Nesta

DATA FOR GOOD

How big and open data can be used for the common good

Edited by **Peter Baeck**
February 2015

Nesta...

Nesta is an innovation charity with a mission to help people and organisations bring great ideas to life.

We are dedicated to supporting ideas that can help improve all our lives, with activities ranging from early-stage investment to in-depth research and practical programmes.

www.nesta.org.uk

DATA FOR GOOD

How big and open data can be used for the common good

Contents

Nesta...

Foreword

Every day across the UK, volunteers, community groups and small charities work to make the world a better place. Their devotion and altruism are one of Britain's most valuable assets. But for the government, they are easy to overlook. The nature of governments is that they pay most attention to what they can observe, measure and count – and grassroots social action is hard for a state to see.

We think there is a modest but important role for data innovation here. Over the past year, Nesta has been working with researchers, charities and civil society groups to look at new insights that data science can provide into what is going on below the radar in communities. Of course, new data sources and algorithms, no matter how promising, are not a substitute for good judgement or understanding. But we think the projects described in this report show that better data can help provide a better understanding of the often neglected but vitally important role of social action.

We hope you find the report useful, and we welcome your feedback.

Stian Westlake,
Executive Director of Policy & Research, Nesta

1. Introduction

New ways of capturing, sharing and analysing data have the potential to transform how community and voluntary sector organisations work and how social action happens.

However, while analysing and using data is core to how some of the world's fastest growing businesses understand their customers and develop new products and services, civil society organisations are still some way off from making the most of this potential.

Over the last 12 months Nesta has grant funded a number of research projects that explore two dimensions of how big and open data can be used for the common good. Firstly, how it can be used by charities to develop better products and services and secondly, how it can help those interested in civil society better understand social action and civil society activity.

- Citizens Advice Bureau (CAB) and Datakind, a global community of data scientists interested in how data can be used for a social purpose, were grant funded to explore how a datadriven approach to mining the rich data that CAB holds on social issues in the UK could be used to develop a real–time dashboard to identify emerging social issues. The project also explored how data–driven methods could better help other charities such as St Mungo's and Buttle UK, and how data could be shared more effectively between charities as part of this process, to create collaborative data–driven projects.

- Five organisations (the RSA, Cardiff University, The Demos Centre for Analysis of Social Media, NCVO and European Alternatives) were grant funded to explore how data–driven methods, such as open data analysis and social media analysis, can help us understand informal social action, often referred to as 'below the radar activity' in new ways.

This paper is not the definitive story of the opportunities in using big and open data for the common good, but it can hopefully provide insights on what can be done and lessons for others interested in exploring the opportunities in these methods.

Nesta...

2. Citizens Advice Civic Dashboard: how DataKind UK helped Citizens Advice Bureau get more from our data

Emma Prest, Datakind and **Laura Bunt**, Citizens Advice Bureau

Introduction

Through its network of local community bureaux, digital advice services and national phone lines, Citizens Advice sees first-hand the issues that people are worrying about and the problems the country as a whole is facing. As a charity we help millions of people every year to solve problems, through offering advice, advocacy and coming together to campaign for change. We help people work through issues with debt, welfare, housing, employment, relationships, justice, discrimination and many others, responding to individual and community needs.

The data Citizens Advice collects on these issues is rich and complex, but comes in a variety of formats and does not directly match up. As an organisation already invested in using our data to achieve change and improve services, we wanted to explore whether data science methods could help us to understand more about Britain today. We are well versed in identifying trends, but could this data help us to spot new or emerging problems? What could we learn about social and economic issues if we analysed it in new ways?

This chapter documents a year-long partnership between Citizens Advice and DataKind UK to learn how data science, collaboration and hard work helped Citizens Advice do more with its data. The outcome from this project was a real-time, responsive 'Civic Dashboard' that reveals the dynamics of social and economic hardship in Britain today.

Our starting point

On average, about 5,700 people walk into their local Citizens Advice Bureau every single day; one of over 3,000 local Citizens Advice offices in communities and high streets across England and Wales. Everyday 69,000 people look for help through Advice Guide, our digital advice service, and 4,300 more reach us on the phone. You might have come to us yourself, to seek advice, to find out information or access other services. Right now, at 7.06pm on a Monday evening in January, 106 people were on Advice Guide in the last minute, and 'basic rights at work' was the most popular content searched for.

Citizens Advice helps people to solve problems. We do this by offering confidential, impartial advice and support with any number of day-to-day and difficult issues; anything from home repairs, faulty goods, dealing with debt, finding work, claiming benefits, registering to vote, coping with relationships or even taking people to court. Our network of volunteer advisors

work with millions of people every year to help them find a way forward, and we come together to influence policy, campaign for change, and devise local solutions to problems to make society fairer.

The scale and extent of these daily interactions mean that Citizens Advice sees first hand the issues that people are worrying about and what the country as a whole is facing. We know the most common concerns for people in Leicester, or Doncaster, or Merthyr Tydfil. We see questions people are most frequently asking on a Sunday night before work, or at coffee time on a Monday, or in the darkest point in the night when everything feels much worse than in the daytime. Through our services we have a window into the messy, complex, changing pressures of modern life. And it is our responsibility to use this insight to make things better for people.

DataKind UK was an obvious ally to help Citizens Advice get more out of this data. DataKind UK's mission is to help third sector organisations benefit from advances in data science. Using private sector volunteer data scientists, DataKind designs and manages data projects to help charities understand what is in their data and use it for better decision making and greater social impact. Unlike many of the third sector organisations that DataKind UK tends to work with, Citizens Advice is fairly mature in its use and handling of data.

Citizens Advice has an in–house team of data analysts that crunches, analyses and maps our data, and regularly reviews the trends in issues from across the service. This team works with our researchers, campaigners and communications teams to turn this data into action; to showcase trends, insight and evidence to policymakers, service providers, regulators and other stakeholders to help address the issues facing Britain. But though we already stretch ourselves through our data work, we knew that there was an opportunity for it to do more.

This is where we started. What more could we learn about social and economic issues from Citizens Advice data if we analysed it in different ways? For Citizens Advice, we wanted to undertake an in–depth analysis to better understand how our different data streams fit together, to better identify and analyse problems and therefore improve our (and other's) response. For DataKind UK, this was an opportunity to work with an organisation with a vast, rich and varied dataset (and quite a mature approach to working with data), and to use it to help us work better. And so a natural partnership formed.

Setting the challenge

The project started with a big, bold idea: could data science enable Citizens Advice to anticipate or even predict changes in the issues affecting people everyday, to act sooner to prevent problems escalating?

In 2012 Citizens Advice spotted a problem with payday loans which was causing too many people getting into debt. By 2014 we had worked with the Financial Conduct Authority to establish new rules that aim to stamp out bad practice from the industry, rules that came into force in January 2015. But who are the next payday lenders? We know that many people are having problems with insecure work, set on zero–hours contracts or without full employment rights. How can we make sure we are spotting other examples of sharp practice – and act before they get worse? If we know the signs for someone facing eviction, could we identify people at risk of homelessness and target resources to their support?

Our volunteer and staff advisors see these sorts of problems daily. With payday loans, lots of people within our network were talking about this as a problem and suggesting that something should be done. What we wanted to do was find more ways to make this insight visible to others, and turn stories that could come across as anecdotal into analysis that was hard to avoid as fact.

Nesta...

Fundamental to making this happen was understanding the correlations between issues, the timing of problems and how the datasets fit together. By doing so, could we build an early warning system to show emerging structural problems across England and Wales?

The DataKind project

Supported by Nesta, Citizens Advice and DataKind UK joined forces to deliver a project that had three goals:

1. Design a tool to harness Citizen Advice's data so that the organisation could better identify and react to emerging social issues in the UK.

2. Build awareness among Citizens Advice staff of new methods for mining and using data, and encourage them to open up their data for others to use.

3. Develop lessons for other charities on how to better manage and use data.

The design of the project

DataKind UK uses several different formats to support partners. Given that Citizens Advice had so much data in varying formats, both an iterative, exploratory process to understand the data and test ideas, and structured time to focus on building and refining the final product was needed. Therefore the project included two 'DataDives' and one 'DataCorps'.

WHAT IS A DATADIVE?

A DataDive is a two-day event that brings the data science community together with three or four selected charities to tackle their data problems. These events are free for both the data scientists and the charities. They provide direct data support to charitable organisations, and make the social sector aware of the power of using data in their work.

WHAT IS THE DATACORPS?

DataCorps are longer projects where a team of data science volunteers work with a charity to tackle a tough data problem. The DataCorps team work with a charity for three to nine months to clean, analyse, visualise, and otherwise make use of data to make the world a better place.

This twin-track approach allowed for:
- In-depth exploration of the data.
- Quick insights into what questions the data might be capable of answering.
- Inspiration for the charity to make them aware of what is possible.
- Relationship development between the Data Ambassadors, Citizens Advice and DataKind UK.
- Based on the trust established, a slow move within Citizens Advice to open up their data.
- Longer-term project planning possibilities through the DataCorps.

An additional benefit of the DataDive model is that it is designed to incorporate numerous charity projects. At the first DataDive three other charities took part – Buttle UK, The Access Project and Shooting Star Chase. At the second DataDive, two charities, St Mungo's Broadway and the North East Child Poverty Commission, were able to tap into and access Citizens Advice data. This meant the project helped multiple charities to develop new methods to mine their data, identify hidden issues and become more data savvy.

The Data

At the heart of this project is the rich and complex data collected by Citizens Advice. We worked with three separate datasets throughout the project.

- **Bureaux issue data:** this is a record of the issues that people bring to Citizens Advice. Each issue is recorded at three levels of detail – tier one being the most general (e.g. benefits), tier two is more specific (e.g. employment support allowance), tier three is the most granular (e.g. employment support allowance appeals). This information is recorded whenever contact is made with Citizens Advice, whether face–to–face in a bureau or over the phone.

- **Electronic Bureaux Evidence Forms:** called EBEFs. These are forms completed by bureaux staff when they want to flag an issue. These records are detailed write–ups of an individual's problem. They are also categorised as tier one, two and three.

- **Web data:** this data comes from Google Analytics and shows which pages people visit on the Advice Guide site and what search terms they use.

Citizens Advice at the First DataDive: understanding the data

At the first DataDive in June 2014, over 100 volunteer data scientists descended on the Mozilla space in central London to donate their time and skills to help charities tackle complex data problems.

The DataDive participants come from all walks of life and are brought together by a desire to do good with data and work on charity projects. Almost all of them have full–time jobs in the private sector and enjoy donating their data expertise to good causes in their spare time.

There were three other charity projects involved in the DataDive – The Access Project, Buttle UK and Shooting Star Chase. Their findings are summarised later in this chapter.

Each charity was assigned three volunteer data scientists, called Data Ambassadors, to help them prepare their data and hone their project idea ahead of the DataDive. This minimises the time during the DataDive that needs to be spent on data preparation – in analytical projects this can be as much as 70 per cent of the work. Instead the volunteers who come to the DataDive can focus on performing the analysis, and discovering new insights.

The Citizens Advice project was led by three volunteer Data Ambassadors[1] who used the event to develop a deep understanding of the Citizens Advice's data, whilst generating usable insight.

One of the main pieces of work over the course of the weekend was to begin matching up the datasets.

The web data shows which pages visitors land on, such as a page on Council Tax Benefit. The graph below shows web pages that have been matched to the Citizens Advice issues classifications and how they vary over time. The DataDive was a great way of flagging up when the web data did and did not match up with the tiered classifications.

Figure 1. **Unique page views by classification - benefits**
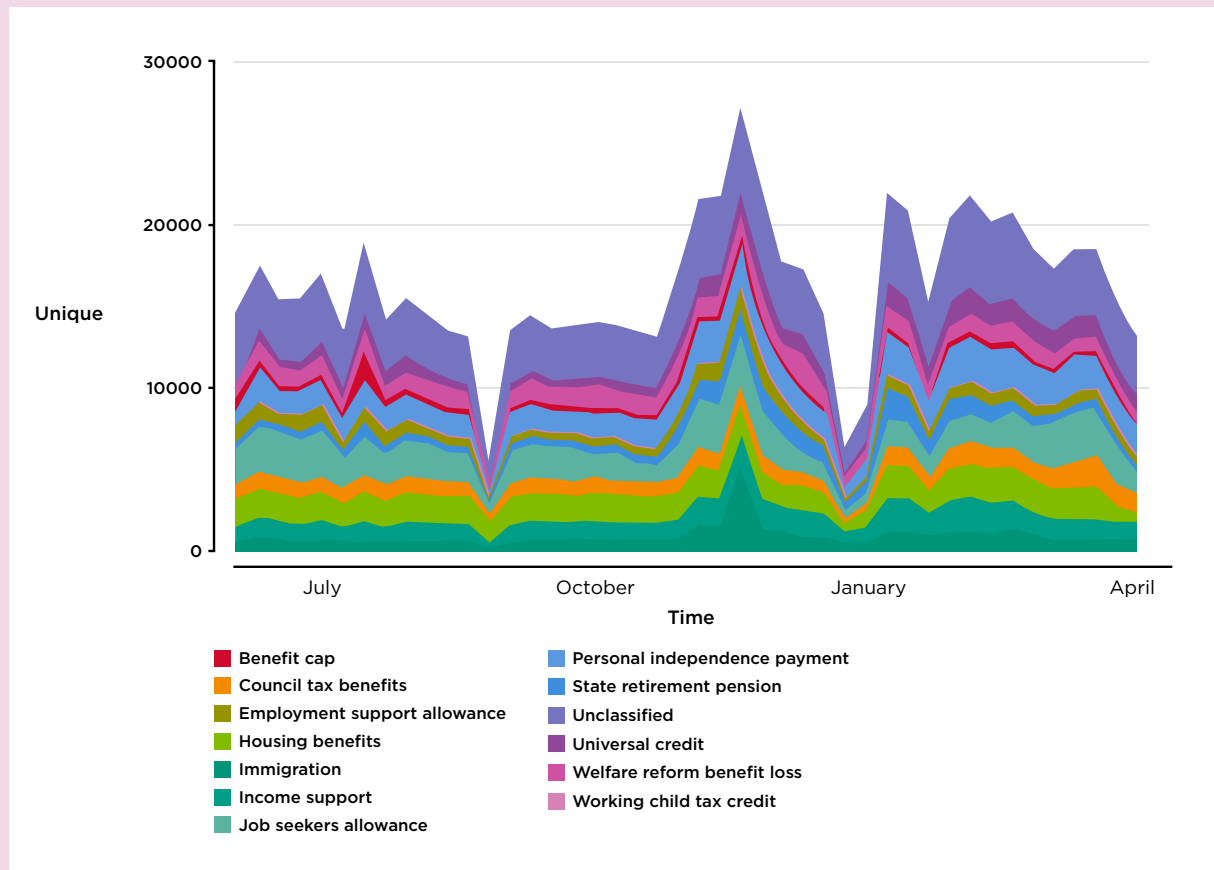
The EBEF text data contained 40 fields with a variety of demographic information about the client and six free text fields. Citizens Advice advisors fill these in to provide detailed insight into what they consider to be critical cases.

Using natural language processing – an advanced technique that gives insights from textual data – a group of the volunteers analysed the EBEF data and discovered a new term appearing frequently in the 'Other' category: 'blue badge'. This refers to the scheme used for parking by people with a disability. As a direct result of this analysis, Citizens Advice has set up a new category that allows this issue to be appropriately and distinctly tagged, rather than grouped with 'Other'.

Though these insights were not new to Citizens Advice, the fact that they confirmed what we knew already gave the team confidence in the method. Volunteers had the opportunity to explore the data and gain a deep understanding of what was in it, and began to see what could be achieved by matching disparate datasets.

## Citizens Advice at the second DataDive: comparing datasets

At the October DataDive – held at the headquarters of the Royal Statistical Society – Citizens Advice opened up their data for the other charities to use.

This marks a potential sea change in the way that third sector organisations treat data – recognising for the first time the additional social value that can be had by linking data held by more than one organisation. In addition to a continuation of the Citizens Advice project, two other organisations participated – North East Child Poverty Commission and St Mungo's Broadway.

Citizens Advice Data Ambassadors[2] began to combine the data, and to test different options for an analytical model that could predict new advice trends. They had made all data available in Elasticsearch, an open source search engine based on Lucene (an open source information retrieval software library) with an accessible and flexible interface. For this project, Elasticsearch was a perfect choice because of the combination of data types we needed to be able to search across and discover the patterns of new topics that might emerge.

The DataDive participants also explored the web data to show Citizens Advice the most frequent issues on the website. They found that tier one issues yielded few insights – Citizens Advice already knew that benefits and employment were the most popular. However, looking at the second tier of issues provided new insights. The graph below shows that requests for advice on higher education are rising, while searches related to housing benefit have notably declined.

Figure 2. **How the relative importance of issues changes over time**
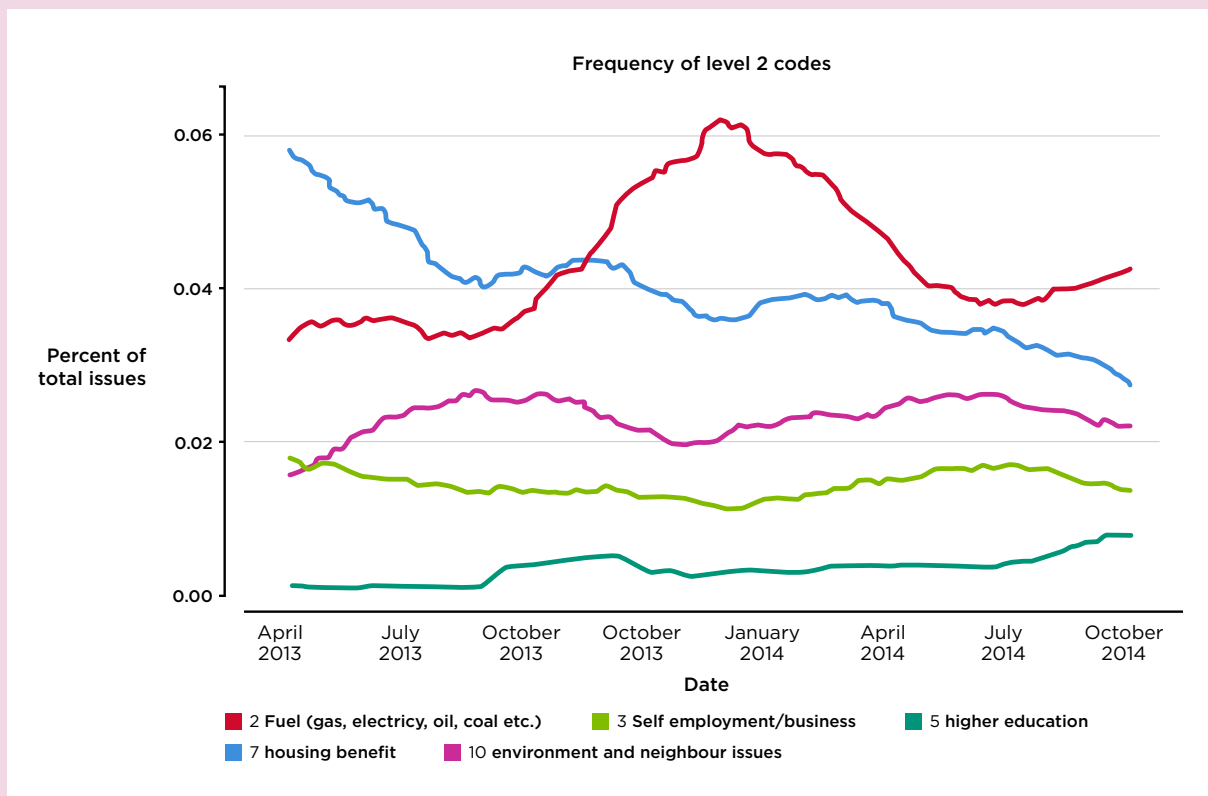
As the weekend progressed, the DataDive participants began joining up the datasets and comparing them. This was the first time Citizens Advice had been able to see the two streams of data and compare the numbers of people accessing the two channels for different issues. Below is a graph showing the web data and issue data for Universal Credit.

