Ipsos MORI
Scotland

APPROACH

November 2018

ARTICLES

NEWS & PUBLICATIONS

Welcome to *Approach* the newsletter from the social research team at Ipsos MORI Scotland

Welcome to this winter's edition of our Ipsos MORI Scotland newsletter. It was a real pleasure to see so many of you at the Future of Research event we held at the awe-inspiring Royal College of Physicians in Edinburgh earlier in the year. Continuing that theme of innovation, this newsletter takes you through some of the latest developments in research and – most importantly – how they can help us understand how best to address some of the thorny policy issues that Scotland faces.

In a context of declining response rates, how can large-scale population surveys adapt - while still delivering the high quality data that's essential for National Statistics? Push-to-web methodologies are one way of doing this, which we've been trialling at Ipsos MORI with our partners in ONS and elsewhere. Chris Martin and Patten Smith take us through when to use push-to-web surveys, and how to minimise survey error as much as possible when you do.

Much public policy requires changing people's behaviours – from measures aimed at encouraging people to eat more healthily, donate blood or get vaccinated, to increasing the proportion of journeys made by public transport. But how can we best combine behavioural science and research to deliver insights that inform policy and strategy? Sara Davidson (@SaraMORIScot) and Colin Strong cover Ipsos MORI's approach to behaviour change and the steps involved in identifying, developing and assessing the effectiveness of potential behaviour change interventions.

The last – but certainly not the least – innovation we cover here is text analytics. We are drowning in data – as Rachel Ormston (@rachelormston) and colleagues highlight, the volume of data we're generating has increased at an astonishing rate. Text analytics is one way in which we can make sense of huge volumes of unstructured data – Rachel shares some examples of how we've been applying this method to make sense of social research problems, and looks at how it could develop in future.

I hope you enjoy reading this edition of Approach – if you would like to discuss anything that's mentioned (or not mentioned!) here, please get in touch. Have a very merry Christmas, and a happy Hogmanay.

Emily Gray
Managing Director, Ipsos MORI Scotland
@Emily_IpsosMORI

The development of Push-to-Web surveys

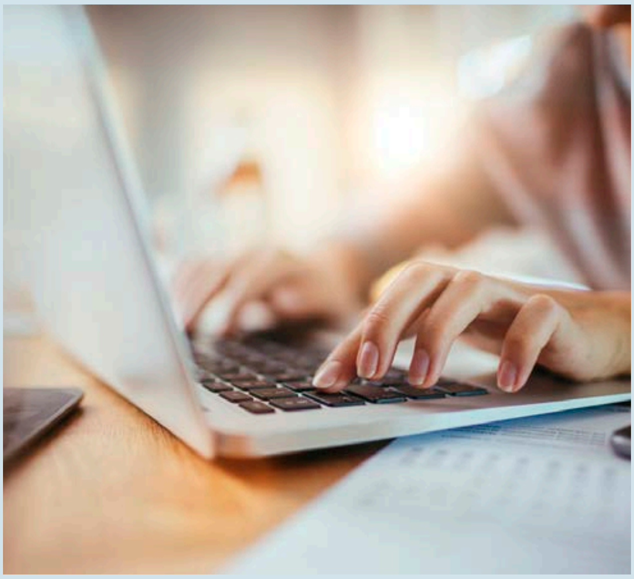
By Chris Martin and Patten Smith

Push-to-Web will become an increasingly used survey approach. But what is Push-to-Web and when is it used? Push-to-web is the use of offline contact to encourage people to go online and complete a web questionnaire. It is used when we want to implement a high quality random probability survey in a population for which we don't have a sample frame with high email address coverage. In the UK the kinds of surveys which are typical candidates for push to web are ones which in the past have largely used face-to-face data collection methods.

