important role in future, in managing consumer feedback and providing insights that can help to improve the design and delivery of services.
- *Monitoring relevant publicity.* Digital data analysis can be used to monitor what the public and key influencers and groups may be saying about the organisation, in order to manage the organisation's reputation. This may involve a proactive reputational strategy, where the

organisation aims to capture and publicize positive sentiment, as well as a defensive one, where it seeks to identify and manage reputational risks.

- *Identifying and analysing public conversation and opinion on local issues.* Digital data analysis can be used to identify and analyse what the public is saying about local matters of concern, capturing public conversations and views that may not otherwise find their way into the organisation via conventional channels, such as consultations and other feedback mechanisms. These insights can then inform organisational decision-making.
- *Identifying and engaging the public in policy development.* Related to the above, but more ambitiously, some participants flagged the possibility that digital data analytics could be used to support a more ground-up approach to policy development. This would entail moving from a top-down model of engagement towards what was referred to as the 'co-design' of policy, where the organisation involves the public more actively in policy formation and policy discussions.

3.2.2. There are challenges and barriers to use

The potential uses of digital data analysis are likely to be limited in practice by the resources and expertise available to organisations. Given significant financial pressures and budget cuts at local level, resource constraints weigh heavily on the ability of the organisations to invest in digital data analytics. A lack of relevant knowledge and expertise within organisations can also pose challenges. Not only does the use of some of the more complex digital analytics tools (such as NodeXL and Gephi) require an investment of time in order to develop appropriate levels of expertise, but there can also be a lack of understanding and support in relation to these methods within organisations more generally. These organisational factors shape the use of digital analytics in practice and make more resource-intensive uses of digital analytics unlikely.

Digital data analytics tools and practices also have weaknesses and limitations that diminish their likely application in practice. Participants in the project, for example, raised concerns about whether data obtained through digital analytics can be representative and complete. Data may capture the views of some social media users, or the digitally enabled, but not of all local communities. Concerns were also expressed regarding how the accuracy of data, with some participants keen to point out that the quantitative reach of a tweet may tell us little about how influential a user is or what impact s/he may have. Moreover, on the topics chosen by our partners, relevant data was not found to be abundant, as we have already noted. The fact that digital data analytics tools are not designed for public sector organisations interested in local issues may explain why data about local publics is hard to access. Some participants were sceptical of the value of digital data analysis given this paucity of relevant data. However, they did not exclude the possibility that at a later stage such methods and data may come to play a more important role within their organisations.

3.3. What are the implications of the ways publics are made visible through digital data methods?

Given our specific interest in local publics and public engagement, the following questions are particularly pertinent in relation to the implications of the ways in which publics are made visible and constituted through digital data methods: (1) how *inclusive* are the representations of the public generated by digital data and (2) what *types of relationship* with the public and *forms of public engagement* do digital analytics enact?

3.3.1. Digital data analytics is a supplement to other forms of public engagement

As already noted, while digital data analytics provides new insights into the public, the data is also partial and incomplete. The unrepresentative character of the data is particularly concerning

when it comes to the use of digital analytics in the public sector, as compared to commercial contexts. Public organisations such as councils and museums are meant to represent and serve the public as a whole, rather than any particular social group, and so *inclusion* is a central normative principle. It would be troubling if unrepresentative data was taken as a faithful representation of the public and if this then formed the basis of decisions about how public resources are allocated or public services are run. Fortunately, our project partners were aware of this danger: they generally acknowledged the limits of digital data (concern over the incomplete and non-representative character of data was frequently expressed, as noted above) and saw digital data analytics as a supplement to (rather than replacement of) established methods of public engagement, such as consultations, citizens' panels, petitions, and so on. More positively, it was hoped that digital data analysis might provide access to some groups and views that may not be captured by these more conventional methods.