3.
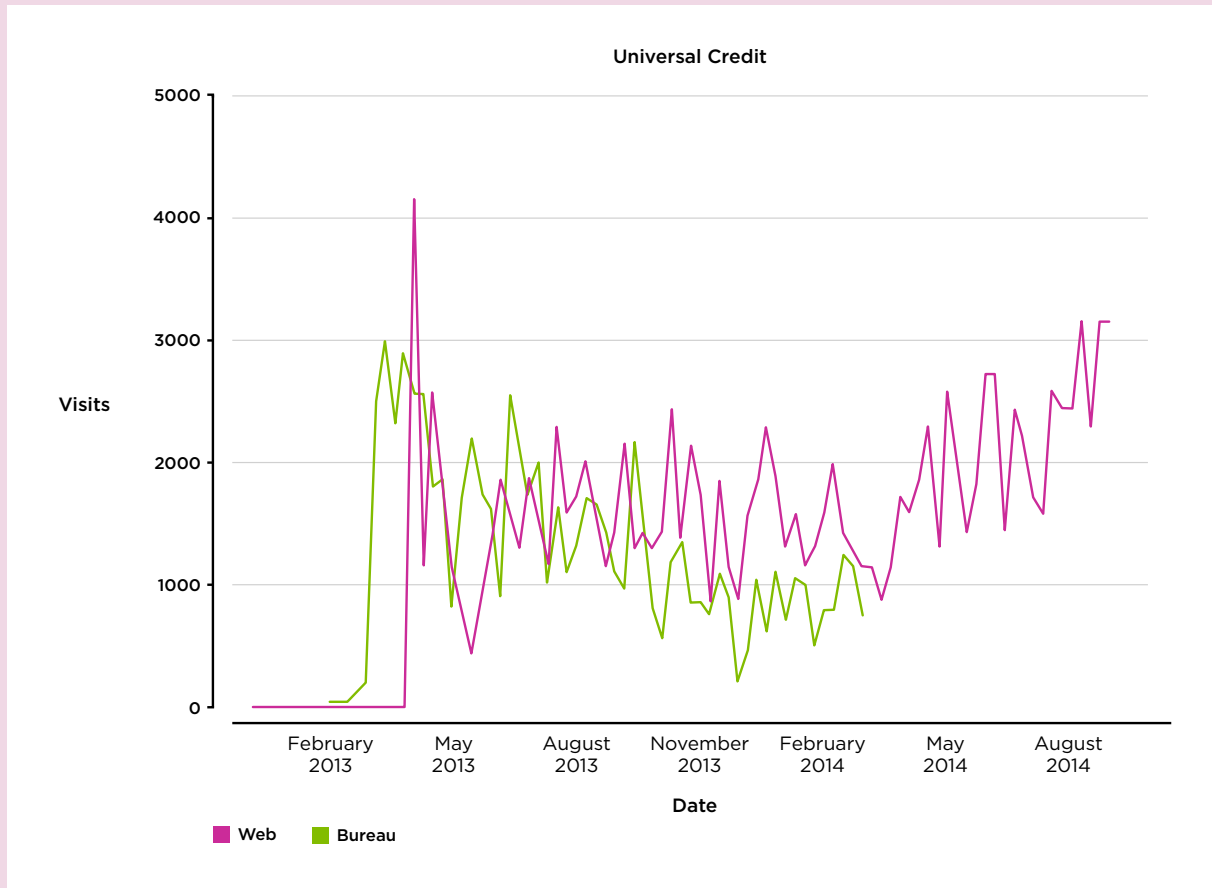
A key element of the DataDive was when one volunteer[3] created an interactive Kibana dashboard based on 44,000 EBEFs. Kibana is a web–based analytics and search tool that can be used with Elasticsearch.

This dashboard brought the otherwise impenetrable text data to life, making it searchable and understandable for the first time. Citizens Advice and the volunteers immediately saw the potential for this approach, and it became the inspiration for the final dashboard the team created.

Figure 4.

The DataCorps

In parallel to the DataDives, the core team of volunteer Data Ambassadors worked closely with Citizens Advice to understand and refine what they wanted to achieve, and to extend the work done at the DataDives.

The initial idea was to create a predictive model to identify new trends and issues. The first step in that model was to match up the datasets and load them into a single repository – Elasticsearch – to get a clear picture of all client interactions across the different channels. The Bureau issue data and the EBEF data was already categorised using Citizens Advice's issues tiers. The web data, however, was not.

Matching up the web data

The Data Ambassadors needed to join the web activity data with the Bureaux issues data and EBEF data – something that had never been attempted before. Volunteers from DataShaka – a company that specialises in unifying datasets – donated time and expertise to match specific web addresses to their corresponding Citizens Advice issue classification codes, which were to be the common data fields linking the three datasets.

Having considered several approaches, the team settled on the idea of using configurable Google Analytics segments to split up the web traffic. An advantage of this is that the segments can easily be re–configured in future when Citizens Advice update their classification codes. The web data consisted of two datasets: daily visits to the Advice Guide website and Google natural search phrases leading users to the Advice Guide site.

The team used the DataShaka platform to configure a series of transformations on the data (also known as a 'data pipeline'), to be executed automatically every night. After acquiring the datasets for the previous day from Google analytics, they converted the data into DataShaka's

data fabric for easier manipulation. They then added in the relevant tier one and two codes to the data and uploaded the data to a remote Elasticsearch cluster.

The result was that whenever someone went to the Citizens Advice website and searched for benefits and Employment Support Allowance (ESA), this could be matched up to someone who phoned up asking about ESA, or visited a Bureau with an ESA inquiry. This enabled the team to understand and compare the issues people were searching for online with the issues dealt with in person or over the phone at the Bureaux.

Tackling the EBEF data

Citizens Advice staff knew they were not getting the most from the tens of thousands of EBEFs submitted each year due to the very large volume of text and the manpower needed to examine them. But understanding what is said in these text boxes can be key to spotting new emerging issues – as was the case with payday loans. Topic analysis of the EBEF data was done to discover which words were appearing most frequently in the free text boxes using natural language processing – an advanced technique that gives insights from textual data.

To understand what was present in the EBEFs, the Data Ambassadors collected all the text from each text fields in every form and analysed the text data to identify the most significant one to three word topics. This resulted in a set of 300 topics that commonly appear, for example 'under occupancy charge' or 'legal aid'.

From prediction to a dashboard

Once the data was more or less matched up for the second DataDive, and the DataDive participants had tested out various ideas, one idea rose to the surface: an interactive dashboard to understand social issues in near-real-time.

This shifted the focus from the original concept of a predictive model to something potentially much more useful – the ability to understand Citizens Advice data across all channels, explore hunches, test ideas and spot new topics.

This was extremely liberating for Citizens Advice staff as the present system is highly dependent on the informational management team, with staff asking questions of the data and the analysts pulling out spreadsheets.

The DataCorps team ran a user meeting with stakeholders at Citizens Advice to determine what they wanted to be able to do with the data and which bits of data were most useful to them. The communications and policy teams were especially excited to get a better idea of how many people were raising which issues, what their stories were, and being able to explore more up-to-date and accurate data on which to base policy research work.

The work of cleaning and normalising the data continued so that the datasets matched up seamlessly. The team also recruited some new expertise[4] to develop the dashboard.

Building on the success from the second DataDive, the team choose Kibana as the tool for the dashboard. The main reason for this that it is designed to be integrated with Elasticsearch, where the data was stored, is open source, flexible and easy to use. The core functionality was extended through the use of D3.js (a visualisation tool that uses a JavaScript library and framework) for some of the custom visualisations. Elasticsearch staff gave the DataCorps team additional *pro-bono* support along the way – this was very helpful as the dashboard was built in a beta version of the software.

The team presented the end result to Citizens Advice staff, who immediately saw the possibilities that a more responsive, visual analytical tool would bring. The volunteers continued to adapt the tool based on feedback and ran training with users, before finally handing over the product to Citizens Advice to form a part of their information infrastructure. This is the Civic Dashboard.

What is the Civic Dashboard?

One of the primary outcomes from the collaboration between DataKind and Citizens Advice is the Civic Dashboard. The Dashboard brings together all the data from the three channels – the bureaux issue data, the web data and the EBEF data – in a searchable interface for the first time. While some trends may immediately spring out, it will be most useful for provoking new thoughts and ideas, rather than providing all the answers. It allows for exploratory analysis and the ability to drill down to see which issues are most prevalent in any particular bureau or time period.

The Dashboard will give Citizens Advice staff an immediate overview of how issues are changing across the UK. Though not an open or a public tool given the confidentiality of the data and granularity of the dashboard, it gives teams the ability to quickly visualise and analyse social and economic trends and use this to inform and shape the public debate and strengthen the impact of service delivery. This will raise awareness of what problems are happening where, and where Citizens Advice will need to target or improve interventions to make society fairer.

The following paragraphs describe the Civic Dashboard in more detail.

The Dashboard is made up of a series of separate panels. The first one gives an overview of all three datasets; the second one shows the web data; the third, the Bureaux issue data; and the fourth the EBEF data. Within each panel the graphs are all linked, so that when you change an element on one, such as the issue or the time period, the other charts respond.

## Combining the three data channels

Figue 5 brings together the three streams of data so you can compare which issues are more prevalent on which channels. The image on the left shows three striped bars: the left bar shows the web data, the middle bar the Bureaux issue data and right bar the EBEF data. This graphic shows how benefits (green) are the most common issue across all three channels, but more so in the bureaux issue and EBEF data. While it is difficult to directly compare the datasets on the same scale as one click on a website is not the same as one visit to a Bureau, this graphic shows each channel as 100 per cent and the stripes show the percentage that an issue is raised in that dataset.

The three smaller graphs on the right show the changes over time for all the data streams by week. Web data at the top, bureaux issue data is in the middle, and EBEF is at the bottom. You can therefore see whether there has been a surge of requests on one channel for an issue, or equally a drop.

Nesta..
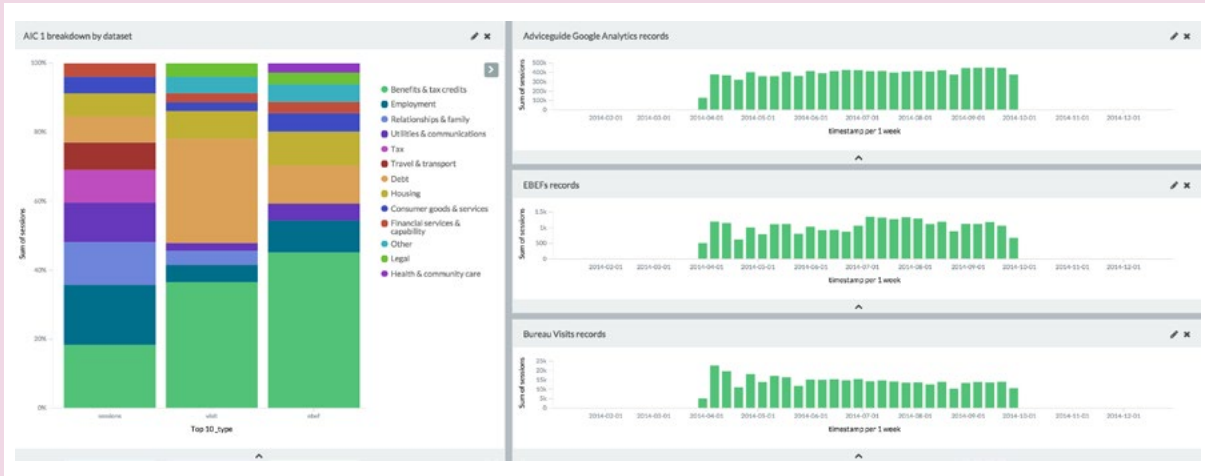
Figure 5.

A window into the web data

Figure 6 shows an overview of web traffic to the Advice Guide site. This panel has been filtered to show employment issues. The donut graphic shows Terms and Conditions of Employment and dispute resolution are the most popular pages. In the right hand stacked bar graph we see top five tier two codes over time on the website. The line graph in the bottom right shows the number of users on the site, filtered for employment. The overlapped area chart in the bottom right shows the most common Google search terms that lead to the Advice Guide.

Figure 6.
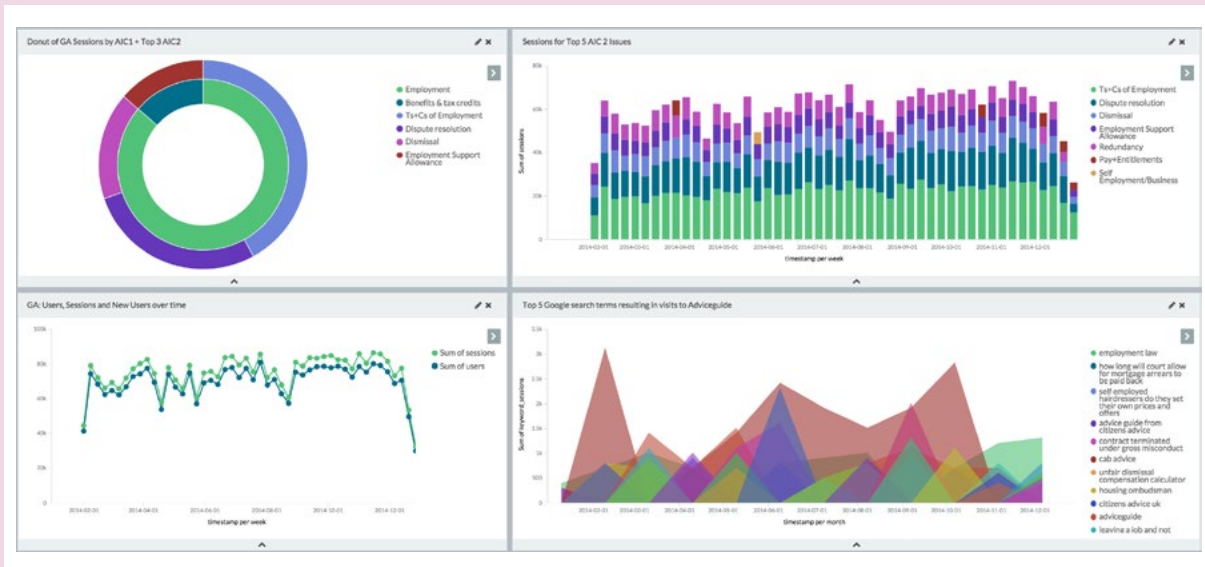
Summarising the Bureaux issue data

Figure 7 showing the Bureaux issue data is made up of three donut charts showing tier one, tier two and tier three issues across the top. The graph on the bottom left shows the most common level one issues by region. The bar graph on the bottom right shows uncommonly common issues – this means that there has been a change in the amount that the issue is appearing in the dataset compared to the average. For example the amount that people are asking about benefits and tax (dark blue) has deviated from the normal amount recently – either a lot more or a lot less reports.

Figure 7.
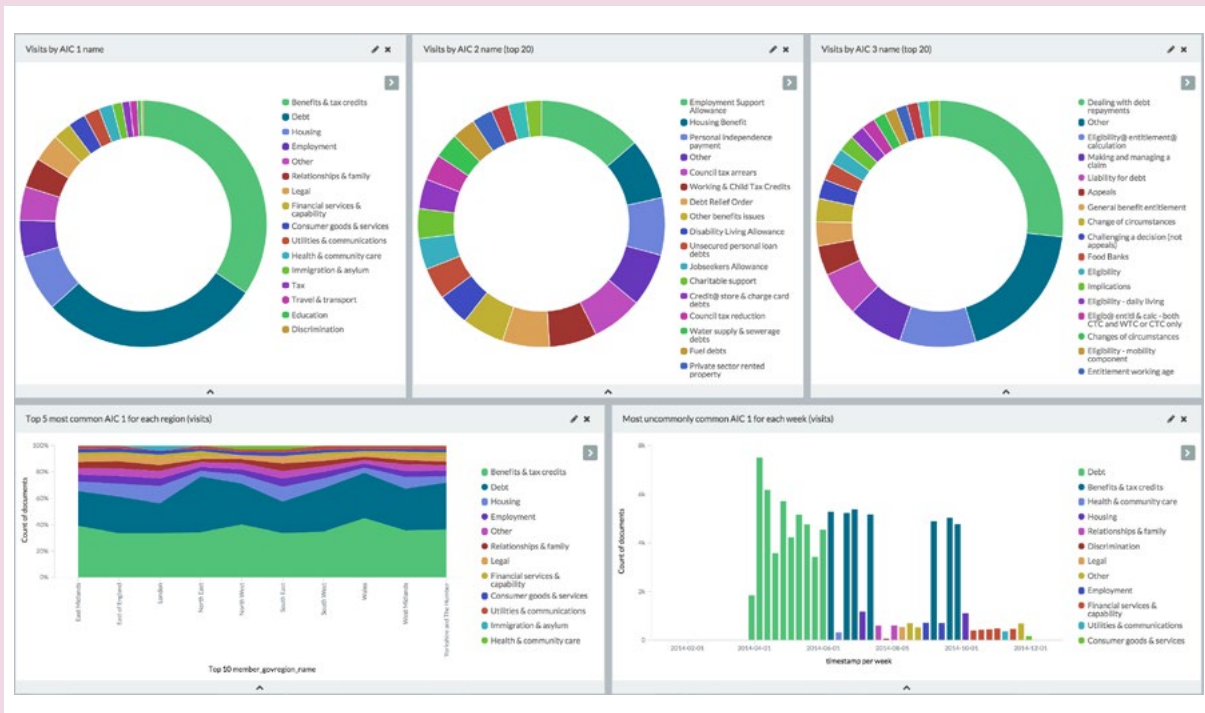
Exploring the EBEF's data

To understand what was being said in the EBEFs, the team examined all of the text word by word using natural language processing, finding what the significant one to three word topics were across the entire set. This resulted in a set of 300 highly significant topics that can be explored using any combination of time, frequency, location, demographic fields and issue codes. This means you can very quickly drill down to the information you are looking for.