And why is push-to-web used? The survey world is changing, and both push and pull factors are driving surveys towards online data collection. On the push side, face-to-face surveys are subject to both declining response rates and rising costs, whilst simultaneously research budgets have been declining. And more recently, the Government Digital Transformation has strongly encouraged the adoption of digital methods for Government sponsored work. Simultaneously, strong pull factors have been operative: online survey methods are cheaper and faster than their offline equivalents, and the public is now coming to expect online official contact. Furthermore, recent methodological evidence shows that the relatively low response rates associated with online surveys do less damage to estimate accuracy than has been previously thought.

And if we want online data collection and we want random probability general population samples, then we are forced to use push-to-web methods, simply because none of the good coverage sample frames include online contact details.

Ipsos MORI has developed an approach to the implementation of push-to-web surveys for general population surveys, an approach that is geared towards minimising survey error and that is based on our extensive



knowledge of the methodological literature and on our own experiments. Our approach has four central elements:

1. We draw address samples from the postcode address file (PAF).

2. We send up to four mailings to each sampled address – an initial letter with a URL, a first reminder letter with a URL, a second reminder letter with a URL and an accompanying mail questionnaire, and a final reminder with just the URL.

3. We ask all adults at the sampled address to participate, to a maximum of two.

4. We offer a conditional incentive for participation.

How do we ensure our approach minimises errors of representativeness and measurement? Starting with representativeness, it's important to acknowledge that one in ten adults lives in a household with no internet access (although this is decreasing, and nowadays half of even those aged 75+ are web-connected), that push-to-web surveys have lower response rates (rarely more than 25%) than the face-to-face ones they are coming to replace and that, compared to the general population, web-only samples contain too many of the middle-aged, the well-educated, high earners, and (unsurprisingly) frequent internet users. The Ipsos MORI approach deals with this through the simple expedient of including an offline (postal questionnaire) reminder. Postal questionnaire respondents tend to be older, less well educated, poorer and use the internet less that do the online respondents. Including a postal questionnaire goes a long way to compensating for the bias in online samples.

Another representativeness issue that has emerged in push-to-web surveys concerns how we get from a random probability sample of addresses to a random probability sample of individuals. This is an issue because recent methodological research has demonstrated that the conventional random adult selection methods used in face-to-face and telephone surveys, Kish grids and last/next birthday methods, simply do not work for self-completion surveys. But what alternative is there? One method that has been tried is to forgo selection altogether and ask all adults to take part. But if conditional incentives are used (as they must to maximise response rates) this approach provides an obvious temptation for fraud: because a household completing four questionnaires will receive an incentive four times larger than one completing a single questionnaire, there is an obvious temptation to invent household fictional members. And there is circumstantial evidence that this does indeed happen.

Our approach lies somewhere between the single random selection and take-all adults approaches and avoids the worst difficulties of each. In households containing one or two adults we simply ask all adults to take part, and in households containing three or more adults, we ask any two to participate. We no longer need to worry about people not following random selection instructions because we no longer ask them to do so. We have also considerably considerably reduced the temptation for fraud by limiting the total amount a household can claim in incentives. Of course, giving up on random selection feels all wrong, but as a matter of simple arithmetic it cannot have a major impact on overall sample estimates: only 15% of households in the UK contain more than two adults and any biases in these will be hugely dampened by the data from the other 85% of addresses in which there is no selection bias.

What about measurement errors? The good news is that methodological work has shown us that in general online questionnaire deliver high quality data. However, knotty problems remain. How can we be confident that respondents' answers are not affected by whether they happen to answer online or by means of the mail questionnaire? And how can we ensure that online answers do not differ by device used (desktop, laptop, tablet or smartphone)? The former question is relatively easily dealt with by adopting what Dillman terms unimode design principles: use identical wordings and maximise similarities across modes in question and answer code formats, and in visual design. Answering the second question is the subject of much current methodological work but the emerging consensus is that we should adopt "mobile-first" design principles which are based on the finding that if we get a questionnaire right for smartphone administration, then tablet/PC administration will look after itself. What this means in practice is that we need to keep questions and answer lists short, and to minimise cognitive burden. These principles are hardly new – they have been accepted as good practice canons since the 1950s!