3.3.2. Different uses of digital data analytics enact different relationships with the public & forms of public engagement

As well as shaping who is included as a member of the public, different uses of digital data analytics also enact different types of *relationship with the public* and *forms of public engagement*. A basic distinction can be drawn here between engagement activities involving: a) information provision ('a one-way relation in which government produces and delivers information for use by citizens'); b) consultation ('a two-way relation in which citizens provide feedback to government'), and c) active participation ('a relation based on partnership with government, in which citizens actively engage in the policy-making process') (Macintosh 2003: 32). Digital data analysis can be used to assess the success of information provision, for example, if it is used to determine how messages from the organisation to the public may be distributed most effectively. Digital data analysis can be used in relation to consultation, to gain feedback from the public about services and issues, although the public may not always be aware of the fact that their views are being harvested and analysed. Most ambitiously, it can also be used by organisations to support forms of active participation, where the public is centrally involved in the development of policy in partnership with the organisation. Which if these ways the public is enacted through digital data analytics will be determined ultimately by how digital data analysis tools are used by organisations, something which will be structured in turn by the resources available to them and by their organisational aims and existing practices. As noted above, some uses of digital analytics —and so representations of the public — are more likely than others.

3.3.3. How to make analytics public

What are the broader conclusions and implications of DDA? While more limited than the hype suggests, our research indicates that digital analytics can have some value for public sector organisations. We can also reasonably expect digital analytics to become more important in future as social media and web use grows. However, in order to serve the public good, we argue that digital analytics would need to become more public itself, in three main ways.

Firstly, and most obviously, for analytics to be public, it needs to be *available to the public to use*. Analytics tools, data, and expertise need to be accessible to public organisations and social/community groups and analytics tools need to be designed in ways that remain open to varied public uses and purposes. Secondly, if analytics is to serve the public good, it needs to be *open to public supervision*. Whereas the code, algorithms, and methodologies behind analytics tools and software are often proprietary and black-boxed, we argue that if they were made more public, they could be scrutinized and debated, and more public confidence in them could be established. One way of achieving this could be through open source approaches. Another might be through regulation, so that code, algorithms and methodologies are transparent and open to public scrutiny. Another approach could be regulation; the involvement of the public in setting

appropriate rules and standards may help to ensure that digital data analytics (and digital data analytics industries) serve the public interest.

Thirdly, we think that the process of analytics should *allow space for public reflexivity and participation*. Analytics could be viewed not just as a way for experts to track and know the public with ever-greater precision, but could be understood in more participatory terms, as offering the means of representation by which publics can come to know and constitute themselves in new ways. We suggest that thinking of analytics as offering a way for publics to constitute themselves means, as John Durham Peters (1995: 16) argues, that 'in acting upon symbolic representations of "the public" the public can come into existence as a real actor'.

But these proposals raise a number of questions. What are the political and normative consequences of analytics becoming more public as we suggest here? Given the possibilities analytics offers for regimes of governance and control, for privacy invasion and transparency evasion, is analytics an appropriate tool with which to overcome emergent digital divides based around the capacity to do digital data analysis? Do we wish to see publics and other communities (such as social movements or activist groups) digitally enabled through analytics in the same way that other digital technologies have formed a part of grassroots digital enablement? These new research questions are suggested by the findings of our pilot study.

4. Key Issues, Impact, Next Steps

4.1. Key Issues

In conclusion, the key issues raised in our study are:

- We need a more nuanced and realistic understanding of what digital data analysis can make possible, based on actual uses rather than speculations about what is possible.
- Digital data analysis offers opportunities for public sector organisations to know and engage their publics, but there are also a number of challenges and barriers to use, and usage limitations.
- There are ways in which analytics could become more public and so serve the public good, but these strategies raise a number of normative questions in relation to the democratization of analytics, or in making analytics public.

4.2. Impact

As the project has only just drawn to a close at the time of writing, it is very early to talk about impact. To date, we know that all of our partners have circulated reports that we shared with them within their organisations. The first was a partner-specific report, summarising what data the analytics experiments generated about each partner organisation. These reports were intended to provide a snapshot of what is possible with digital data analysis tools, rather than a comprehensive account of available data. The second report was a general guide to tools, listing some tools and explaining what they do and how they work. Our key contacts in the partner organisations report that both the reports and the attendance of some of their staff at our end-of-project workshop have fed into the drafting of their social media policies, which each organisation was in the process of writing towards the end of 2013. In this respect, the knowledge of digital data analysis tools which DDA facilitated was particularly useful, according to our key contacts. In addition to this, in one partner organisation, the communications team held workshops based on our general guide to tools, to introduce the staff in the team to a range of available digital data analysis tools.