In Figure 8 the map shows all the EBEFs related to unfair dismissal and the other charts have been filtered for employment issues. On the top right is a list of the 300 two to three word topics that most commonly appear in the employment EBEFs. We can see that the most common topics are housing benefit, food voucher, poor administration. Where we are likely to find the emerging issues is in the idle of the list of 300 topics. The most common ones are likely to be what Citizens Advice staff already know about. The least common rarely appear, but in the middle will be hints of problems that are just below the radar. The bottom graph shows the most common 25 topics by month that appear in the free text fields for EBEFs categorised as employment.

Nesta...

Figure 8.
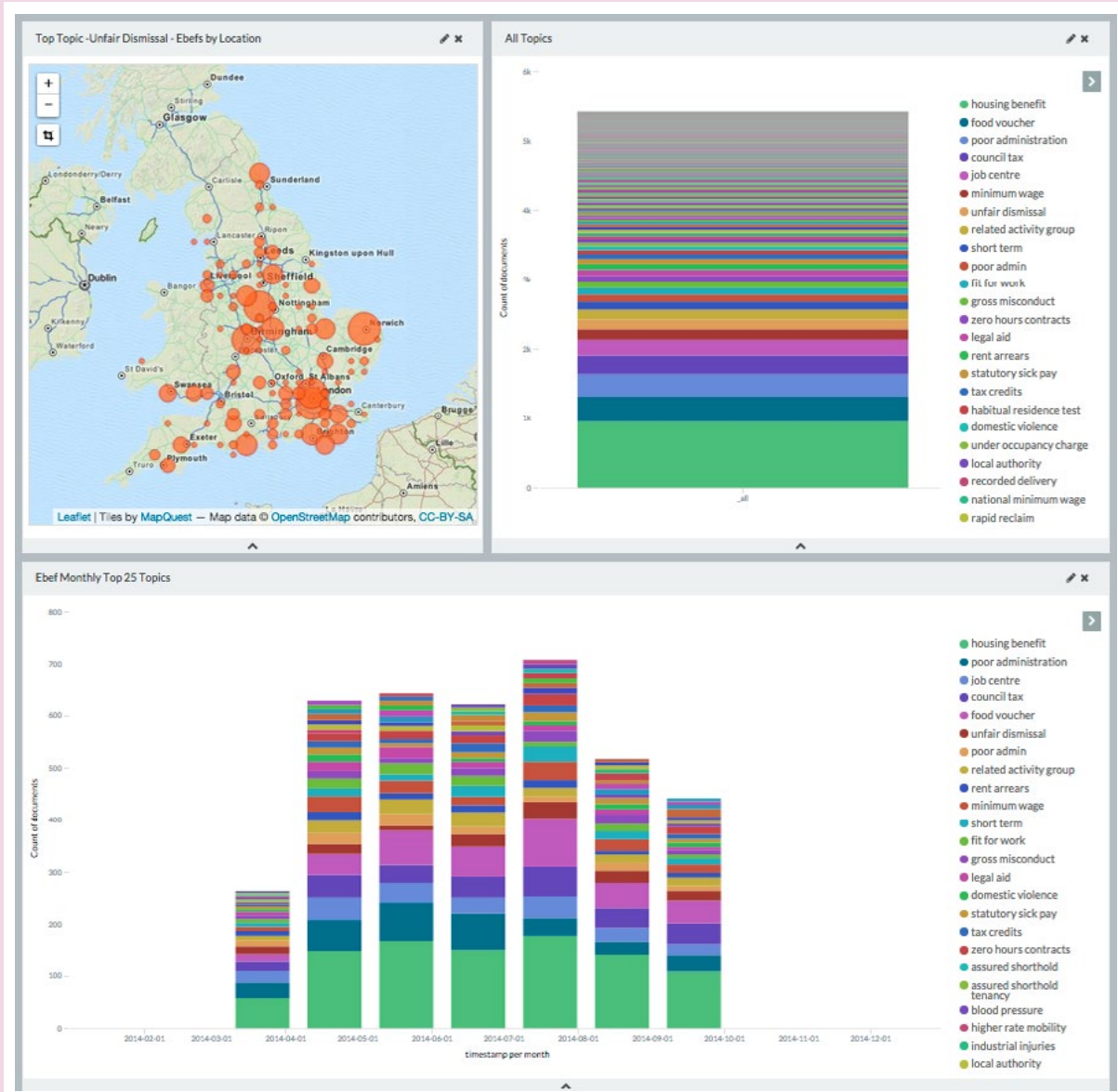
So what's changed?

When we started this project, Citizens Advice already took data seriously as a route to influencing change and improving performance. But this project was a catalyst for a completely different way of thinking about and approaching data as an organisation and as individual teams. Primarily, this happened through the team's collaboration with data scientists who offered new skills, experience and ideas that raised our expectations about what was possible with data. It gave the team energy, confidence and status in a way that was visible to other parts of the organisation, which in itself drove interest and engagement with the project. There are three particular shifts that have come as a consequence of this project:

1. **Embracing openness**
   Citizens Advice takes client confidentiality very seriously. It is one of the pillars of the service, and earns trust from clients. Whilst open to the idea of exploring the data in more depth, at first the teams were naturally and rightly sceptical to the idea of inviting external parties to dive into the data.

   However, though we remained clear that no action would breach our commitments to anonymity, confidentiality and privacy, the teams gradually became more comfortable with sharing data as long as the right protocols were in place. By the second DataDive, we worked hard to devise a safe and responsible way of sharing and linking data with St Mungo's Broadway to test if this would help us learn more about routes into and out of homelessness (and ultimately how to prevent people becoming homeless).[5]

2. **Democratising access to data**
   Perhaps the biggest impact of this project for the data analyst teams has been how it has shifted the strategy for managing information and data across Citizens Advice. They have moved from understanding their role as a team which provides regular analysis 'products' such as monthly stats breakdowns, maps, static dashboards and trends – as well as providing expert, ad–hoc guidance and consultancy to different teams – to providing the tools, platforms and methods that will allow others to interrogate data on their own terms.

   The dashboard from the DataCorps is core to this, as the visualisation tools and cloud–based access make it much easier for anyone in the organisation to access and analyse data in real–time. This is combined with the team offering coaching and running workshops about what it is possible to do with data, and even hosting their own internal DataDive inviting teams across the whole organisation to pitch ideas.

3. **A greater emphasis on questions and exploration**
   Ultimately, the dashboard makes Citizens Advice data more visible. This is prompting questions, and inviting more people to make connections between issues and areas; clusters on a map are much clearer than points in a data table. This greater visibility is encouraging more people to explore the data and ask questions, rather than being reliant on the data analysts team to come forward with their own analysis. Collectively, we are more alert to what is going on and better equipped to explore more.

And what was difficult?

Privacy

Throughout the project we were acutely aware of the sensitive and personal nature of the data we were working with. We were often dealing with data about vulnerable individuals – whether that was sick children in the Shooting Star Chase data or homeless people in St Mungo's Broadway's case – the work required both an awareness of data protection laws, as well as an emotional sensitivity to the stories in the data.

The risk for the charities involved was that someone could potentially identify an individual in a dataset and misuse that information. DataKind UK has established structures to support charities that take part in DataDives. During the DataDive weekends the charities worked with their Data Ambassadors to anonymise the datasets before opening them up to the room of data scientists. We asked the DataDive participants to sign non–disclosure protocols. We also asked that they not publish or share the data with anyone outside of the venue and to delete it once the event was over.

Nesta..

Citizens Advice was particularly worried about sharing the EBEF data as it had small amounts of profile information, along with free text fields which are harder to clean and fully anonymise. However, Citizens Advice mitigated the risk by sharing that data in a strictly regulated environment (participants could only access the data using an SSH tunnel) and decided that the benefits – being able to learn what was in the free text fields – outweighed the potential risks.

The St Mungo's Broadway project presented a particularly challenging data management problem. Linking up St Mungo's Broadway data with Citizens Advice data meant sharing their data in its raw form to match up individual records by name, data of birth and national insurance number. St Mungo's Broadway's information management team was concerned about the riskiness of exposing an individual. In order to minimise the risk they established a protocol for how to share the data, and ultimately they decided that the results would be worth it. To begin with, St Mungo's Broadway asked Citizens Advice to sign a data–sharing agreement. St Mungo's Broadway then shared the relevant data on a secured pen drive with one Citizens Advice staff member who matched up the data, reanonymised it and deleted St Mungo's Broadways original raw data. There is little guidance on how to link data in this way between organisations or what best practice in this area looks like.

Getting the most out of data without compromising confidentiality and privacy is tricky. It presents some genuinely difficult judgments as organisations weigh the risks and benefits, and is an area where society needs to catch up with changing technology.

Consent

As organisations collect individual level data they ask people to complete a data consent form. The form outlines the specific uses to which personal information will be put, one of which is usually analysis and service improvement. With St Mungo's Broadway the implicit assumption is that the analysis they conduct using individual level data is done internally. This was the first time that they had interpreted consent in a different way, and raises questions about the correct wording of consent for future projects.

Practical problems

Parts of the project were slow to move along, in part due to the heavy cogs of working in a large organisation. Asking two large organisations to share and link data, something neither had done before, proved to be a time consuming process and in some ways, perhaps it is surprising that it happened at all.

Other problems arose around sustaining momentum in the volunteer team to continue to work on the nine month project. The odd volunteer dropped out and others were recruited. All the volunteers had full–time jobs, and we would like to thank their employers, many of whom were flexible and allowed them to use work time to complete aspects of the project.

We would also like to mention the challenges we did not face. There was buy–in from all the partners, from the senior levels of all the charities, and importantly an incredibly dedicated and engaged information management team at Citizens Advice. Internally, Citizens Advice had also started using Elasticsearch so there were some happy coincidences in terms of tool consistency.

How did the DataDives help the other charities?

The two DataDive weekends were not only run for Citizens Advice. Three charities took part in the June event – Shooting Star Chase, Buttle UK and The Access Project – and two more in the October one – The North East Child Poverty Commission and St Mungo's Broadway. Below are the summaries of their findings and insights.

For all the charities, the DataDives were an eye–opening experience into how they could be making more of their data. Some of them took away interactive visualisations that they continue to use with stakeholders (Shooting Star Chase and The North East Child Poverty Commission), others used it to refine their processes and systems (Buttle UK and the Access Project). The most ambitious project involved linking St Mungo's Broadway data to Citizens Advice data on an individual level. This was pioneering for the sector in that it showed that it was possible to match up datasets across organisations.

Many of the data scientists who attended the DataDives and the Data Ambassadors continue to work with the charities. The DataDives were not ends in themselves, they were the start of many of these organisations' data journeys.

CASE STUDY: **SHOOTING STAR CHASE**

Shooting Star Chase, a leading children's hospice, was keen to discover where thousands of children in need of support from children's hospices around the country might actually be – a real concern as many hospices are sited for historical rather than needs–based reasons.

During the investigation, the Shooting Star Chase volunteers streamlined the referral paths of how children come to be at the hospices saving up to £90,000 for children's hospices around the country by refining the referral system. These savings are the equivalent of hiring three nurses – a significant discovery for such an underfunded sector.

A group of volunteers continued working with the charity after the DataDive to extend the work by building interactive visualisations. These enable Shooting Star Chase to better understand where the children with life–limiting conditions are and which health facilities they visit with which sicknesses. They now use this dashboard to persuade clinical commissioners of the high demand for children's end–of–life services and the need for more funding for the sector.[6]

CASE STUDY: **BUTTLE UK**

Buttle UK is a charity that provides grants to some of the most vulnerable children and young people in the UK. At the DataDive they used text analysis to unearth useful insights about the most common shared experiences among the families they support. They are now using the results from the DataDive to guide them in a new organisational strategy, to show which issues their clients face and how they are interconnected, and what kinds of longer–term support Buttle UK should provide.[7]

Nesta..

CASE STUDY: THE ACCESS PROJECT

During the DataDive, The Access Project, an organisation that provides tutors for young people – mapped the lifecycles of their volunteer tutors to better understand what makes for a successful pairing between tutor and student. They were able to predict when volunteers were likely to drop out of the programme and which factors did, and did not, affect their involvement.

However, the main takeaway was the realisation that they were not collecting the right data. The volunteer data scientists quickly refined the reporting process so that The Access Project could collect more useful data and get to the root of the problem, something they immediately implemented.[8]

CASE STUDY: THE NORTH EAST CHILD POVERTY COMMISSION

The Commission seeks to reduce child poverty in the north east of England. At the Autumn DataDive they wanted to explore how they could gain a more real–time understanding of the issue and communicate the data in a more actionable way to encourage immediate responses. The team of volunteers first looked at whether the Citizens Advice data could be used as an indicator of child poverty. However, they found that it was not statistically significant. The team had a lot more success with creating an interactive visualisation to help the public understand the data around child poverty. The volunteers had an epiphany when they discovered that a DataKind project in the US was remarkably similar. They repurposed code from the DC Action for Kids project[9] and, within a matter of hours, had produced an interactive map,[10] which lets the viewer choose any one of the loaded datasets and see how it plays out across the map.[11]

CASE STUDY: ST MUNGO'S BROADWAY

St Mungo's Broadway wanted to look at what kinds of issues people went to Citizens Advice for help with, before they ended up homeless in St Mungo's Broadway's system. The team of 30 volunteers who worked on the St Mungo's Broadway project said that St Mungo's Broadway's clients were most likely to have approached Citizens Advice for help with employment support allowance. They also began creating a predictive model to establish how likely it is that a client will return to St Mungo's Broadway for help and if so which type of service they are likely to require.

Looking ahead

The Civic Dashboard is up and running as a responsive, real–time tool to understand the social issues people are facing in Britain. Citizens Advice teams are using this as a way to spot trends, interrogate questions and identify emerging challenges.

Our data analyst teams are hosting workshops to walk people through the Dashboard, and using this as a springboard for coming up with new ideas and opportunities for data–driven innovation. This is only augmented by the incredible and inspiring generosity of the DataKind UK volunteers who now act as an expert advisory group and informal mentor network for the teams and individuals internally.

But the work doesn't end there. Even since the Dashboard was first shown, we have explored the possibility of new projects such as: linking Citizens Advice data with healthcare data to more fully understand the relationship between health problems and social problems, and therefore design more holistic interventions; supporting more civic action through making data open as to what's happening where (e.g. plotting instances of scams, mis–selling or poor advertising); allowing for more responsive monitoring and evaluation, tracking demand and moving resources to where they are most in need; combining census data or other data with Citizens Advice data, or using Citizens Advice data as a benchmark to track impact of interventions in other areas.

Through exploring our data, we have laid the foundation for much more innovation in using data to address social needs.

Conclusion

This chapter shares some insights and lessons from Citizens Advice, DataKind UK, and the other charities that are using data science methods to understand more about social issues. And we are not alone. There has been a real surge in interest in recent years, in the potential of open data and data science to help charities make the most of their data, and open up their data to others.

Much like any innovation project, we have learnt that making this work requires hard graft, and lots of time and commitment from dedicated teams. Whilst technology has introduced the possibility of new methods and faster processing, ultimately the 'magic' of data science comes down to people and organisations with the determination to make things work and improve outcomes for the people we are here to support.

Nesta...

3. Using data-driven methods to understand hidden social action

Understanding the dynamics of social action is one of the key components of Nesta's work on supporting public and social innovation. In reports such as the *Compendium for the Civic Economy*[12] and *People Helping People*[13] and practical programmes such the Centre for Social Action Innovation Fund[14] and Innovation in Giving[15] we have sought to understand where social action happens, how it happens, who the people and organisations supporting and enabling social action are and what can support more social action to happen at scale. This combined with research from organisations such as the Charity Commission,[16] the Third Sector Research Centre (TSRC)[17] and the NCVO Civil Society Almanac,[18] mean that we already have a very rich understanding of social action and civil society activity in the UK.
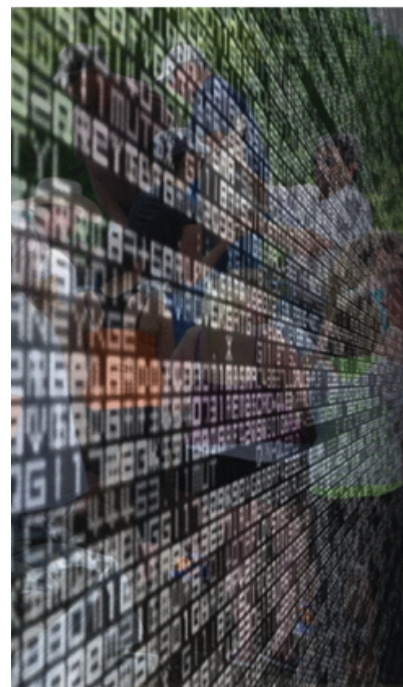
Funding social action through crowdfunding
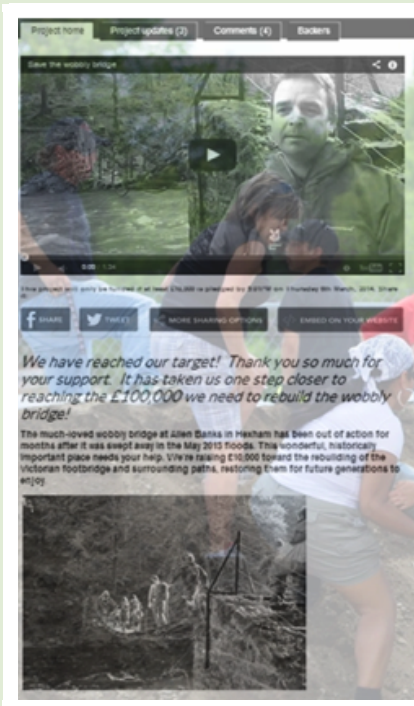
Mobilising social action through social media

Discovering hidden social action through open data

However, most of our understanding to date is based on studies and registers of established civil society organisations such as registered charities or social enterprises.

From research by the TSRC and others we also know that looking at established social economy organisations does not give us a full picture, as there is an enormous amount of social action happening 'below the radar' (BTR) in informal groups coming together to identify and address a social need. This type of social action has traditionally been, and still is, very hard and resource intensive to research.[19] However, as more of what could be considered BTR activity moves online and we develop more and better methods for capturing, opening up and analysing data we are interested in understanding, the potential in developing data–driven methods to better understand this phenomenon. Examples of this type of activity include citizens using crowdfunding platforms to finance social action projects, or using social media such as Twitter to address social needs.

This led us to fund five research grants exploring different methods for answering some of these questions:

- **Demos' Centre for the Analysis of Social Media (CASM)** – looks at the role of Twitter in mobilising social action, examining two case studies, the Somerset floods and the Step up to Serve campaign.

- **The Royal Society for the Encouragement of Arts, Manufactures and Commerce (RSA)** – Examines how traditional methods of mapping used in RSA Connected Communities research in Hounslow compare to data driven methods when mapping local below the radar activity.

- **Cardiff University** – Investigates how we can begin to analyse and understand new types of community mobilisation and social action through social media analysis, with a specific focus on the events surrounding the murder of Lee Rigby.

- **European Alternatives** – Study projects on civic crowdfunding platforms, to explore if this can help identify new types of collaboration and funding for projects not connected to established organisations.