Push-to-web surveys are a good choice for survey topics that are interesting or important for the study population, and where the questionnaire is relatively short and simple. However, push-to-web surveys that use long questionnaires and cognitively burdensome questions will increase break-off rates and reduce data quality compared to interviewer-administered modes which benefit from interviewer help and encouragement.

The use of push-to-web surveys is likely to grow considerably over the next few years, and we will continue to test different strategies to maximise quality and value for money.

Chris Martin

Patten Smith

For more information, contact **Chris Martin**

The best of both worlds: **Combining behavioural science and social research for optimal insights**

By Sara Davidson and Colin Strong

Changing behaviour is at the heart of much public policy. While this is perhaps most evident in the health sphere – for example, in the form of measures aimed at encouraging people to eat more healthily, stop smoking, get vaccinated, and so on – it is increasingly so in other areas too. Indeed, many current Scottish Government's National Indicators and strategies relate to behaviour change, from increasing the proportion of journeys made by public transport; to reducing household waste production, and increasing levels of volunteering. Of course, it's not just among the general public that policy makers seek to promote behavioural change – Ipsos MORI is currently undertaking a study for the Scottish Government aimed at identifying how it can encourage growth behaviour among the country's Small and Medium Enterprises, for example, as a part of its broader ambition to strengthen Scotland's economy.

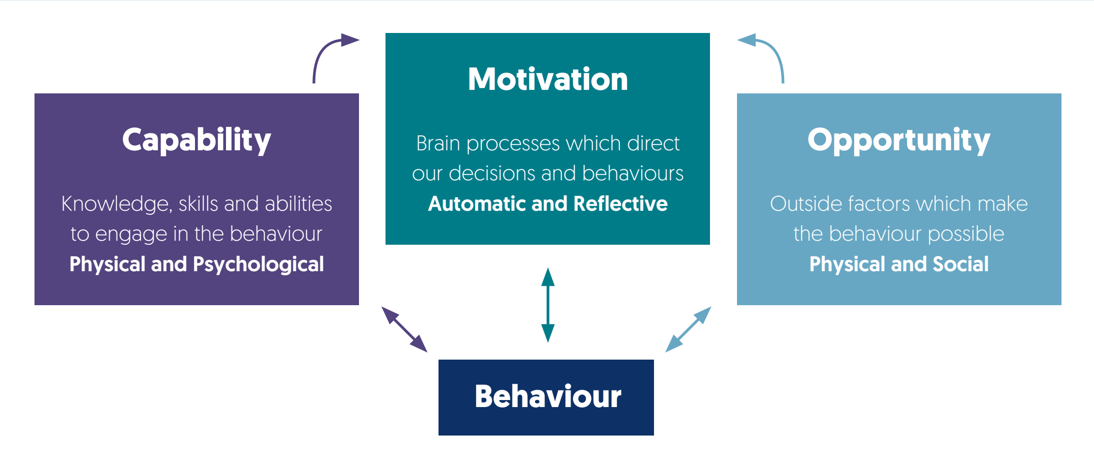
Notwithstanding such diversity, common to all behavioural change programmes is a need to first understand *current* behaviours in order that appropriate interventions can be identified. This is where research is crucial. But how exactly do we combine behavioural science and research to deliver optimal insights that can inform policy and strategy?

Before considering this question, it is important to recognise that changing behaviour is not always easy. We find it hard enough to change our own, after all. Everything from biting finger nails and turning up late to family events, to the more serious examples of drinking too much alcohol or failing to do enough exercise; these are all ways in which we know we could do better, and indeed often want to, but simply fail to do so. Then there are the times when we fall into habitual behaviour and don't have any particular desire or motive to change, even though we are vaguely aware that there are other (perhaps better) options available – a good example is not switching energy supplier, even though we know people who have switched and saved money as a consequence. This is because behavioural patterns become ingrained and even for one-off behaviours there can be strong psychological, social and environmental factors which discourage change.