4.3. Next steps

Given the concluding comments in (3) above, a future research priority is to research social/community uses of analytics empirically, and also by addressing the normative questions set out at the end of section (3).

The visceral reactions of our partners to the data visualisations that some of the analytics tools produced and that we shared with participants leads us to propose that, in order to enhance understanding of potential public uses of digital data analytics, there is a need to explore and understand how data visualisations get received, perceived and consumed. This subject has formed the basis of the successful Seeing Data project bid outlined in (6) below.

5. Dissemination

5.1. Papers

Kennedy, H., Moss, G., Birchall, C. and Moshonas, S. (submitted) 'The numbers do not speak for themselves: the rhetoric and practice of big data analysis' (*Information Communication and Society*).

Moss, G., Moshonas, S., Kennedy, H. and Birchall, C. (in preparation) 'Enacting 'the public' through digital data: digital analytics and public engagement' (for submission to *Journal of Information Technology and Polity*).

Kennedy, H. and Moss, G. (in preparation) 'Can analytics be "really useful"?' (for submission to *Big Data and Society*)

5.2. Presentations

Kennedy, H. (2012) 'Social media insights as expertise', Digital Expertise Workshop, EPSRC *Digital Transformations of Community & Culture Network+* event, Brighton, November 2012.

Kennedy, H. and Moss, G. (2013) 'Digital data analysis, public engagement and the social life of methods', Digital Practices, In/visible Communities, EPSRC *Digital Transformations of Community & Culture Network+* interim meeting, Brighton, February 2013.

Kennedy, H. and Birchall, C. (2013) Accessing and Using Big Data to Advance Social Science Knowledge workshop, Oxford Internet Institute, Oxford, March 2013.

Kennedy, H., Moss, G., Birchall, C. & Moshonas, S. (2013) 'Digital data analysis, public engagement and the social life of methods', Digital Practices, In/visible Communities, EPSRC *Digital Transformations of Community & Culture Network+* annual meeting, Leeds, September 2013.

Kennedy, H, (2013) 'Social media intelligence work and expertise', *Expertise with/in Digital Media* panel, World Social Science Forum Conference, Montreal, October 2013.

Kennedy, H. (2013) 'How can we research social media data mining / big data analytics?', School of Creative Studies and Media, University of Bangor, invited seminar presentation.

Kennedy, H. (2013) 'Big data analytics: reflections from the social science/humanities borderland', School of Computing, University of Leeds, invited seminar presentation.

Kennedy, H. and Moss, G. (2014, to be confirmed) 'Making analytics public: really useful analytics and public engagement', *Really Useful Analytics and the Good Life* panel, ICA Conference, Seattle, May 2014.

6. Funding

DDA Principal Investigator Dr Helen Kennedy has been awarded two grants from the AHRC which build on the research carried out on this project. These are:

- (January 2014 – March 2015) **Seeing Data: are good big data visualisations possible?** (AH/L009986/1) £280,067.49 FEC; £224,053.99 actual contribution, to explore the reception and perception of data visualisations.
- (February 2014 – August 2015) **Understanding Social Media Monitoring** (AH/L003775/1)**,** £168,893.00 FEC; £135,114.40 actual contribution, AHRC Fellowship, to develop understanding of the role of social media monitoring in the production of social life.

7. References

Macintosh, A., 2003. 'Using Information and Communication Technologies to Enhance Citizen Engagement in the Policy Process', in: *Promise and Problems of E-Democracy: Challenges of Online Citizen Engagement.* Organisation for Economic Co-Operation and Development.

Peters, J.D., 1995. Historical Tensions in the Concept of Public Opinion, in: Glasser, T.L., Salmon, C.T. (Eds.), *Public Opinion and the Communication of Consent.* Guilford Press.