- **The National Council of Voluntary Organisations (NCVO)** – Considers if mining open data from grant funders such as the Big Lottery Fund and Wellcome Trust can identify grantee organisations that do not appear on any existing registers of civil society organisations.

The following articles summarise the different projects, their methodologies and what they found. The full length papers with more in–depth descriptions of the methods and datasets can be downloaded from www.nesta.org.uk/publications.

Finally, it is important to note that this research is focusing on a very small subset of BTR activity, looking at people and activities that are mobilised online. This naturally excludes interesting offline activity and people not online. We acknowledge this and hope these methods can be used to complement other types of research looking at offline activity.

Nesta...

3.1. Mining the grant makers

David Kane, National Council for Voluntary Organisations (NCVO)

What did we set out to do?

This project set out to develop and test a methodology for identifying below the radar organisations within data held by grant makers (both statutory and private) on who they fund. By matching data about who these grant making organisations fund with data on registered organisations, we hoped to identify the remainder as below the radar. At NCVO, we believed this approach would enable us to pick up organisations outside the sphere of known, registered organisations. However, we realised from the start that the method would only pick up particular types of organisations – namely, those that have an interest or an ability to seek out grant funding. In this sense, we thought this method would lower the radar, rather than bypass it entirely.

What do we mean by below the radar?

In the context of the project we defined below the radar organisations as associations of people with a charitable or social aim, but which don't have a formal registration as a charity, company or other legal form. This definition focuses on organisations that may not be formally constituted and are usually unregistered due to their small size (as measured by income, expenditure, assets or employees). However, it does not capture some below the radar activity, particularly new forms of activity such as social media.

How did we do the research?

We identified below the radar organisations in three stages. First, we gathered the source data from grant makers (either directly or through the 360 Giving initiative).[20] Secondly, the data was matched with official registers of organisations based on the name of organisations – a process known as 'reconciliation'. Groups of registered organisations such as universities, local authorities and schools were also found using keyword searches. Lastly, we tested the process and produced our results. Our methodology was set up to be flexible and iterative, with improvements and refinements identified as the project progressed.

How many below the radar organisations did we find?

After applying the matching process we identified over 111,000 grants out of our population of 240,000 grants that were not matched to a registered organisation, or identified as a registered organisation or other above the radar organisation (such as a local authority or school) through keyword searches.

After deduplication, we were left with 125,000 organisations in the grants database. Of these organisations, 63,000 had received one or more grants that had been previously identified as registered organisations, indicating that 49 per cent of organisations in our dataset were unregistered (around 62,000 organisations).

A further check of the results was done by looking at the amount awarded. As the threshold for charitable registration in England and Wales is an income of over £5,000, this can be used as a proxy for the size of the organisation in the dataset. Of the 62,000 unmatched organisations, 33,000 received no grant greater than £5,000. This would indicate that just over half of the unmatched organisations could be classed as below the radar.

Figure 9. **Matching results flowchart**

Initial population: **240,012** grants (**125,149** organisations)

Filter by matching with registered organisations or identifying as registered organisation

Reconciliation

Registered charity → **55,510** grants

Keyword searches:

Companies → **24,475** grants

Public sector → **19,787** grants

Universities → **13,903** grants

Registered charities → (not matched) **10,839** grants

Person → **4,444** grants

Reminder: **111,054** grants **61,885** organisations

Matched or identified as registered: 128,958 grants (63,264 organisations)

Filter by size of largest grant

Grant amount over £5,000 → **28,668** organisations

Reminder: **33,217** organisations → Most likely to be 'below the radar'

Nesta.

What did we find out about these organisations?

Activity

The largest category of below the radar organisations is culture and recreation (including arts and sport organisations) with 30 per cent of organisations, followed by social services (24 per cent), which encompasses a wide range of social activities, including children's clubs and support for the elderly. When comparing the distribution of organisations in our below the radar population with those for registered charities, we found that culture and recreation organisations are not only the largest group, but also disproportionately represented in below the radar organisations.

Figure 10. **Distribution of organisations by ICNPO category**
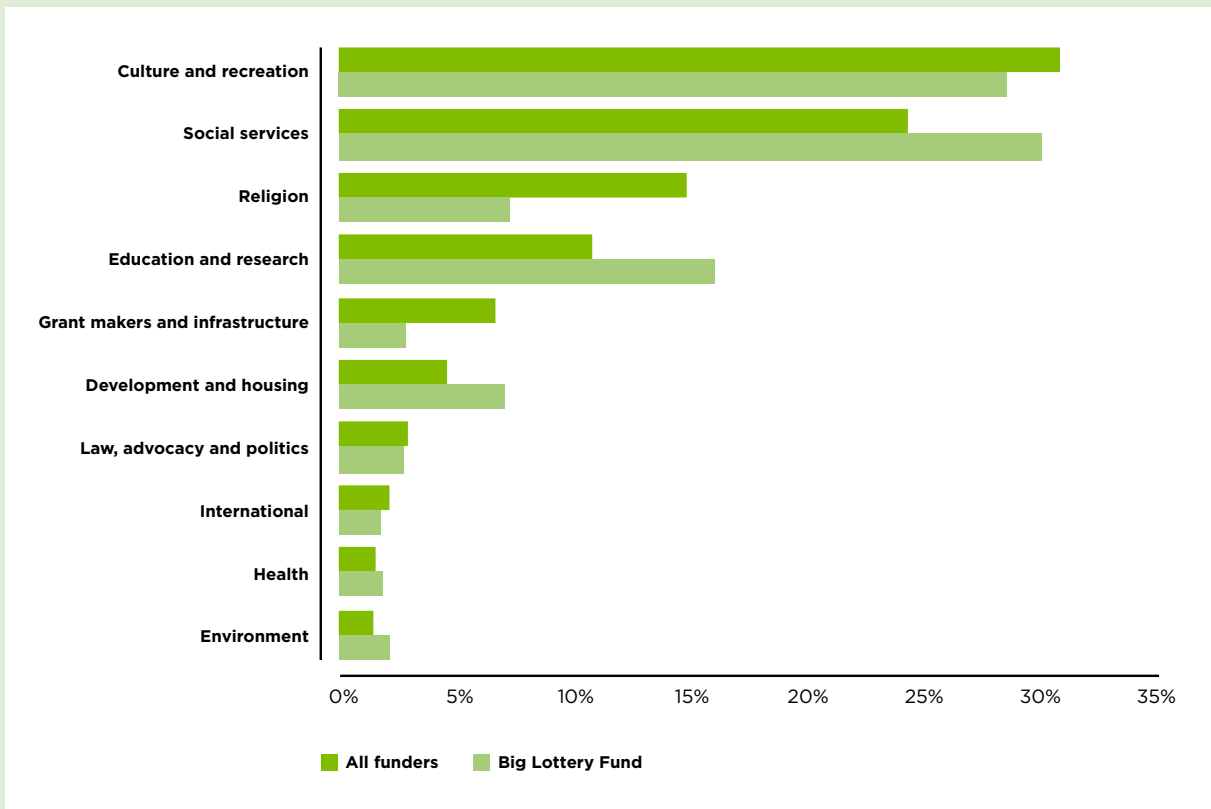
Geography

The distribution of below the radar organisations by country largely reflects the distribution of registered charities, with the exception of Wales, which has 12 per cent of below the radar organisations but only 5 per cent of registered charities.

Distribution of organisations by country

Country	Below the radar	Registered charity
England	71%	75%
Northern Ireland	3%	4%
Scotland	14%	15%
Wales	12%	5%
UK–wide	0.1%	1.5%
International	0.0%	0.1%

Looking just at England, we found a more even distribution across the country than for registered charities. Organisations identified as below the radar do not appear to have the same 'headquarters effect' that registered organisations do, whereby looking at the registered address of an organisation shows over–representation of London and the South East, where many larger organisations have their headquarters. The North West, North East and East Midlands appear to have a greater proportion of BTR organisations when compared to registered charities. This may provide some support for the idea that below the radar organisations are more active in areas where charities are less common.

Figure 11. **Distribution of English organisations by region**

Nesta..
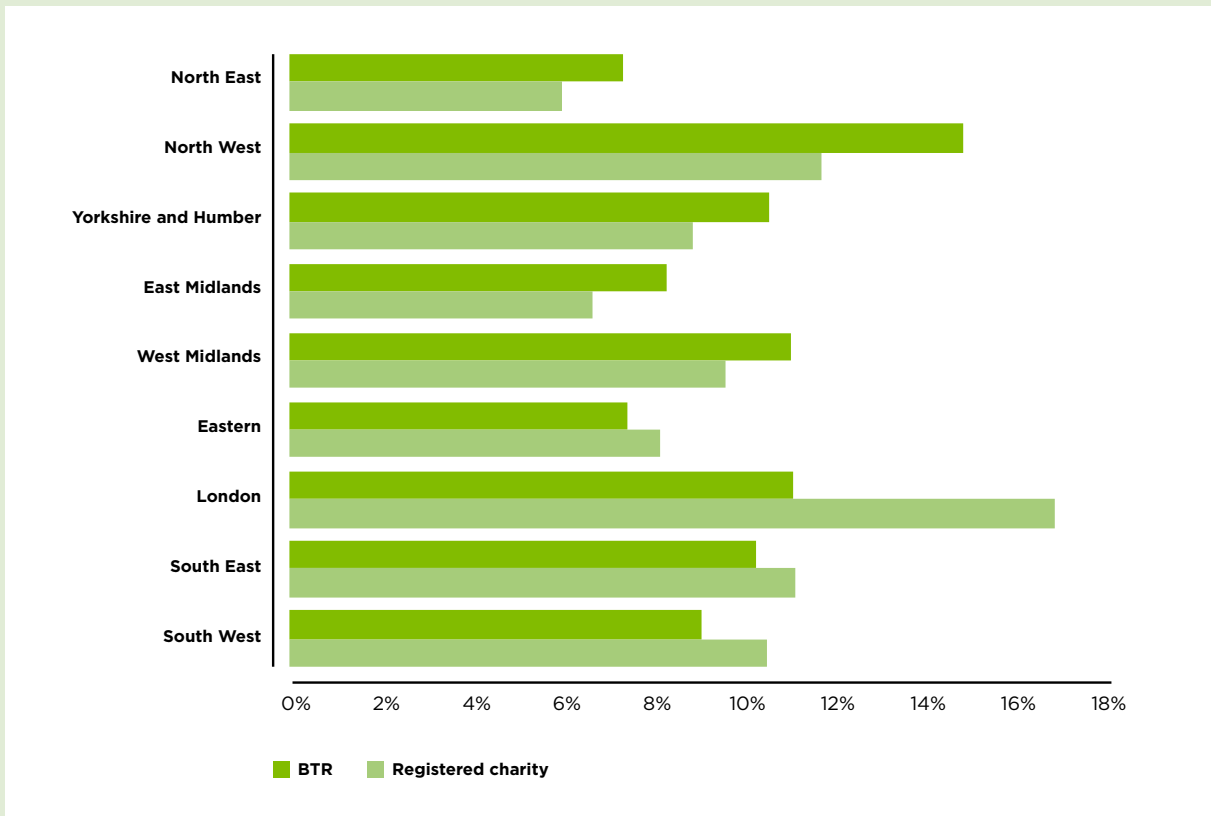
Organisational size

The data used did not have a field that specified the size of the organisation. However, we were able to use the size of the grant that organisations received as a guide to the size of the organisation. In general, the organisations that were identified as registered received larger grants than those that could not be matched. However, there were a number of large grants (over £10,000) made to organisations that were not identified as a registered organisation – indicating that the matching process has likely missed these organisations (as organisations receiving a grant of this size are likely to be registered).

Figure 12. **Distribution of English organisations by largest award size**
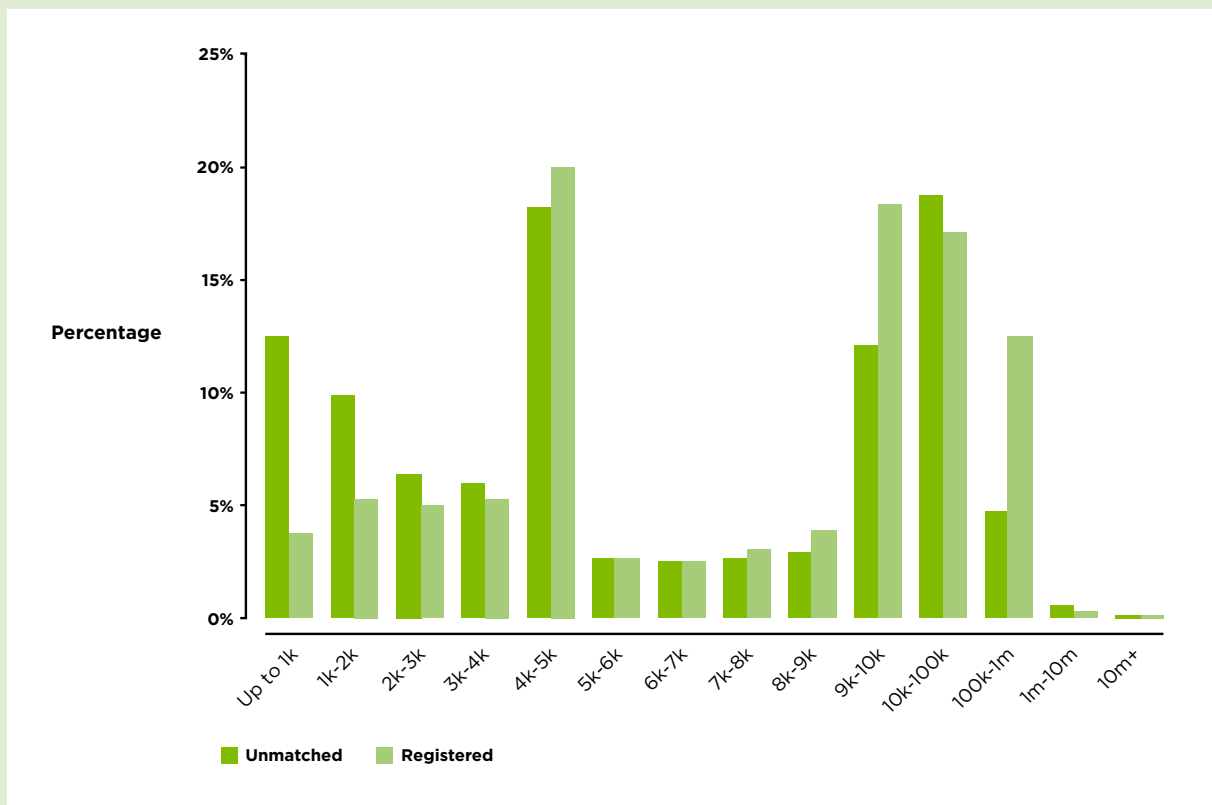
So was our approach successful overall?

The methodology we used had mixed results:

The process of matching organisations that have received funding to official registers of organisations (notably the Charity Commission register) appears to have gone well. But the picture is complicated by the difficulties in finding other registered organisations in the dataset. In particular, there is no comprehensive register of public sector organisations which includes, for example, parish councils, schools and other smaller organisations. Even looking for discrete groups of organisations such as universities or local authorities through keyword searches is made difficult by the way these organisations are described.

However, it was possible to identify organisations that could be considered as more likely to be below the radar. To that extent we have been successful in lowering the radar. These organisations that have been identified as likely to be below the radar do exhibit characteristics that might be expected from smaller unregistered organisations:

- They are likely to be smaller in size (as measured by the size of the grant they received) on average than those identified as registered organisations.

- They are also more likely to be arts and sports based, often running small arts projects or sport clubs.

- They show a much more even spread throughout the country.

What next?

This project has highlighted a number of possibilities for further work and development:

- A comprehensive register of organisations and reconciliation services needs to be available. Resources such as Opencorporates[21] and Opencharities[22] provide examples of what is needed. These will help researchers and others repeat similar exercises using large lists of organisations.

- There is a particular gap around public sector organisations. There is no official list that contains every public organisation, and no unique identifier or URI scheme for organisations. This is vital when looking at a data resource that covers the boundaries between sectors as this one does.

- There are other gaps in available data on civil society organisations. Scottish and Northern Irish charities are not currently included, although Scottish charities are now available as open data from OSCR[23] and the Northern Irish charities register has only recently been set up. Other types of bodies could also be incorporated such as royal charter bodies and Community Amateur Sports Clubs (CASCs).

- The 360 Giving programme is providing useful data which is of value and offers insights that cannot be found elsewhere. We would encourage the continuation of that programme, and greater use of the data and improvements to the quality of data imported will make it a more useful resource. Of particular use would be the inclusion of charity numbers and company numbers which would remove the need for an imperfect matching process.

Nesta

3.2. Listening in: using social monitoring tools to understand the social economy

Rowan Conway, the RSA, and Jimmy Tidey, the RCA, with thanks to Gaia Marcus

Going where the (social) action is

Whether it is local mums coming together to support each other or a group of neighbours joining forces on a guerilla gardening campaign, there is 'below the radar' social action happening everywhere. As well as the official services delivered by charities and social enterprises, it is the informal or semi-formal activities led by community groups or individuals that provide the 'renewable energy' which powers the social economy

The transformative potential of below the radar social action is enormous, but it is also hard to track. Understanding grass roots community activity outside of the established social economy (such as registered charities or official databases) has the potential to redefine community engagement methods; transform social investment models; or even challenge established local governance systems by widening the pool of talent to include a range of 'unusual suspects'.

The Connected Communities research team at the RSA has been exploring how social network analysis can be used to combat isolation at a community level since 2010. In a seven-site Big Lottery–funded study called 'Building Inclusive Communities' we have been using traditional social research methods to gain a rich picture of the localities we work in across the UK. Through social network analysis and asset mapping, our data has provided a snapshot of each community at a hyperlocal level, providing a platform for co–production and inspiring social action in each locality.

The programme uses asset–based community research methods and surfaces below the radar activities with tools such as surveys, interviews and focus groups. Local community volunteers might survey a local area through door knocking to capture data about its resources, skills and talents. Or community groups might come together in a participatory process to compile an asset map and use the map to build collective plans for an area. Social network surveys provide another lens through which we view community assets, building a picture of a community's relationships as a social graph. In using these research methods, the RSA has observed that the process of research itself can be empowering, as the journey of discovery of local assets sometimes inspires social action in itself, as well as building relationships and thickening local social networks.

While these research methods offer a multi-layered picture of a community – way beyond the official stakeholder databases or registries of voluntary sector bodies – getting this data is not simple. The process of recruiting and training community researchers, managing local volunteers, and processing the data can be very resource intensive. The data is also time-sensitive as it shows a community's assets or an individual's social network at a particular moment in time. These limitations are largely outweighed by the benefits, but the RSA was interested in exploring alternative methods to see if they might yield similar results in a less resource intensive and more dynamic way.

The Community Mirror

As more and more community activity moves online into hyperlocal blogs and onto social media sites, the opportunity to apply digital research tools for local community data gathering is becoming a reality. The RSA wanted to test digital research methods to map community assets to see if they could deliver similar results to the traditional face-to-face methods that we deploy in the Connected Communities work. With Nesta funding, we took on a micro-research project to see if social media analytics might provide an alternative way to view a community's assets and activities.