Nevertheless, attempts to change behaviour are often made without any analysis of existing behaviour, or any theory of the mechanisms underpinning it. Even when theoretical models or theories are chosen, they often do not include the full range of influences on behaviour, and exclude key drivers. So, for example, the 'Theory of Planned Behaviour' (Ajzen, 1991) and 'Health Belief Model' (Nancy & Becker, 1984), though valuable in many contexts, do not address directly the impact of impulsiveness, habit, self-control, associative learning and emotional processing (West, 2006). Improving the effectiveness of behaviour change work, means adopting a systematic approach that provides not only an understanding of the behaviour that is to be changed, but a framework for determining how interventions can be used to make use of this intelligence.

Our approach draws directly on the work of academics led by Susan Michie from University College London who undertook a systematic review of the literature and engaged with experts to identify frameworks of behaviour change interventions. The result was a 'Behaviour Change Wheel' (Michie, van Stralen & West, 2011; Michie, Atkins & West, 2014) which provides a multidimensional approach to understanding what is driving a given behaviour or set of behaviours, and identifying relevant evidence-based interventions. At the heart of this approach is the COM-B system, which is used to provide a behavioural analysis of the targeted behaviour. From this analysis, intervention functions that are linked to key barriers of a targeted behaviour are considered and then developed through considering a number of policy categories. A key strength of the approach is that it maintains a focus on intervention in all its possible forms, rather than narrowing in on a subset of intervention types. It ensures that we consider not only 'external' influences (such as 'nudges'), but also those which relate more to personal agency, such as education and training.

Figure 1. COM-B sources of behaviour and their relationship



Another strength of the approach is the way it integrates context. There are many cases of attempted behaviour change which have failed because this has been poorly considered. Social and market research has always been the study of the broad range of ways in which our behaviour is shaped, and the context in which the behaviour takes place is critical. For example, whilst it is tempting to consider that a mobile-based app can help engage those with a medical condition to take their medication, there may be side-effects of the medication – like lethargy – that interferes with or prevents an individual from engaging in their regular social activities. Hence, regardless of the number and quality of reminders, that intervention will simply not be effective as its development didn't consider this broader context and the barriers that prevent an individual from engaging in the desired behaviour.

Our approach involves four stages of 'identification': identifying the behaviour (current and desired); identifying the barrier to the desired behaviour; identifying intervention options; and identifying means of improvement.

Stage 1: Identifying the behaviour

All too often there is a desire to change behaviour that is not clearly defined. We may want to address health problems such as too much pollution from traffic, high rates of obesity, or increase adoption of lower-risk smoking products, and so on. But these are not 'behavioural' targets per se. Weight loss is not behavioural, but increasing physical activity and reducing calorie intake are broad behavioural targets. The following questions can be helpful in terms of specifying the target behaviours:

- Who needs to perform the behaviour?
- What do they need to do differently to achieve the desired change?
- When do they need to do it?
- Where do they need to do it?
- How often do they need to do it?
- With whom do they need to it?

Another decision that needs to be made is which target behaviours to focus in on – there may be multiple options (for example, different segments of the public, those buying a product for the first time versus encouraging others to switch product). A useful means of prioritising is to consider:

- The likely impact of the behaviour, were it to be changed.
- How easy it is likely to be to change the behaviour
- Possible spill-over effects on other behaviours (could it create a virtuous loop?).
- Ease of measurement – can change actually be measured?

Stage 2: Identifying the barriers

The Behaviour Change Wheel identifies three attributes that are necessary and sufficient to change behaviour (the 'COM-B' model of behavioural change, see Figure 1):

- **Capability** – including physical capability (eg. the skill, strength, stamina) and psychological capability (knowledge or psychological skills to engage in the necessary mental processes).

- **Opportunity** – including physical opportunity (as afforded by environmental factors such as time, resources, locations, cues) and social opportunity (as afforded by interpersonal influences, social cues and cultural norms that may influence the way we think).

- **Motivation** – an intention to perform and automatic processes that guide the behaviour.

The question then becomes how we go about understanding these different components of behaviour and identifying where barriers exist. This is where social research comes into its own – as we have a broad range of different tools to examine these issues. The methods that are used will depend on the nature of the issue – for example, where the choices are semi-automatic in nature, then observational/ethnographic approaches may be best suited. If the topics are more accessible to human understanding, then a 'behavioural survey' may be more appropriate (this is a survey which focuses on behaviours and their drivers and barriers rather than on attitudinal measures).

Equally, in some cases a lot is already known about the drivers and barriers of behaviour and it is not necessary to undertake further research. Instead, applying the COM-B lens to existing evidence may provide many of the answers (and any gaps can be addressed by smaller scale research, if necessary).

Stage 3: Identifying interventions

Moving from an understanding of behaviour, through to identifying appropriate interventions will always be something of a conceptual leap. Nevertheless, it is important to do this in as systematic and evidence-based a way as possible, to avoid a scatter-gun approach. The analysis of the barriers for a specific behaviour provides a starting point to develop interventions in a more systematic way. That is, according to the Behaviour Change Wheel process, barriers that fall under certain sources of behaviour can be best addressed by specific 'intervention functions'. For example, barriers of reflective motivation might be best addressed through intervention functions that focus on Education (increasing knowledge or understanding), Persuasion (e.g. advertising), Incentivisation (creating an expectation of a reward), or Coercion (creating an expectation of a punishment or cost). Building on this, the Behaviour Change Wheel process then helps to guide the selection of policy categories through which interventions are developed, which include Legislation, Regulation, Fiscal, Environmental or Social Planning, Service Provision, Guidelines and Communication/Marketing.

Stage 4: Identifying improvements

Much of the time interventions are implemented without any real thinking about evaluation. But this is a vital step that should not be ignored, as there is a chance that the interventions may not work – or have an impact that was not anticipated. Evaluations are typically undertaken using a range of primary research methods, from survey work to qualitative interviewing/focus group and data or documentary analysis. The findings from this stage can be used not only to identify which interventions work well (and should therefore be rolled out), but also how to improve them – feeding back into the beginning of the overall process and representing a process of continuous improvement.

Perhaps the most important strength of the COM-B model is the transparency it offers: In a time of fiscal constraint, in particular, change will only happen if stakeholders understand, support and are part of the process. The Behaviour Change Wheel process offers an accessible, evidence-based means through which potential interventions can be identified, developed, and assessed – and one that can be used from the start to the end of the policy process.

Sara Davidson **Collin Strong**

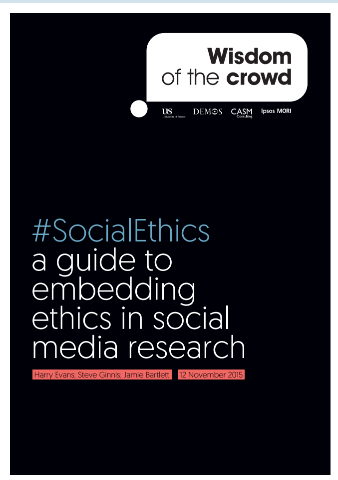
For more information, contact **Sara Davidson**

Unlocking unstructured data: **How text analytics can extract meaning and provide new insight**

By Rachel Ormston, Sylvie Hobden & Josh Keith

We are, as we are constantly being told, living in a world of 'Big data'. The volume of data we are generating as individuals and organisations has increased at an astonishing rate. Each time we buy a new jumper online, post an ill-advised late-night tweet, or use our council's website to register a complaint or request a service, we are adding to the vast store of data that companies, governments and other organisations hold about us.