This collaborative research project, which we called Community Mirror, was carried out to understand the ways in which online data-gathering techniques compared to door-to-door research. We worked with the Royal College of Art (RCA) using their LocalNets. org application to gather the online data and contrasted the findings with one of the RSA's existing community research projects in the London Borough of Hounslow that used face-to-face methods.

Mining the Twittersphere

The LocalNets.org app mines the social web by aggregating Tweets and blog posts from selected Twitter accounts and blogs. This activity is in effect a 'social listening' platform as it provides the capability to systematically collect online conversations about specific phrases or words. Each item is coded with topic tags, geolocation where applicable, and any entities that are relevant to it. Words that are captured include people, places, organisations and events – the same type of data that we seek through face-to-face surveys.

As it is a microblogging platform, Twitter is a rich source of social data and is a good starting point for social web mining because of its inherent openness for public consumption (Facebook is a more closed network). Twitter data is particularly interesting because it is so dynamic. The LocalNets.org web app was live over a four month period in 2014 and collected data about the online activity in Hounslow in order to produce a visualisation to show the following:

- A list of the community's assets, including people, places, organisations, and events.

- A geographic map of places that the locality's community members are talking about online.

- A network diagram of how the community assets are connected to one another.

The resulting online community map looked like this:

Figure 13. **Hounslow social network map of the community assets identified by the LocalNets.org web app**
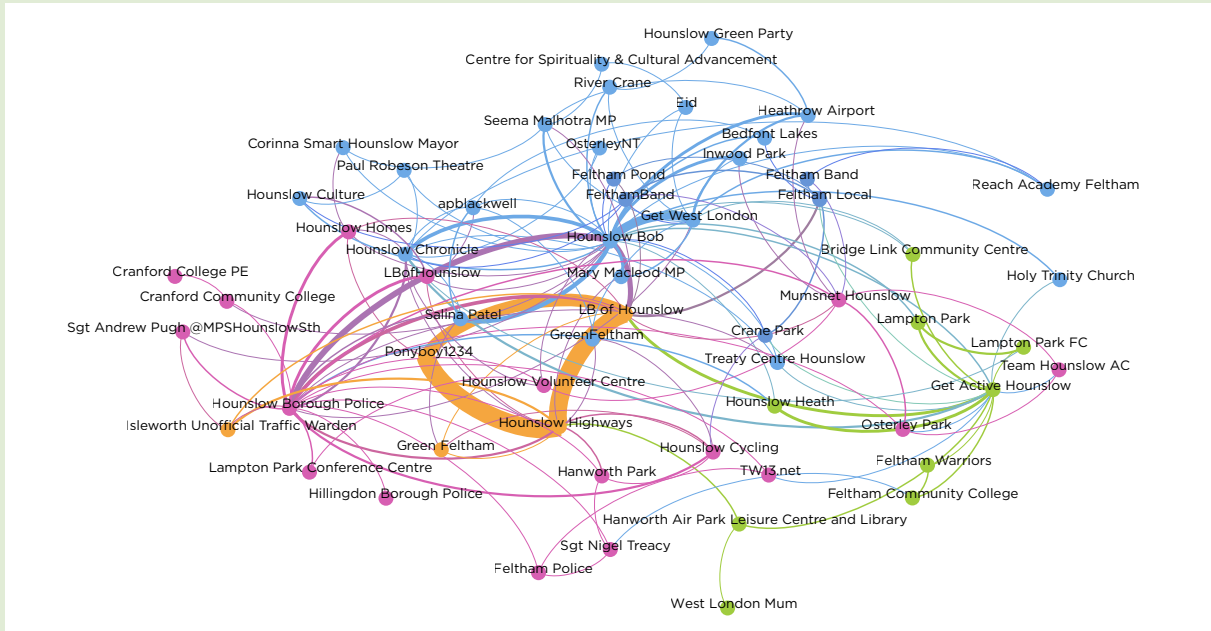
The offline research (door–to–door surveying) that happened concurrently in the Cranford ward of Hounslow, captured the people or resources that local community members look to for social support. Respondents were asked to identify people, locations, organisations and events that help them feel part of the community. These community assets and the connections between them and the respondents were then visualised on the social network map below:

Figure 14. **Social network map of the community assets mentioned by survey respondents in the offline door-to-door research**
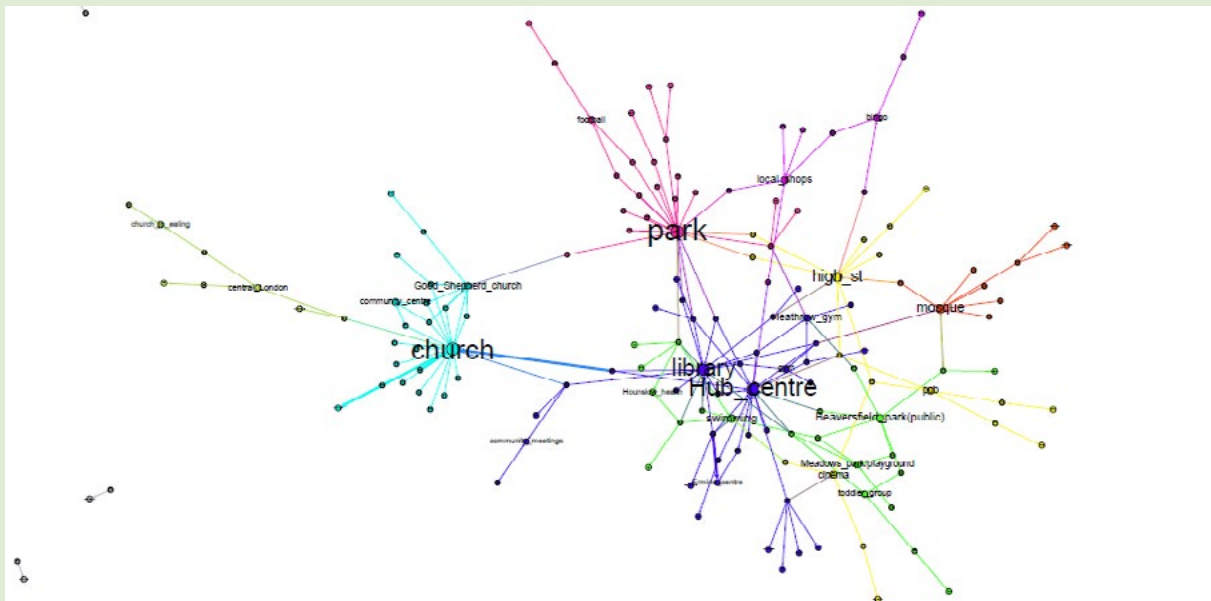
Note: Different clusters of connected assets are shown in different colours for clarity.

Findings

In the four month period of data collection, the online dataset generated a list of 294 community assets, categorised as people, locations, organisations and events. These assets were from across Hounslow, not specifically in the Cranford ward. It showed – perhaps unsurprisingly – that when they were online, Hounslow residents often identified community assets outside of their own locality. When it came to events and people, there were no direct correlations between the offline and digital data. The online asset mapping discovered 25 community events and 72 people who might be seen as influencers – including MPs and local footballers.

It was with the local organisations and places where correlations could be found. The digital asset mapping found 168 locations and organisations and the offline asset mapping found 51 locations and organisations. Twenty per cent of the assets mapped offline were picked up by the digital tool, with ten clear matches between the two datasets. With a further six probable matches, it is reasonable to say that 31 per cent of the place and location assets mapped offline were discovered by LocalNets. This suggests that the LocalNets digital tool has validity in surfacing relevant local data, while at the same time supplementing this with a large volume of additional assets not mentioned in the offline research.

Whether these items can accurately be described as constituting the below the radar social economy, however, is open to question. Most of the assets included in both datasets – hospitals, parks, churches – are well known to authorities and are key parts of the formal social economy rather than below the radar. What the LocalNets data can provide, however, is a map of how these institutions engage with each other online – and it is this relational aspect of the social economy that is often below the radar, and one of the key benefits of LocalNets. The tool can produce up to date and locally valid assets maps, with some insight into how these institutions interact with each other – or not – online.

Overall, the study found that LocalNets.org software is an efficient way of collating hyper–local information about community activities and organisations in a given area. With further user interface development, it is possible that is could become a reasonably inexpensive way of gaining insight into the below the radar activity in a local area as part of a social media analytics process.

Whose radar is it anyway?

While the study did prove that the LocalNets.org app can surface additional data about the people and organisations that are supporting and enabling social action, what one does with this intelligence is another matter. Digital methods like social listening can provide a relatively inexpensive way to bring a new, more dynamic, dimension to a community asset map, but the tricky questions arise when you ask who exactly will use these methods? Will it be the local authority, local community groups themselves or even central government? Whose radar is it that we are going under?

Mapping things that are below the radar sounds inherently covert, like somehow we are smoking out a shadow economy. Certainly, there are ethics and privacy questions that need to be explored when social media monitoring is applied to communities. Lessons can be learned from the corporate experience where social intelligence tools such as Brandwatch are routinely used by corporations to track social media and online conversations about their brands in order to understand consumer behaviour. Companies have vast datasets about consumers, but many still struggle with how to engage with customers about the data. The process of interacting with users of social media for organisational purposes like customer

Nesta...

service or sales brings up awkward questions such as: do Twitter users know you are listening? Will they find it acceptable? Does regular social listening turn you into an online Big Brother? These questions, when applied to a local community context, will require thorough consideration.

In our conclusions, we find that digital methods do offer a promising approach to mapping the below the radar social economy assets, but that traditional methods are key to engagement with communities. It is human connections that take an asset mapping process beyond a simple audit to a platform for social action. While the on the ground research certainly requires significantly more resource than digital tools, our study did not suggest that the digital tools could effectively replace offline methods – rather they were complementary. Combined, they can provide a powerful source of insight. Indeed, they can feed each other, and digital asset mapping can be 'seeded' by existing local knowledge and used to complement offline community asset mapping approaches. LocalNets.org can be primed with keywords that come from other asset mapping processes, and the output of LocalNets.org can be used as a starting point for deeper investigation.

So digital tools that go beneath the radar can indeed produce useful and useable insights about community. This study certainly informs the Connected Communities work going forward, and you never know, it might even evolve into a new way of tackling that old chestnut of finding the 'hard to reach'. But that is another conversation entirely...

3.3. Soft facts and spontaneous community mobilisation: the role of rumour after major crime events

Colin Roberts, **Martin Innes**, **Alun Preece** and **Irena Spasic** Cardiff University

This study examines how social media is transforming processes of community mobilisation in the aftermath of major crime events. In so doing, it illuminates processes that are 'below the radar', in terms of being frequently 'seen but unnoticed', owing to how public attention fixates upon the response of the police and other security agencies. The analysis focuses upon two specific issues: (1) how rumours disseminated via social media platforms work as 'soft facts' to influence patterns of collective reaction in such circumstances; and, (2) how these communication platforms are being used to organise particular forms of spontaneous community mobilisation.

The analysis pivots around four case studies of community mobilisation following the murder of soldier Lee Rigby in Woolwich, London in 2013. It distils some more generalisable ideas about how social media is transforming the ways in which contemporary collective action and social reactions are organised. The key insights derived from the analysis are:

- Social media communications function as both an 'engine' and a 'camera' – they propel social reactions, but simultaneously leave 'digital traces' that can be used to develop a picture of these reactions.

- Rather than 'big data', the key quality of social media following major crimes and disasters is it's functioning as 'fast data', with information about what is happening travelling rapidly out from the scene.

- The 'velocity' of data influences both public sense–making and processes of collective action, and comes close to outstripping the capacity of existing social networks to respond. Consequently, rumours play an important role in shaping public sentiments on social media following major incidents.

- Concurrently, forms of community mobilisation spring up, often reflecting polarised political and ideological interpretations of what has happened and what should be done about it.

- The evidence suggests that effective spontaneous community mobilisation frequently has to be scaffolded by existing social networks and institutional structures, which are repurposed in moments of perceived crisis.

A key idea developed through the research is that of 'spontaneous community mobilisation'. This is intended to capture how, in the volatile wake of a major event, individuals, communities and groups come together to engage in forms of urgent and improvised collective action and sense–making. This is spontaneous in that it was not (and often could not be) planned, and social media, because of the networked communication it represents, plays an important role in enabling these mobilisations to occur.

Nesta...

Research design and data

The data analysed are drawn from a larger dataset of social media materials collected in the wake of the killing of Lee Rigby. In excess of 34 million data points were collected using a text–mining platform developed by the authors and their research team. These stretched from the first Tweet from a witness at the crime scene, through to the conclusion of the court case where the two suspects were found guilty. The scale of the data is important, because of the 'high resolution' picture it affords of processes of social reaction. It enables us to track and trace in fine–grained detail what happens following a major crime, in ways that have not been possible previously.

For this study, samples from this larger dataset were extracted for more detailed analysis, on the grounds that a preliminary look had identified the presence of potential rumours and community mobilisation. Framed in this way, the research engaged with three principal aims:

1. To use data–driven analysis to develop an understanding of how rumours about crime risks and threats encourage forms of community mobilisation that are not dependent upon formal organisations.

2. To bring forward new concepts that account for the role of social media data in seeding new forms of social organisation.

3. To consider the policy and practice implications of these modes of spontaneous community mobilisation for agencies involved in community impact management work following the occurrence of major crime events.

The data were collected and analysed using a bespoke suite of software tools called Sentinel.[24] Sentinel can collect social media data from a variety of sources, however this study focuses on material collected from Twitter. Analyses of these data were guided by application of a 5W (who, what, when, where, why) framework:

• Who: Identify participants in events, and relationships between them.

• What: Characterise events including crimes and communities mobilisation.

• When: Maintain a timeline of linked events.

• Where: Determine locations by both geo–tags and place names.

• Why: Apply social science models to uncover causal links between events.

Figure 15 below provides a conceptual representation of the platform's architecture and Figure 16 shows how data are displayed by the system:

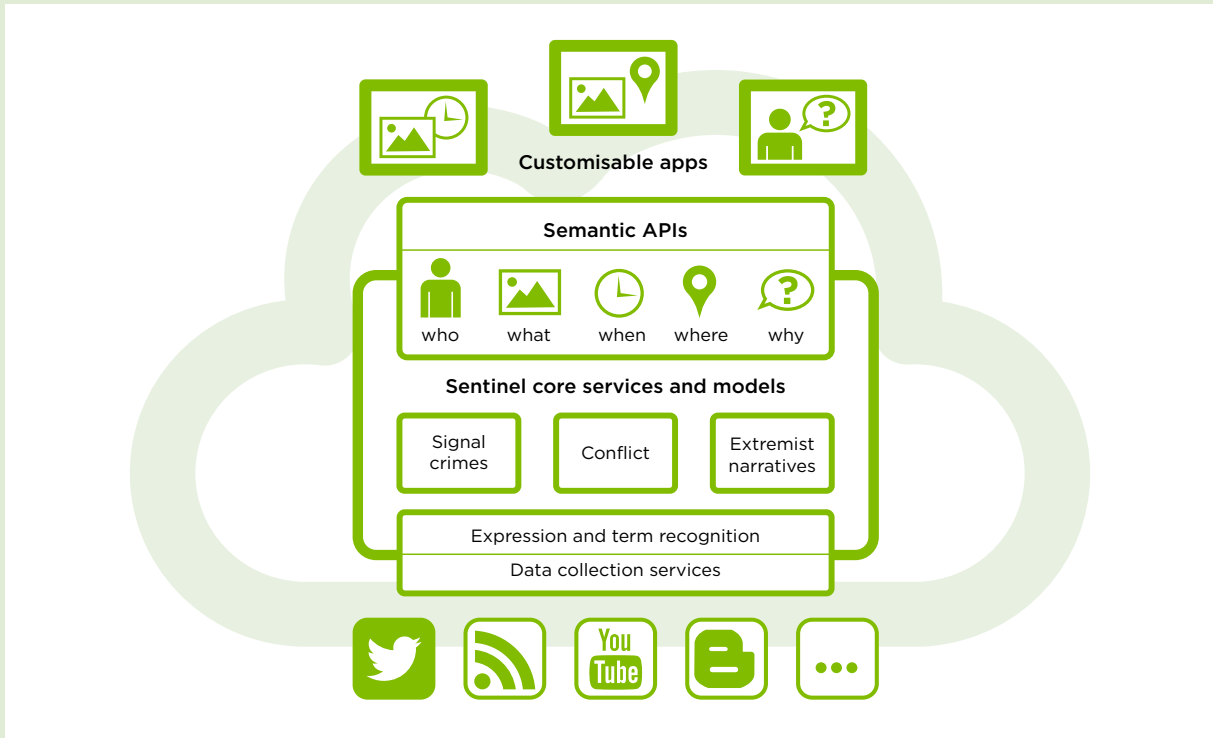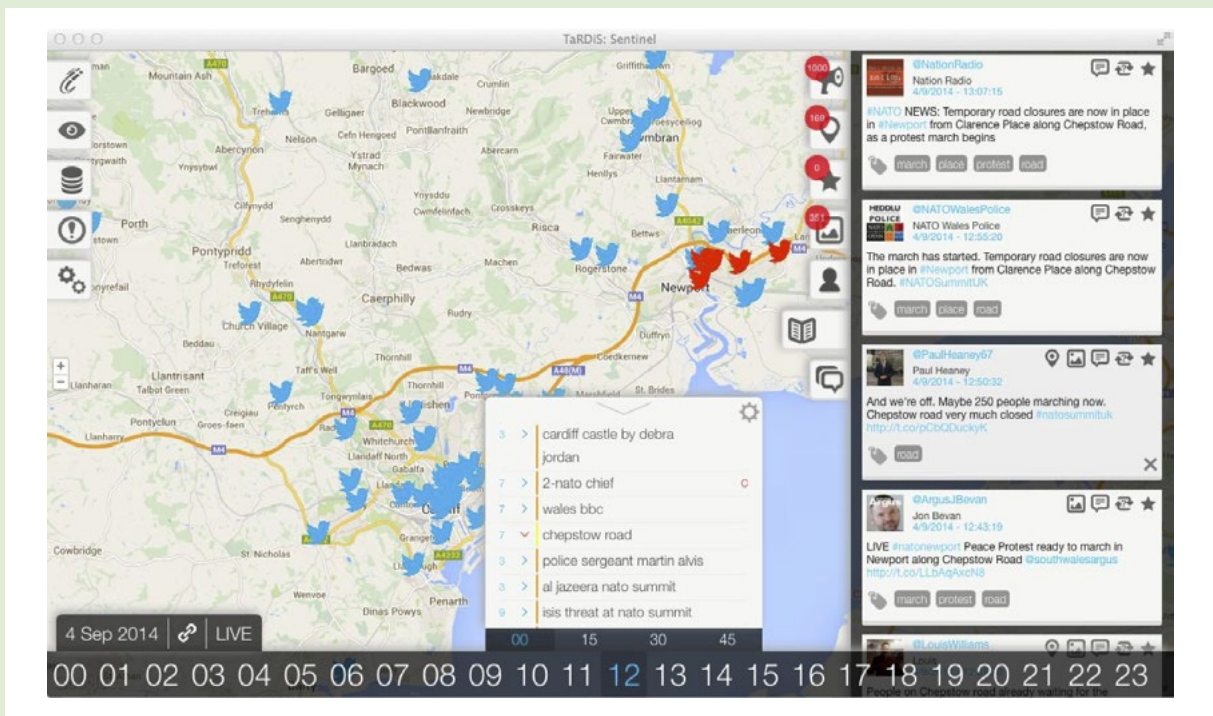## Figure 15. An Overview of the Sentinel Platform

Figure 16. Example of a Sentinel What/When Data Visualisation


Nesta...