The advent of big data raises many questions – not least around the ethics of how this data is collected, stored, accessed and used. The Cambridge Analytica scandal only served to highlight what was already an area of concern for researchers trying to establish appropriate ethical boundaries around the use of social media data – Ipsos MORI and Demos published a guide to embedding ethics in social media research in 2015. But even when appropriate consideration has been given to individual rights and privacy, using big data still poses substantial challenges. New types and volumes of data demand new approaches. Careful manual review and coding of each response might remain the best way of coding open-ended questions to a modest-scale survey, but it is unlikely to be the best way to approach a dataset of, say, 60,000 tweets. At the same time, our clients are looking for cost-effective ways of analysing more traditional data – including consultation responses (the volume of which can sometimes catch organisations by surprise), open-ended questions to surveys (particularly where these are repeated over time), and unstructured feedback data (collected by many organisations, but often without the resource to analyse it systematically).



It is the desire to extract meaning and insight from varied and, in some cases, vast sets of unstructured data that has led to the development of 'text analytics' as a field. This article aims to:

- Provide a succinct introduction to what we mean by text analytics and how it works.
- Share some examples of how text analytics can be applied to help address social research problems.
- Reflect on both the current limits of text analytics, and how its uses might be developed in the future.

What do we mean by text analytics?

'Text analytics' is used to describe the process of identifying and classifying text-based data to derive patterns, trends and insight. While, in principle, this could be done manually, the term is more commonly used to refer to a combination of **human coding, natural language processing** and **machine learning** to replicate human coding decisions across large datasets in a **time-efficient** manner. A variety of software applications now exist to support this process – including IBM Modeler (an add-on module within SPSS) and Method52 (developed by the Centre for the Analysis of Social Media). Each of these applications uses a slightly different combination of techniques to classify and analyse data, including:

- **Key word matching** – in which the programme identifies instances of key words and phrases chosen by an analyst.
- **Probabilistic coding** – in which the programme uses machine learning and natural language processing to replicate coding decisions made by an analyst.
- **Probabilistic categorisation** – in which the programme classifies data into 'concepts' it identifies as common combinations of nouns, verbs and adjectives. These concepts are then grouped qualitatively into appropriate themes and sub-themes by analysts.

- A key distinction here is the extent to which the programme is designed to be more **'human-led'** – with researchers specifying which words or phrases to search for, or undertaking initial coding which the programme then attempts to replicate – or **'machine-led'**, where the programme itself decides how the data should be grouped and coded, using probabilistic models to group words or phrases based on how often they co-occur and in what proximity to each other. Text analytics programmes will vary in how exactly they combine human and machine-led approaches – something that should be considered in choosing which programme is most appropriate for any given research project. However, in practice, all text analytics-based analysis will involve some human intervention – it is not a 'push button' solution. Typically, there will still be a need for analysts to:

- **Review whether the data requires any initial cleaning** – e.g. spell-checks, dealing with 'don't know', blank, or atypically long or short responses.

- **Develop or review initial code frames**.

- **Review and refine code-frames and/or rules** based on interim and final outputs.

But although there is no 'magic button' that can be pressed to instantly generate insight, text analytics does offer huge potential savings of time and money, particularly where datasets are either very large, or data collection is recurrent, since once a text analytics framework is set-up and validated, it can be (almost) instantly reapplied to subsequent datasets.

What do we mean by text analytics?

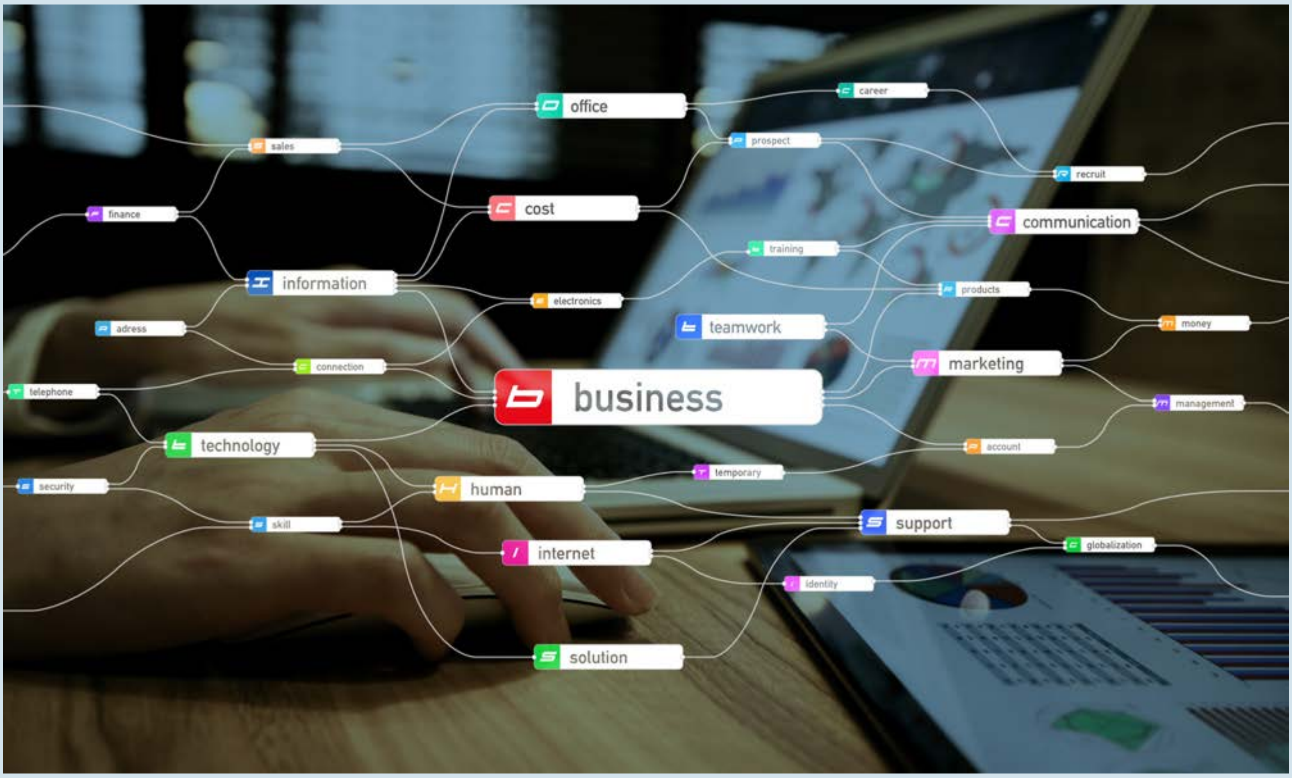
The use of text analytics is still relatively new in social research. However, in recent years it is increasingly being used to provide insights that would not otherwise be accessible given available resources. At Ipsos MORI, we have used the approach to analyse a variety of different unstructured text data, including:

- **Consultation responses** – We have used text analytics to analyse responses to a number of large-scale government consultations, where more traditional human-coding would have been prohibitively time consuming. For example, our analysis of the Schools that work for everyone consultation, on behalf of the Department for Education (England) used text analytics to establish the range and relative weight of issues raised across 80,099 individual answers to 30 open ended questions.

"although there is no 'magic button' that can be pressed to instantly generate insight, text analytics does offer huge potential savings of time and money, particularly where datasets are either very large, or data collection is recurrent"

- **Staff feedback** – Staff are often asked to give feedback or reflections on their experiences at work, particularly in response to training courses. The NHS Leadership Academy in England has collected thousands of impact statements written by staff who have been on their courses, but the length and volume of statements meant that systematically analysing them was challenging. We used text analytics to help identify the ways in which participants felt the courses had impacted on their professional skills and practice, and to identify suggestions for improvement.