Key findings from the four case studies

The first case study focused on processes of social reaction in the first 24 hours after the killing of Lee Rigby, and how right–wing political groups rapidly tried to mobilise their supporters for public protests using social media. In response to which, a coalition of groups opposed to their political views also sought to rapidly mobilise a counter–demonstration. Again, much of the work of contacting and trying to persuade people to get 'feet on the streets' was conducted using social media.

The key insight developed into how mobilisation happens is the interaction between the two sides. Each was adapting and responding to messages broadcast by members of the groups to whom they were opposed. These messages were 'soft facts' in that it was difficult to corroborate them and establish their validity given the fast moving nature of the situation. But they were sufficient to act as indicators of the presence of a perceived 'threat' needing to be countered. This seems to be a key ingredient in terms of persuading supporters of the need to physically mobilise.

Analysis of the Twitter data showed that it peaked on the first evening with nearly 900 Tweets per minute relating to the Rigby case, as people flocked online to try and find out more about what was happening. More fine–grained analysis though revealed that embedded within these large volumes of data, were defined 'thought communities' actively using these platforms to try and rapidly mobilise their political support.

Figure 17. **Tweet Volumes Per Minute for Lee Rigby Murder During First 24 Hours**
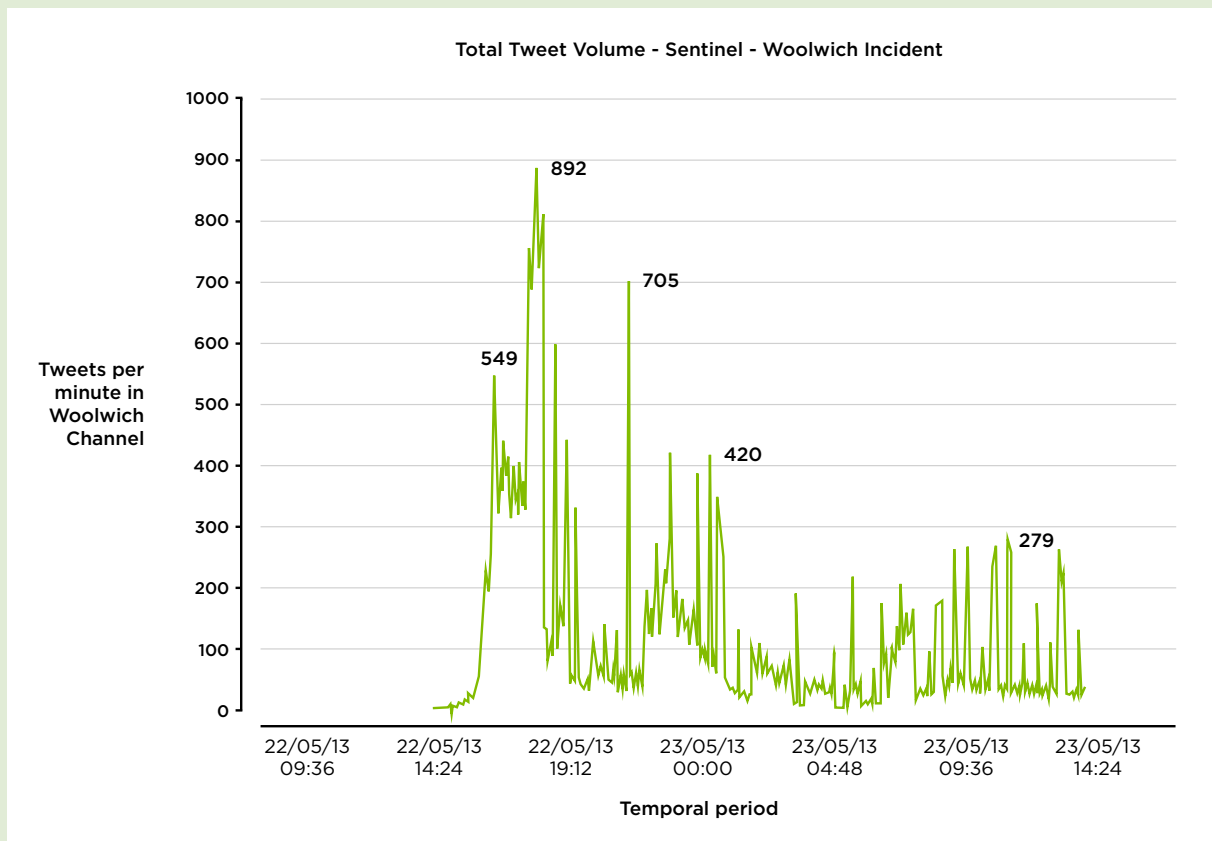
This case exemplifies just how social media can be used to track and trace elements of complex and unfolding patterns of social reaction. Initially the community of Twitter users mobilised online, using the broadcast function to interrogate their social network about what might be happening, and to share what they knew. Much of what was communicated in these exchanges was 'soft facts' in that the validity and reliability of the information was clearly uncertain. But the interesting thing is that collectively some of the participants did a good job in establishing an accurate definition of the situation. Their 'soft facts' were progressively 'hardened up' into a coherent public narrative, often as a result of interventions by mainstream media outlets reporting 'official sources'. When LBC radio in London began to suggest that the murder might be a terrorist incident, their statements were treated with a degree of authority and their intervention was widely retweeted.

Following on from such sense–making processes however, other voices became more active on social media. Notably, the English Defence League, an established community of interest, became a significant actor in the story once the incident was defined as Islamist extremism. They used specific rhetorical techniques to try and 'spontaneously' mobilise their support, as did those opposed to their ideas. The analysis clearly evidences that both sides struggled in this respect. Social media and the soft facts communicated by it were partisanly deployed, but converting interest and online support into a physical presence and action proved difficult.

In the days and weeks following the murder of Lee Rigby, reports were received from around the country of hate crimes being committed against people and symbolic buildings. Community tensions were running high. One story though 'cut through' this general climate of unease and concern. It told of how tensions relating to an EDL march in York had been diffused when members of a local Mosque had come out and talked with the marchers over a cup of tea, and a friendly game of football had ensued. This story was picked up by many national media outlets and was repeatedly recounted by journalists. It became almost iconic in the narrative of the Lee Rigby murder, conveying a morality tale of how community impacts were managed and a sense of order restored. For implicit in the narrative that was worked up was how symbols of English identity (a cup of tea and biscuits, and football) were effective at diffusing the potential for racial conflict.

However, our analysis of social media data around this event evidences that what actually happened, was significantly different from the 'mythologised' narrative propagated by broadcast media and the press. Very few far–right political supporters actually mobilised despite calls for them to do so on Twitter, possibly because there was a major national event elsewhere in the country the day before. A far more successful counter–demonstration was organised though. From the point of view of the interests of this study, two particular aspects stand out:

- The coalition of left wing and anti–fascist groups that quickly mobilised in response to the threat of a right wing anti–Muslim march was scaffolded by several key local institutions that provided an infrastructure and centre of gravity for the action.

- On both sides, potential participants were actively using social media to try and test the level of support that was actually going to show up so that they were not in danger of being isolated.

In a study of 'below the radar' activities the latter behaviour is especially intriguing as those involved were effectively using social media platforms as a 'sonar radar' – 'pinging' a signal to see what response they received, which would tell them something about the social environment they could expect. This was not about being above or below the radar, but about how social media affords a network of lots of micro–radars.

The final case study in the empirical sections of the report shifts focus to attend to how a particular online community mobilised its resources to intervene in the aftermath of the Rigby

case. This is especially interesting because it is a purely online action, potentially pointing to new directions that need to be taken in studies of what actually counts as community mobilisation. On the 27 May 2013, the webpage of the English Defence League was hacked by the online collective Anonymous UK. All data relating to personal information of EDL leaders and supporters, including addresses and telephone numbers had been copied and over the coming days was publicly posted on Pastebin. It is not clear how 'spontaneous' this action was from the available data, but it was certainly 'below the radar' in terms of coming unexpectedly from a non–traditional civil society organisation.

Conclusions

Taken together the case studies provide important insights into the role of social media in how and why community mobilisation does and does not occur, as well as the wider concepts involved in thinking about 'below the radar' social action.

Social media provides 'fast data' about emerging events and unfolding situations. But the speed of communication often means that the provenance, and validity and reliability of the data cannot be confirmed. As such, the information is often treated by potential users as contingent 'soft facts' – in due course it may have to be updated – but in the absence of more authoritative sources it 'fills' an information gap.

'Soft facts' can be especially influential when they provide a stimulus for collective action, in terms of a perceived problem or threat that needs to be countered. In the aftermath of the Lee Rigby killing, politically opposed groups sought to use social media to mobilise their supporters. It was interesting that both sides struggled to convert 'soft support' online into actually mobilised support in real life. Community mobilisation efforts did become more successful when they were supported by established institutions. This is in keeping with recent research identifying the importance of institutional infrastructures in providing a platform for social action.[25] But in part, it may also reflect the particular challenges of the circumstances being examined, in that in the aftermath of a dramatic crime where emotions were elevated, direct involvement entailed personal risk, given the conflict dynamics unfolding.

These findings connect to some wider implications the analysis has for thinking about civic action being below the radar and for the application of social media analytics to such issues more generally. With the widespread adoption of social media, being below the radar is not the same as being completely 'off the radar'. In terms of our analytics, we found we had to 'tune' our radar to be able to detect community mobilisation that was taking place. Given the volumes of data circulating through the social media ecology, the digital traces of how these mobilisation events were being organised were difficult to see with standard instrumentation, but they were there.

This reflects a more general point that many orthodox data science techniques for 'crunching' large volumes of social media data are probably insufficiently calibrated to 'sense' the kinds of complex social processes involved in community mobilisation. For example, our analysis shows how polarised forms of social support developed around different ideological standpoints, and these interacted in driving some individuals to physically mobilise. These interactive processes would not have been picked up by standard sentiment analysis.

How communities react and mobilise informal social control following major crimes is not something that has been studied very often. These are processes that are 'below the radar' in the sense that they have been largely neglected by the authorities. Some attention is now paid to conducting community impact assessments and consequence management of any community tensions, but this is far less significant in terms of public service agendas than finding the perpetrator. An interesting finding in this regard concerns the important role played by some institutions in scaffolding responses performed by civil society organisations and other networks. In this sense, an important question to ask when talking about being 'below' an organisation's radar is 'whose radar are we talking about?' The community mobilisation actions may have been below the radar for the police, but they plainly weren't below the radar for the individuals and groups engaged in rapidly and urgently mobilising.

Developing this line of thought, possibly the most important thing about the new information environment is how it has significantly widened the scope of the radar. Accepting the point about needing to calibrate the radar properly given the issue of detecting small signals of interest in the vast volumes of 'noise', it nevertheless remains the case that:

- Social media is directly engaged in enabling rapid forms of community mobilisation to be performed.

- The digital traces of these mobilisations that social media cast allow us to study and understand aspects of the social reaction to major events that were previously seen but unnoticed.

Nesta

3.4. Social action on social media

Carl Miller, Centre for Analysis of Social Media, Demos

Social media platforms are often seen to be nasty, even dangerous places. Misogynistic and racist language, cyber–bullying and hateful abuse all appear on social media platforms in large quantities. Extremist groups from across the political spectrum, even terrorists, have found a voice on these digital platforms, and use them to spread their message and find new recruits.

However, these new digital platforms are also used to help others. These are vibrant new places which people use to volunteer, organise, share skills, mentor, fundraise and donate, participate in local civic projects and work as activists to change laws and minds. Taken together, this is digital social action – a new class of social good and community resource.

Demos set out to understand the emerging contours of social action on one of the most important social media platforms: Twitter. We wanted to learn how it happened and who conducted it. We wanted to understand the contexts that inspired or provoked social action and the problems it was directed towards. We were especially interested in social action that usually sails 'below the radar'; conducted outside of the formal structures of charities or social enterprises, that runs on little or no money and – because of this – is often missed. Ultimately, we were interested in how this emerging class of social action could be supported and encouraged.

We looked at Twitter's reaction to two events. The first was the Somerset Floods during January and February 2014. Following the wettest weather on record, the UK suffered widespread flooding. Thousands of homes and offices were flooded, causing hundreds of millions of pounds of damage. The second was the launch of the Step Up To Serve Campaign,[26] launched on 21 November 2013. Supported by Prince Charles and the leaders of all three main political parties, it aimed to increase the number of young people routinely conducting social action in their local communities.

Both events prompted large reactions on Twitter, and we collected over 120,000 Tweets in total. The sheer scale of data has become a perennial problem to those interested in studying social media: these platforms now routinely create more data than any researcher can hope to read themselves. We used new technology that CASM has developed to respond to this problem, called 'method 51'. It allows us to train algorithms to automatically analyse Tweets. We give the algorithm examples of different kinds of Tweet, and it learns the kinds of language use that most indicate each. The advantage: we can analyse many more Tweets than we could ever read. The disadvantage: it's never perfectly accurate, and there will always be some noise in the analysis that we make using this technique.

We built eight algorithms overall for this research, and used them in a number of ways. Some sorted Tweets relevant to the events in question from those that were not. Some sought to find Tweets that represented social action in some way, and others to sort Tweets representing social action into those that came from charities and institutions from those that came from individuals. On average, the classifiers got it right about 78 per cent of the time.

The Somerset floods

After a period of extremely wet weather, widespread floods hit the UK in January and February 2014, causing millions of pounds of damage, and severe disruption to Britain's transport networks. This was selected as a largely unanticipated event that caused widespread hardship and disruption to communities throughout the UK. In the shocked aftermath of the floods, as affected communities across the country were reeling from the impact, people took to Twitter for a number of reasons.

Of the 116,000 Tweets that we collected between 5–10 February, 84,000 were about the UK floods and 39,000 were related to social action. These were sustained throughout the time of the period studied.

Figure 18. **Number of social action Tweets versus total number of Tweets relevant to the flooding, over time**
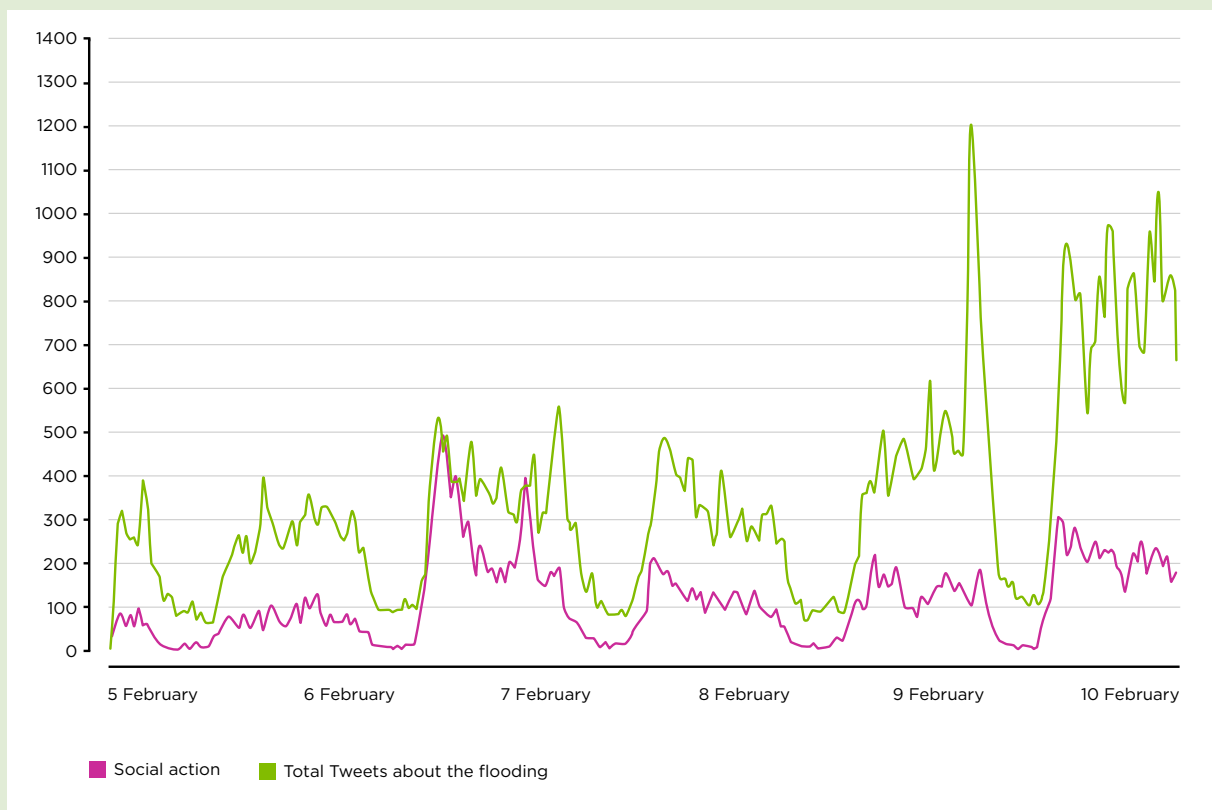
The vast majority of these Tweets were online social action, where people were using Twitter to try to help others by sharing information, updates, warnings and advice. These included people were sharing information on which roads and rail services were disrupted by the floods, weather reports, official warnings and advice and to report on the extent, damage or disruption of the flooding first–hand, often linked to a photo of the scene.

A smaller body of Tweets (about 3,000) represented offers or mentions of offline help to those affected by the floods. This included mentions of volunteers working to contain the floods: pumping water, laying sandbags and transporting supplies. Others offered to rescue, move or stable animals – both livestock and pets. People also used Twitter to offer a range of other help, such as sewage treatment, the provision of fodder or relief using 4X4 vehicles, and donating money to relief organisations. The reverse was also true. Some people needed help, and didn't know who could provide it. Strictly, of course, these Tweets did not represent social action, but instead the opportunity for it to be done.

Nesta...

Figure 19. **Proportion of Tweets by category of offline social action**

■	Animals
■	Volunteering
■	Specialists
■	Sympathy
■	Donations
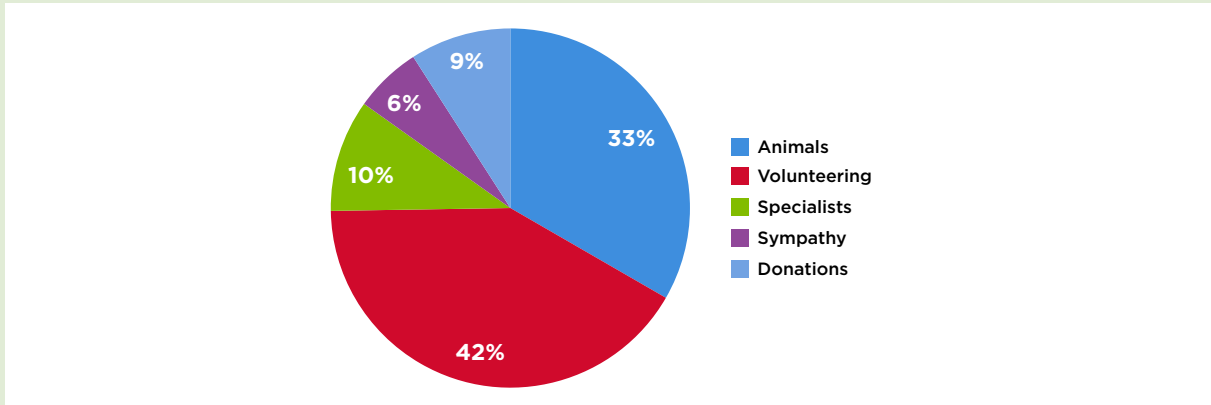
Of those that could be geographically located, a very broad pattern emerged. Tweets reflecting offline social action – of volunteering, donations and offers of help, tended to come from areas affected by the flooding. Tweets reflecting online social action – of sharing information and advice – tended to come from areas – especially London and central England – that were not affected by the flooding, but were densely populated. Most of both kinds were below the radar. These were not Tweets sent by large, organised charities. These were individuals taking to Twitter, largely outside of any organisational context, trying to help, or find help, however they could.

Figure 20. **Tweets related to social action. Orange = 'online social action'. Blue = 'offline social action'. Shaded = area affected by the floods. Size of colour = quantity of Tweets.**

Overall, these Tweets represented a form of digital community resilience. In the aftermath of a largely unanticipated disaster, communities across the UK reacted spontaneously to try to help one another. This research suggests that there is now an important digital dimension to this reaction. During the floods, Twitter became an important forum for important information, real–time updates from the ground, and a key exchange to make offers for help and to appeal for it.
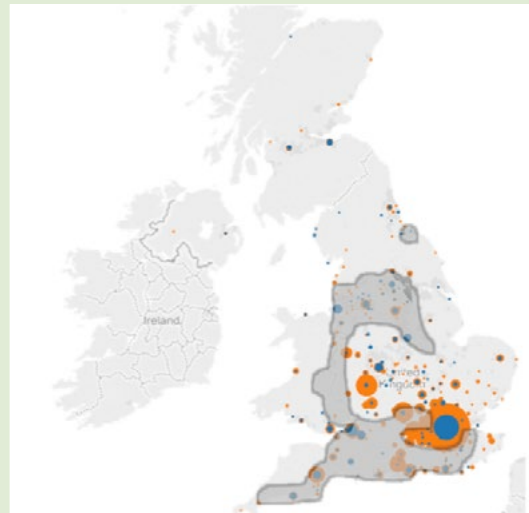
Step Up To Serve Campaign

The Step Up To Serve Campaign was a high–profile call to get over half of young people regularly conducting social action by 2020. It was launched in Buckingham Palace by Prince Charles, David Cameron, Ed Miliband and Nick Clegg, received major support from businesses and third sector organisations and was widely covered in the media. The Campaign asked for people to use Twitter to tell people how they planned to help young people volunteer using the hashtag '#iwill'. It was selected because it was an opportunity to study an event that has caused many people to speak about social action, their plans, views and priorities.

Over the first month of the Campaign, we collected 10,665 Tweets, of which 3,780 were relevant to the Step up to Serve Campaign. People used Twitter to respond to the campaign in a number of ways. Some Tweets were attempting to increase awareness of the campaign, and sharing links to news stories covering it. Others spoke about the campaign, discussing its objectives and intentions. Importantly, some people used Twitter to make pledges to conduct social action or help others to do so.

The nature of this response was markedly different from the first case study. In responding to the Step Up To Serve campaign, people did not use Twitter to look for, or offer, specific help. Instead, people offered their longer–term views and plans related to social action. Of the 3,780 responses to the launch of the campaign, 1,884 Tweets contained a 'pledge' to support or conduct social action; 690 of these came from, as far as we can tell, normal people acting outside of any kind of structure or organisation. The highest proportion of Tweets pledged were a commitment to social action in general. They contained pledges that supported the concept of social action and generally promised to help, but did not specify what this help would be. Some Tweets contained offers to help young people to find work as a form of social action, or to allow young people to improve their employability through volunteering themselves. Others specifically mentioned social action to support the local community, disadvantaged groups, the education system or to increase how social action is understood and appreciated.

Overall, Twitter was used to talk about people's experiences, what they have done that day. It was also used to mention what people find inspiring and laudable. Hence, some Tweets contained information about specific acts of volunteering that the person had done themselves, or had seen happen. This shows that Twitter can be studied to gain insight into attitudes towards social action – its prestige and sense of importance, the kinds that people think are a priority and how they intend to do it.

Conclusion

Two different case studies produced two very different reactions from Twitter, but overall they tell us an important truth: Twitter is a significant new forum for social action. It reflects, mobilises and unlocks new forms of mobilisation, of people helping others.

There are two important opportunities presented by the rise of Twitter, and social media more broadly. We can use these new platforms to understand social action better, and also to help encourage it.

First, there's a research opportunity. Social media platforms – and especially Twitter – produce huge amounts of data, much of it in real–time. Alongside the rise of these valuable new sources of social information are the emergence of powerful new analytical techniques that are capable of crunching and making sense of all this information. Even within the scope of this limited pilot study, 41,211 pieces of information relevant to social action were found. None

of these new technologies and techniques are perfect, and many struggle to produce insight as robust as that you would expect from conventional sociology. However, social media research methods are improving all the time, and, in my view, this is a valuable new research coalface to be worked on.

Second, there's an opportunity to proactively increase the extent and effect of social action. People used Twitter to both offer help and also ask for it, but the process was *ad hoc* and chaotic, and it wasn't clear whether suppliers and providers effectively linked up. Valuable offers for help are buried within huge quantities of other, irrelevant Tweets, and it's very difficult to find them. We need to do more to better support the growth and increase the effect of social action on social media.

I recommend an 'Ebay for social action' – a powerful application of digital platforms in other areas – especially commerce and the provision of services – be developed to directly connect people providing something with those who need it. When social action information is found, it could be centralised onto a real–time online platform, information exchange or brokerage hub, clearly related to a specific event and segmented either into the type of help that people are offering, or where the help is being offered. This would allow people to find and help, and also know what help is needed and who needs it. It's worked for commerce; it could work for social action too.

3.5. Mapping below the radar organisations on crowdfunding platforms

Maria Botello and **Noel Hatch**, European Alternatives

1. Introduction

It provides a methodology to map below the radar (BTR) activity on two crowdfunding platforms: Spacehive and Crowdfunder. Crowdfunding platforms seemed a suitable environment to find data on BTR organisations as they can provide these organisations with the financial resources needed to carry out their work. As BTR groups are characterised by their activity, the crowdfunding projects were the unit of analysis selected to account for this type of organisation.

This methodology has been developed into a software prototype that uses web–crawling techniques to navigate crowdfunding projects through hyperlinks and captures those projects complying with certain criteria. Namely, projects are expected to be run by BTR organisations if 1) they meet the condition upon which co-link analysis is based, 2) have reached their funding target, and 3) are not run by registered organisations.

The application has been used to collect data in order to conduct a case study for each of the platforms. This paper explains the methodological approach undertaken and the development and evaluation of the methods, based on the results obtained in the case studies. The paper will conclude with a summary of the lessons learnt across the two case studies and the next steps to be taken in the development of data–driven methods for BTR research.

2. Methodology to track BTR activity on crowdfunding platforms

What do we mean by below the radar organisations?

The first thing that becomes apparent when conducting any kind of research on BTR activity is the need to establish a working definition for this concept. To that end, a very concrete notion as to what can be considered a BTR organisation was selected from the multiple definitions documented in third sector research literature.

The term BTR organisations is used in the context of this project to refer to those organisations that, while carrying out voluntary activity with a social purpose, remain unregistered and unregulated. That is to say, they do not appear in official registers such as the Register of Charities by the Charity Commission. In summary, it describes those organisations that are under the 'regulatory radar'.[27]

Why are they likely to be found on crowdfunding platforms?

Crowdfunding is defined as 'the practice of funding a project or venture by raising many small amounts of money from a large number of people, typically via the internet'.[28] The fact that BTR organisations lack a legal structure implies that they might not be able to raise funds for their activity through the conventional channels used by organisations with a registered formal status. Therefore, crowdfunding platforms constitute an available alternative to BTR organisations in terms of funding.

Nesta...

As the scope of this project is limited to the study of BTR organisations within the UK, CrowdingIn.com, a directory of crowdfunding platforms that operate in this country, was used to select the platforms. The reason why the crowdfunding platforms Crowdfunder and Spacehive were eventually selected is that they are oriented to projects with a social purpose, often referred to as civic crowdfunding projects, and that presented an appropriate data structure for the method devised to find new BTR groups.

Co-link analysis as a way to identify thematically related projects

The main method used to identify new BTR groups in crowdfunding platforms is co-link analysis, which was implemented for the first time in the Issue Crawler. The Issue Crawler, a web application developed by the Digital Methods Initiative, deploys co-link analysis to enable the location and analysis of 'issue networks' on the web, defined as a set of web pages 'dealing with a common theme that are connected by hyperlinks'.[29, 30]

In the framework of this project, co-link analysis is used to identify a set of crowdfunding projects thematically related, rather than web pages, departing from a single seed project. Any project on those platforms could be used as a seed, that is to say, as the initial project from where to start the search - or crawl - of new projects by following its hyperlinks.

As this implies shifting the focus from crawling the whole web to crawling specific platforms, co-link had to be adapted to fit the particular data structure on the selected platforms. On these platforms, projects do not show hyperlinks to other projects. Instead, each project has hyperlinks to its backers' profile pages, and these from pages hyperlinks to the pages of all the projects they have pledged money to.

Figure 21. **Co-link analysis as deployed in this project**
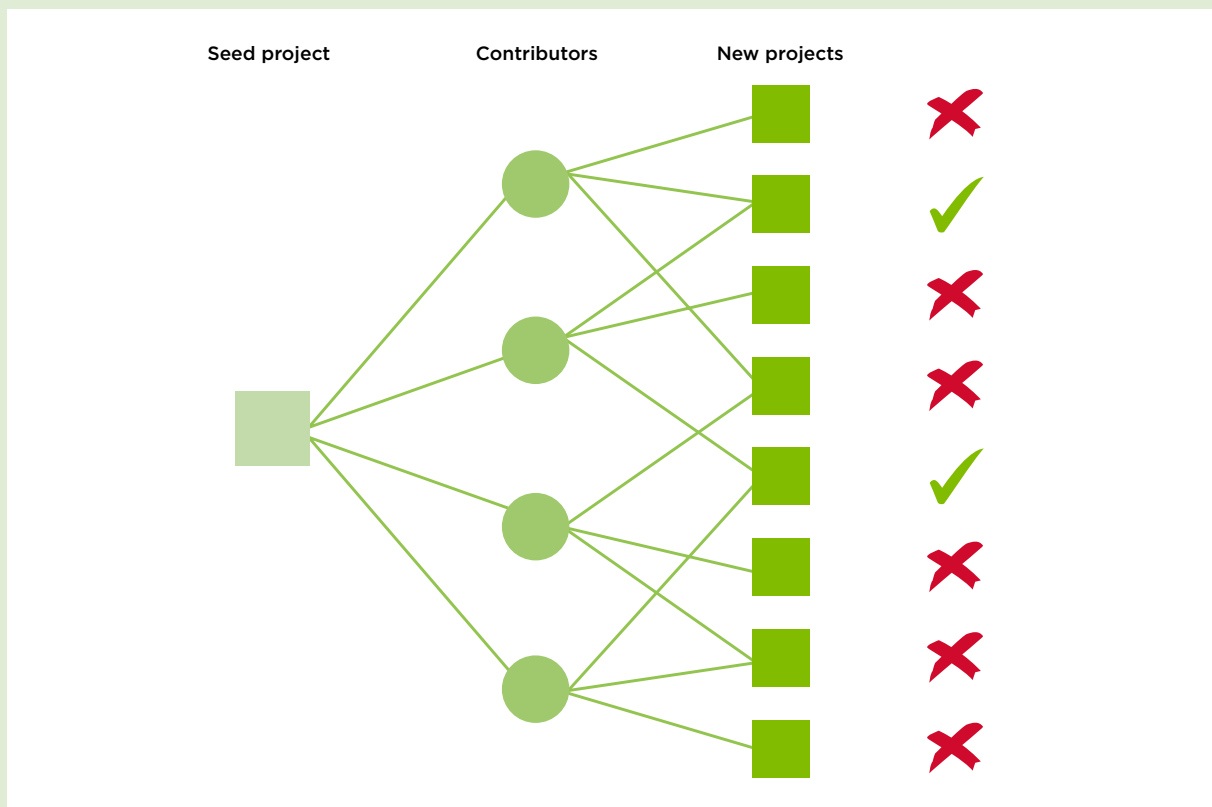
As a result, co-link analysis is deployed to follow the connections or hyperlinks between two different objects: the projects concerned and their contributors. Thus, given a seed project, new projects can be obtained. The premise here is that only the projects that have at least two contributors in common with the seed are captured for further analysis. That is to say, projects that have been funded by at least two people who also pledged funds to the seed project.

The new groups are expected to present a thematic relation with the seed and between themselves. That is to say, if a seed project is selected that pursues a social aim that concerns the environment, the projects obtained from that seed will all be expected to revolve around a similar theme.

Complementary conditions

Meeting the premise above didn't prove to be a sufficient condition to ensure that the new projects were the type of organisations sought for this research. Therefore, extra conditions were established in order filter the projects resulting from the co-link. Namely:

- **Being worth studying.** Only projects that have reached their funding target are deemed worthy of attention, the reason being that only these projects will actually receive the money to be realised. As projects are often developed over an extended period of time during which the organisations behind them can be considered to be active, the successfully funded project seemed a good indicator of sufficient and significant activity beyond the purely digital domain.

- **Being run by BTR organisations.** It is usual to find projects on these platforms that, while aimed at a social purpose, were created by registered organisations, from charities to different types of social enterprises, and therefore, do not count as being below the radar.

3. Deploying the methodology on Crowdfunder

Crowdfunder is currently the UK's leading reward-based crowdfunding platform. By April 2014, it was found to have more projects than all other UK platforms.[31] As a consequence, even if this platform is not exclusively aimed at projects with a social purpose, it presents a significant number of this kind of project or, in other words, enough relevant data.

Development of methods

Both the co-link analysis and the conditions above were implemented into an application intended to find projects by BTR organisations on Crowdfunder. With regards to the co-link analysis, an extra iteration was developed in order to get more new projects from a single seed project. That is to say, the program runs the crawler for the first time based on the seed URL entered by the user, and then uses the URLs of the resulting projects as the new seeds for a second iteration of the co-link analysis. The set of projects output by the app will include those projects obtained from both the first and the second iteration.

The condition that checks for those projects worth studying was implemented so that the app only returns those projects that had reached their funding target within the present or previous year. This was done to ensure that the organisations behind projects are 'alive' at the moment data is collected.

Nesta...

Figure 22. **Crowdfunder application**

The approach undertaken to filter out projects by BTR organisations was not the one originally intended. This alternative solution was implemented based on the fact that registered organisations tend to make explicit their legal status on the description of their projects on the platform. In doing so, a set of keywords was selected that account for the most common ways in which the different legal status is featured on Crowdfunder.[32] While running, the app checks whether the selected keywords are shown on the description of the project, and if so, discards it, keeping only projects that do not contain those specific keywords.

The app also allows the export of the results in the appropriate format for Gephi, an open-source network analysis and visualisation software package. When imported into Gephi, the resulting data is used to create a network visualisation where projects are represented as nodes and the connections between them represent their common contributors.

Case study

The application has been used to collect data for a case study aimed at testing the efficiency of the methods in identifying BTR organisation on Crowdfunder and exploring what insights can be drawn from the analysis of the data on BTR activity.

The selected seed project to run the application for the study was 'Grow a Future for Families', as the number of projects it returned was the highest among those seed projects used to test the app. Namely, a total of 16 new projects presumably created by BTR organisations. Then, the application was run a second time using the same seed but leaving unchecked the keywords filter, a feature enabled by the app. This resulted in up to 61 new projects both by registered and BTR organisations, allowing for a more accurate evaluation of the co-link analysis method.

This brings us to the next step undertaken, visualising the results using Gephi. Two different network visualisations were created based on the set of results above. On these networks, project nodes are coloured based on their category on the platform as a way to render visible thematic relations between them. Results on both networks were used in the evaluation of methods.

Figure 23. **Network of projects**
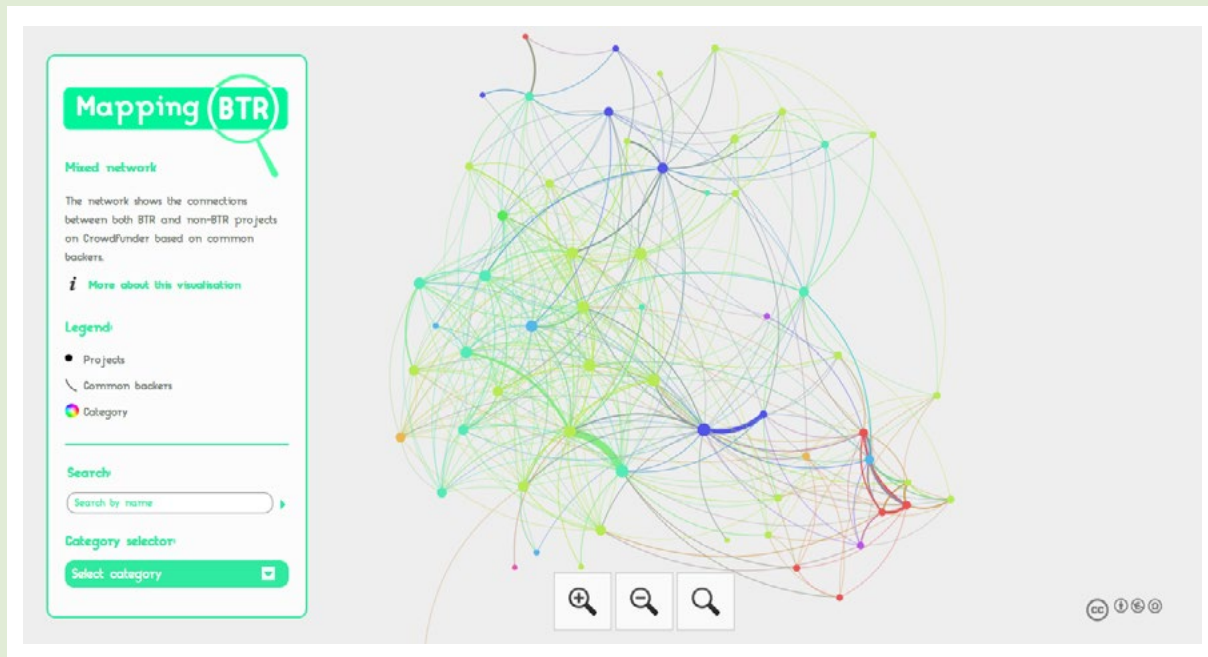
In doing so, the following questions were formulated: do the resulting projects pursue a social purpose? Are these projects thematically related? To what extent can platform categories be used to determine that? With regard to those projects that are supposedly run by BTR organisations, is this actually the case?