- **Social media** – Social media analysis is probably the most widely known application of text analytics. We have used text analytics to:
  - Track the volume and direction (positive or negative) of Twitter discussion about each party leader during UK General Election leader debates. The analytic model was built in advance, so that responses to each leader on Twitter could be shown live during the debates.
  - Understand how food businesses engage with their customers on Twitter, mapping both the volume of interactions and apparent outcomes (e.g. offers of vouchers, confirmation that complaints will be investigated, etc).
  - Understand how current and potential students talk about different Universities.



Limits and future possibilities

Given its (relatively) low cost, speed and scalability, it is tempting to see text analytics as the solution to all our research headaches. Why spend hours meticulously coding open-ended data if your computer can do it for you? But of course, things are not quite so straightforward. As we have already stressed, text analytics is not an instant, 'push a button' solution – the need for human involvement is reduced, not removed. Other key considerations that analysts should reflect on before opting to text analytics as a solution include:

- **Computers are not perfect** – any analysis using text analytics is likely to include some false positives and some false negatives. This is due to programmes under- or over-allocating comments to particular themes depending on how 'cleanly' particular words or phrases associate with those different themes.
- **It is not a solution to poor question design** – if questions are too broad or ill-defined, then a text analytics programme may struggle as much as humans to find clear patterns in the responses.
- **Computers are not (currently) as adept as humans at reading emotion** – while text analytics programmes can use dictionaries and linguistic rules to pick up straightforward expressions of emotion, they struggle to detect more subtle sentiments, particularly sarcasm.

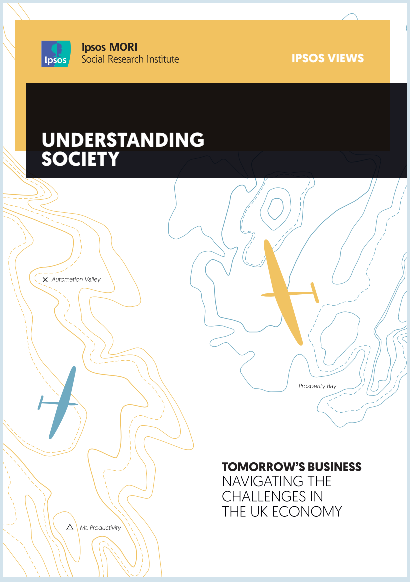
However, while text analytics may not be the answer to every problem, it undoubtedly offers huge potential to extract insight and meaning from data that may otherwise be too large or unwieldy to use. Text analytics could be used to help companies and organisations extract more value from unstructured data they collect which may currently be used to inform actions at an individual, but not a collective level. For example, collections of staff or user feedback reports, or internal reports and reviews could be reanalysed to understand key priorities and issues and how these are changing over time. But this is just one suggestion of many - the potential future applications of text analytics are almost as numerous as the datasets out there awaiting the time and resources for further analysis.



Rachel Ormston **Sylvie Hobden** **Josh Keith**

For more information, contact **Rachel Ormston**

News & publications



Understanding Society

The latest edition of the Ipsos MORI Social Research Institute's Understanding Society delves into the landscape of UK businesses – not only the contemporary and upcoming challenges that threaten their success, but also some of the opportunities that await them:

New justice research

We've recently been conducting a number of interesting studies in the justice field, two of which have just been published:

- Our review of the Aberdeen Problem Solving Approach and;
- Our evaluation of Moving Forward Making Changes - a group-based treatment programme for sex offenders.

Quantifying kindness

Quantifying kindness, public engagement and place - findings from the first ever quantitative survey on kindness in communities and public services, conducted by Ipsos MORI for the Carnegie UK Trust.

It's all in the presentation

See the Scottish Government's innovative approach to presenting survey information in an accessible, engaging format – the Scottish Household Survey Data Comic.