Based on the evidence provided by the biggest network, the method seems to work successfully. The majority of the projects returned have a social purpose and are thematically related to those other projects with which they share connections, even if they do not match the same platform category. The reason is that projects often do not stick to a single theme but rather have several thematic dimensions, which are not all reflected by the assigned category.

Thus, for example, the purpose of two projects on the platform 'Snact' and 'The Happy Pig' reflect strong environmental values, even though those projects were assigned categories other than 'environment'. And in fact, both projects, especially 'The Happy Pig', present a significant number of connections to projects under that category.

For the purpose of checking whether the organisations behind the 16 projects on the BTR network were indeed 'below the radar', a more detailed analysis of the pages of those projects on the platform and their own websites had to be carried out. In doing so, it can be concluded that even if the keywords filter works successfully, in that the selected keywords did not appear on the description of the projects, this method presents significant limitations, such

Nesta..

as the fact that it does not take into account the context of the keywords. Furthermore, it was found that registered organisations do not always reveal their legal status on Crowdfunder.

As a result of the above, only six out of the 16 projects on this network, proved to be by BTR organisations. Both the 'Enchanted Acres' and 'The Happy Pig' projects were intended to build spaces where people can explore and learn about permaculture and sustainable ways of living at no cost. Similarly 'A tipi for earth education', aimed at raising funds to build a tipi for school visits where children can learn environmental values and connect with nature.

The purpose of the 'The Big Sun Flower project' is to send out seeds to people at no cost, in order to raise awareness of centronuclear and myotubular myopathy by growing sunflowers all over UK and beyond. 'Anna Harbour Reopening' was created with the goal of bringing together the necessary resources to clear a local harbour of silt accumulated over 50 years and turn it into a community asset. 'Ask Amy' is an app prototype aimed at increasing the number of young people accessing and getting involved in politics. The project on the platform intended to raise money to develop the prototype into a fully functional application.

4. Deploying the methodology on Spacehive

Spacehive is said to be the world's first crowdfunding platform for civic projects.[33] It is primarily aimed at infrastructural projects at the local level in order to help communities to transform their public spaces. Thus, it seemed a suitable place to locate BTR activity.

Development of methods

Even though the data structure on this platform is suited to deploy co-link analysis, this method was not eventually employed due to the lack of the relational data needed for that purpose. That is to say, that at the time the method was being implemented it became apparent that funders hardly ever contribute to more than one project. This might be due to the local nature of Spacehive projects that results in funders contributing exclusively to projects in their own area.

The reason why it was decided to go ahead with the development of a more simple method to map BTR activity on this platform is that Spacehive is specifically addressed at civic projects. As this type of project is presumed to have a social purpose, it seemed that deploying a method based on the two conditions previously described could still fit the purpose of this research.

In compliance with the first of those conditions, an application was developed that captures all those projects on Spacehive that have succeeded in meeting their funding target, regardless of when, as the platform does not provide this information.

Moreover, in order to keep only those projects run by BTR organisations, the app combines two different filters: the keyword filter already described and another one that checks whether the names of the organisations behind Spacehive projects appear on certain registers.[34] Where this is the case, the project is discarded and otherwise captured for further analysis. This later condition could be implemented because on Spacehive the username of the project promoter usually matches the name of the organisation behind the project.[35]

Figure 24. **Crowfunder application**

Case study

The app was also used to collect data on Spacehive. As the method to detect BTR activity on this platform is not based on co-link analysis, and the relations between the resulting projects based on common contributors were not tracked, there was not suitable data available to build the network visualisation of thematically related projects.

As a consequence, other questions were formulated in order to check for the efficiency of this method: Is the fact that the platform is aimed at civic projects reason enough to ensure that the resulting projects pursue the type of social purpose that characterises BTR organisations? Do the resulting projects actually belong to BTR organisations?

From the resulting 13 projects, nine were found to pursue a social purpose to some extent. As for the second question, the combination of both restrictions seemed to work well for a good number of projects, as it was observed that six out of the nine projects that have a social purpose have been created by BTR organisations. However, the second restriction also presents limitations, as it was observed that among the results there are a few instances of projects whose creators on the platform do not match the name of the organisations behind them for different reasons. Moreover, it should be noted that a proper copy of all concerned registers would allow for more refined results. The purpose of the six BTR projects found in the platform were:

- **'Burghead Tennis'** was aimed at purchasing a portable tennis net and other equipment to be used by a tennis club and local communities in a school yard.

- The **'Northgate Herb and Fruit Beds'** project had goal was to convert a derelict eyesore into small community garden where people can grow and harvest their own fruit and vegetables.

- **'A Child's Dream'** sought funding to build a sensory garden for a young girl with autism as a pilot for a community garden.

The next three projects present a more artistic approach to their purpose.

- **'The Porty Light Box'** was intended to raise funds to turn a decommissioned phonebox into light box from which to display images by local artists, schoolchildren or other groups.

- The project **'Drum Together Brum Street Party'** was about creating a one-day free event for the local community featuring a programme of live music performances and workshops.

- **'After the riots - Happiness in Tottenham'** raised money to set up an exhibition based on a series of proposals intended to address the psychological and financial impact that the riots of 2011 had on this area.

5. Lessons learned across the two case studies

While civic projects on Spacehive have by definition a social dimension, they not always have the wider social purpose characteristic of voluntary activity. And in fact, when comparing the results from both platforms it becomes apparent that co-link analysis, based on crawling the common preferences of contributors, constitutes a more efficient way to check for projects with the kind of social purpose sought.

The proportion of BTR projects returned with the Spacehive crawler was nonetheless greater than for Crowdfunder. To some degree this might be due to the fact that two different filters, the keywords and the registers' restrictions, were used for the former platform. However, it seems that, even if both filters present limitations at present, the filter using the registers implies a more consistent approach to check for the legal status of organisations on crowdfunding platforms.

Furthermore, the notion of BTR organisations as those being under the regulatory radar seems appropriate for the purpose of developing automated data collection methods. Nonetheless, in order to enable a richer understanding on BTR activity on crowdfunding platforms, a more nuanced analysis of the projects and organisations identified from deploying those methods is needed.

For example, when researching BTR organisations behind projects on both platforms it became apparent that relatively often BTR activity appears combined in varied ways with that of registered organisations and governmental bodies. This is especially the case for those crowdfunding projects intended to have continuity over time, for which they receive support from other organisations in the form of funding, material resources or spaces.

6. Next steps for improving the efficiency of BTR methods

In summary, co-link analysis seems a valid method to map new projects thematically related, which, if combined with a suitable method to check for the type of BTR activity sought, can serve to identify BTR groups in crowdfunding platforms. Having the data in a format that suits methodological requirements would be crucial in making the necessary improvements to both methods.

To that end, a form of partnership with the platforms concerned could be established. For example, the data required to deploy co-link analysis could be enabled through an API, which would allow for the development of a more robust application.[36]

With regard to the methods aimed at filtering out projects by BTR organisations, it seems that if further developments were to be made, they should focus on improving the restriction based on the registers. However, platform collaboration would still be needed for that purpose. So, for instance, it would be extremely useful if platforms ask users to specify the name of the organisation they work for whenever they are acting on its behalf.

3.6. What we learned about the potential of using data–driven methods to understand hidden social action

So what do the five research projects tell us about the opportunity in data–driven methods?

Data–driven methods, particularly those looking at Twitter data, can help us understand new types of hidden social action. The work by the RSA, Cardiff University and CASM show that people do have conversations about local social issues online, and that data–driven methods can indeed be used to understand these new types engagement and of hidden social action that results from conversations. The study of the Somerset floods and the Lee Rigby incident show how Twitter in particular is a powerful medium for producing and sharing knowledge about a social issue that can lead to action.

What is unclear, however, is the extent to which activity and engagement online add value and impact, either through the distribution of valuable knowledge between peers or resulting in offline social action. The study of the Somerset floods did prove that this was the case in some instances, with people offering to help others affected by the floods with housing, but the extent to which there was an actual exchange is unclear.

As outlined in the introduction to this chapter, this research builds on the existing studies of below the radar groups. As demonstrated by the RSA study in Hounslow, data–driven methods can be used to create a local map of assets and resources that add to, and complement, existing maps of assets developed using offline methods.

An emerging understanding of what constitutes BTR activity when it is happening online, and methods for identifying this activity in large datasets: One of the big challenges is how to definite the type of activity that people engage in online; how do we define what constitutes a group or a BTR community and not just a few people having a conversation online? This is easier when looking at examples of a common social project that everyone participating in the project joins in with, such as a crowdfunding campaign, but more challenging when looking at more unstructured and varied data, such as Twitter conversations. However, the work on tools and methods such as the Sentinel, Localnets and 'method 51' demonstrate how the analysis of big datasets can identify those incidents where larger groups of people come together to discuss and potentially take action on a social issue. The question remains, as alluded to earlier, as to what the effect of this is and the extent to which the community forming online translates into social action either off or online.

However, the question that follows the ability to map this activity through new data–driven methods is whether or not below the radar is the right label? Firstly, much of the activity that the different methods have mapped can be described as being very different from the traditional offline informal activity, often focused on knowledge exchange and mobilising people around single events rather than ongoing long–term commitment to a social issue. We still know very little about the below the radar experience and how is it different from participating in above the radar activities, or the differences between off and online BTR activity.

Nesta...

There is often a relationship between those above and below the radar. While the purpose of BTR research is to identify and understand activity not driven by established above the radar organisations, such as charities and social enterprises, it is important to remember that this is a classification developed by academics and government to define what is and isn't in public registers, and not necessarily how people think about their own social action, or how charities think about mobilising and managing volunteers.

Case studies such as the study of the Lee Rigby murder and the Somerset floods show how key nodes play a strong role in mobilising online BTR activity. However, those nodes will often be an above the radar organisation, such as local mosque, or a person affiliated with this.

While this challenges the terminology and makes it a more complex field to research, it adds to the growing knowledge around the potential for civil society organisations in mobilising people to take part in social action, via social media in particular.[37]

Following on from this, the NCVO study showed how open data from funders can help 'lower the radar' and develop a more refined understanding of civil society activity through identifying those organisations that are established enough to receive a small grant but don't feature in any official registers.

Increased access to data will enable more and better data–driven BTR research; the challenge is to do this without invading property. All of the methods described in this paper are driven by the increased social action taking place online, and the access to this data through open data, web–scraping APIs etc.

In this context it is important to note that the five projects described above are primarily looking at data from Twitter and other social media platforms, crowdfunding platforms and open datasets in their analysis. Focusing on these data sources means that there is lots of online BTR activity and types of social action that is being mobilised through other platforms that the research methods won't pick up.

As all the research projects have focused on either accessing twitter API's, open datasets or scraping information that is already in the public domain, the issue of privacy has been less of an issue in this research that it was for the CAB and DataKind. However, as these methods become more refined and widely accessible, the civil society research community needs to continue the discussion on how to tread the fine line between using these methods and developing a better understanding of civil society without invading privacy.

4. Where next for big and open data for the common good

As demonstrated by the projects in this paper, there is real potential in using data–driven methods to both help charities develop better services and products, and understand civil society activity. Key lessons and recommendations for future work in this space include:

- **Better investment in data skills in the third sector:** Government and philanthropic organisations should invest in programmes that build collaboration between data scientists and organisations in the third sector, such as the collaboration between Datakind and CAB. While the work by CAB and Datakind has focused on the opportunities for a bigger charity such as CAB, these programmes should focus on how to support the sector as a whole, including smaller community and voluntary sector organisations.

  Data dives is one approach to this. Another example is the Open Data Challenge Series, a collaboration between Nesta and the Open Data Institute. It runs seven challenge prizes that invite businesses, startups and individuals to develop innovative solutions to address social issues using open data in areas such as Crime and Justice, Energy and Environment, Education, Housing and Food. This has already led new solutions such as MoveMaker, an app which uses open housing data to help social tenants house swap.

- **Support the development of better and more open data:** Much of the work on both researching and developing new products is only possible because of access to high–quality data and the sharing of open data between organisations.

  Open data from grant funders can lead to a better understanding of where money is spent and the size of civil society, as illustrated by the experiment mixing data from St Mungo's Broadway and CAB to detect and address homelessness issues, and the work by NCVO.

  However, to reap the benefits of this, potential organisations, from government and big grant funders to big and small charities, need to continue experimenting with new ways of opening up and collaborating around data, without putting privacy at risk. The 360 Giving programme which helps grant funders open up their data has pioneered much work in this space, and a continuation of this programme to support better collaboration around data is recommended.

- **Make better use of social media to engage with and build civil society capacity:** The work done by CASM, Cardiff University and the RSA has provided some early insights into how social media, in most cases Twitter, can be used to better understand new types of civil society activity and local assets. However, following on from this it would be interesting to explore the real potential of social media to help charities mobilise volunteers and communities to address issues and signpost people to local resources in more structured ways. Furthermore, how can public services, and local government in particular, use these approaches as a tool for understanding local conversations around issues and assets, to better respond to public issues, such as the Somerset floods.

- **A new type of civil society organisation, a new way of understanding of civil society:** While the vast majority of civil society activity still happens through established organisations such as registered charities, the below the radar grantees and Nesta's other research in to areas such as digital social innovation and crowdfunding show that people are constantly finding new ways of mobilising time, resources and money to address social issues via online platforms. As this trend continues, we need to keep working towards understanding what a digital civil society looks like.

Nesta

Endnotes

1. The Data Ambassadors: Peter Passaro, Chief Data Officer at Datanauts and a natural language processing expert, Iago Martinez, VP of Data Engineering at DataShaka, and Arturo Sanchez, who at the time was a data scientist at YouGov.

2. Arturo Sanchez Correa was replaced with Sam Leach, a data scientist at Inquiron.

3. Ian Huston, a data scientist at Pivotal.

4. Henry Simms, an Integration Technical Lead at Experian Data Quality UK, lead the design and development of the visualisations, and Billy Wong Principal at Hedgehog & Fox, was recruited to analyse the connections and relationships between issues.

5. St Mungo's Broadway is also open to the idea of sharing data and has approached Citizens Advice about continuing to collaborate on a new research project.

6. For more information about the Shooting Star Chase project see http://www.datakind.org/finding-30000-missing-children/

7. For more information about the Buttle UK project, see: http://www.datakind.org/making-sense-of-text-data-to-help-disadvantaged-families/

8. For more information about the Access project, see: http://www.datakind.org/projects/improving-access-to-education-by-supporting-tutors/

9. http://www.datakind.org/blog/dc-action-for-children-long-term-collaboration-for-long-term-impact/

10. The map can be viewed at: https://datakind-uk.github.io/child-poverty-commission-dashboard/

11. For more information about the project with The North East Child Poverty Commission, see: http://www.datakind.org/projects/communicating-about-child-poverty/

12. http://www.nesta.org.uk/publications/compendium-civic-economy

13. http://www.nesta.org.uk/sites/default/files/people_helping_people_the_future_of_public_services_wv.pdf

14. http://www.nesta.org.uk/project/centre-social-action-innovation-fund

15. http://www.nesta.org.uk/project/innovation-giving-fund

16. http://apps.charitycommission.gov.uk/showcharity/registerofcharities/SectorData/SectorOverview.aspx

17. http://www.birmingham.ac.uk/generic/tsrc/index.aspx

18. http://data.ncvo.org.uk/

19. http://epapers.bham.ac.uk/1550/ and http://www.birmingham.ac.uk/generic/tsrc/documents/tsrc/working-papers/working-paper-29.pdf

20. http://threesixtygiving.org/

21. https://opencorporates.com/

22. http://opencharities.org/

23. http://www.oscr.org.uk/charities/search-scottish-charity-register/charity-register-download

24. Development of Sentinel was supported by the TaRDiS project (Tackling Radicalisation in Dispersed Societies, funded by the European Commission, 2012-2015) and the After Woolwich project (funded by the ESRC, 2014-2015).

25. Sampson, R. (2012) 'Great American City: Chicago and the Enduring Neighbourhood Effect.' Chicago: Chicago University Press.

26. http://www.iwill.org.uk/

27. McCabe, A. and Phillimore, J. (2009) 'Exploring below the radar: issues of theme and focus.' University of Birmingham, Birmingham: Third Sector Research Centre.' See: http://www.birmingham.ac.uk/generic/tsrc/documents/tsrc/working-papers/working-paper-8.pdf

28. http://www.oxforddictionaries.com/

29. Marres, N. (2012) 'The re-distribution of methods: On intervention in digital social research, broadly conceived.' London University, London: Goldsmiths Research Online. See: http://research.gold.ac.uk/7773/1/Marres_redistribution_of_methods.pdf

30. The Issue Crawler deploys co-link analysis in the following way: from a set of seeds or website URLs entered by the user, the crawler returns only the URLs of those sites that receive links from at least two of the seeds. The resulting websites are expected to be relevant actors on the Issue Network of the particular theme introduced by the seeds.

31. http://www.ukbusinessangelsassociation.org.uk/news/crowdfunder-smashes-through-500000-funding-target-just-over-3-hours-equity-crowdfund-campaign

32. The keywords used were: CIC, foundation, charity, social enterprise, cooperative, co-op, co-operative, ltd, association, company, business, trust. It should be noted that these keywords were selected from the observation of a limited sample of projects. Thus, adding other well-documented keywords to that list would help to achieve more accurate results.

33. http://www.bigsocietycapital.com/how-we-invest/spacehive

34. Up to three different registers were found susceptible to contain the types of formal organisations present on crowdfunding platforms: the Register of Charities by the Charity Commission, the Register of Companies by Companies House, which includes regular companies, CIC and other social enterprises, the Mutuals Public Register by the Financial Conduct Authority, which includes organisations such as Co-operatives. Although it was not possible to get an official copy of those registers on a suitable format, I could get through more informal channels two updated files with the names of all registered charities and community interest companies. In order to get more accurate results, a proper version of all the above registers should be used instead.

35. This method was not eventually developed for Crowdfunder due to the fact that relatively often project owners do not register on the platform with the name of the organisation they work for, but use instead their own name or a random username.

36. An API, or application programming interface, enables applications to access data on platforms in an easy and stable way. As a result, apps that use API's do not need to be modified every time the data structure on platform pages changes. This is, however, the case for those applications using web crawling techniques such as the one developed for this project.

37. http://www.carnegieuktrust.org.uk/getattachment/ee2be1a3-8a94-4d8b-963f-94e29c8a9781/Making-the-Connection--The-Use-of-Social-Technolog.aspx

Nesta

1 Plough Place
London EC4A 1DE

research@nesta.org.uk
@nesta_uk
www.facebook.com/nesta.uk

www.nesta.org.uk