Big Talk about Big Data: Discourses of 'Evidence' and Data in British Civil Society

William L Allen, University of Oxford
Centre on Migration, Policy, and Society

Final Report – Seed Project

Working Paper of the Communities and Culture Network+
September 2014

Executive Summary

The terms ‘Big Data’ and ‘evidence-based research’ carry significant popular and academic currency, but understanding their meaningfulness and relevance for British civil society and voluntary organisations (CSOs), particularly those working on issues of migration or social welfare and working with academic researchers on projects, is unclear. When key members of these organisations talk more generally about ‘data’ and ‘evidence’ in their outward facing materials as well as in their day-to-day operational work, what do they mean by them? This seed-funded study from the Communities and Culture Network+ (CCN+) draws upon quantitative text analysis from corpus linguistics to analyse the published materials of a sample of CSOs which broadly work on migration and social welfare issues in the UK. This totalled over 9.5 million words and about 2,700 items from January 2007 to August 2014. Then, supplemented by qualitative interviews with key staffmembers of these organisations who are responsible for policy, research, external communications, or strategic management, the study identifies how they perceive data and evidence as relevant to their overall mission. Some key points that emerged from the study include:

- In their online published materials, the sampled CSOs tended to mention different types or sources of data, whereas mentions of evidence were usually modified by different qualities or strengths. The phrase ‘Big Data’ did not appear at all in the sampled organisations’ materials.
- This was echoed by the qualitative interviewing, which showed that ideas of ‘evidence’ and ‘evidence-based research’ are much more commonly used.
- CSO representatives talked about ‘evidence’ in at least three senses: robust, systematically-collected information that can be used to inform decision-making by others; research that is used to demonstrate or emphasise a pre-determined position; or information which promotes authoritativeness, impact, or importance of an issue. These different constructions subsequently help CSOs achieve their goals.
- Within the sampled CSO written materials, actions associated with ‘evidence’ tend to relate to supporting a given claim, while actions associated with ‘data’ tend to be related to showing or revealing a particular insight.
- Actions associated with ‘data’ as they appear in CSO materials tends to show a greater emphasis on creation or discovery. Mentions of ‘evidence’, meanwhile, appear to be more frequently used in final, presentational contexts.
- Uptake of data and evidence is influenced by several factors, including audience, available skills, and demands of the CSO environment.

Civil society and voluntary organisations that agreed to participate and be named in this study included: The Conversation UK, Eaves, Joseph Rowntree Foundation, Migrant Rights Network, Migration Yorkshire, National Council for Voluntary Organisations, Renaisi, and the Scottish Refugee Council. In this working paper, the set of published CSO texts that was analysed excluded The Conversation UK because of its unique journalistic purpose, although staffmembers’ reflections are included in the qualitative findings.

Contents

Introduction

The terms ‘Big Data’ and ‘evidence-based research’ carry significant currency in business, academic, and computing spheres. Navigating these terms’ meaningfulness and perceived relevance to British civil society and voluntary organisations (CSOs) presents opportunities as well as challenges for the sector (Ross 2013). Although defining a heterogeneous concept like ‘civil society’ is difficult and perhaps inadvisable across all contexts, some previous studies have broadly conceived it as comprising organisations which lie outside of the public and private business sectors, such as: non-governmental organisations (NGOs) or third-sector voluntary organisations; policy influencing bodies including think tanks; philanthropic organisations; campaigning charities; and recreational, cultural, or environmental charities (Bastow, Dunleavy, and Tinkler 2014). On contested issues like migration and social welfare, CSOs often cite data and evidence from many sources and comprising different types. But for what purposes and to what ends? Research from science and technology studies among other sources (Boswell 2009) has usefully shown how policy-making bodies use evidence not only to inform decision-making but also signal certain organisational values or characteristics. However, part of what is missing from existing knowledge about the role of evidence and data in influencing social change is a picture of the language and perceptions that shape CSOs understandings in the first place.

Aims and Objectives

This seed-funded pilot project had two objectives: (1) to examine the discourses around ‘evidence’ and ‘data’ in civil society organisations, particularly those working in issue areas related to migration or social welfare; and (2) to relate these discourses to perceptions about what social research accomplishes in civil society or voluntary sectors contexts. By linking textual analysis with semi-structured interviewing, this project aimed to reveal the different ways that these concepts are discursively employed and perceived. The key findings provide entry points for further discussion and exploration.

Methods and Datasets

This working paper draws upon two textual datasets (plural ‘corpora’, singular ‘corpus’) and a set of 11 qualitative semi-structured interviews. Interviews provided valuable windows into the ways that key members of CSOs perceive data and evidence in the course of completing their research, policy, and advocacy work. Given the exploratory nature of this pilot study, snowball sampling was used to identify relevant UK CSOs that were broadly operating within migration or social welfare topics, or actively facilitating public discussion about these topics as in The Conversation UK. Then, key staffmembers of those organisations whose job titles indicated involvement in research, policy, senior management, or communications activities were contacted for interview. Table 1 reports on the organisations that responded positively. The mission or statement of purpose of each organisation was found on its respective website, generally under a heading titled ‘About Us’. Interviews were transcribed, then analysed using Nvivo software. Examination of these statements reveals several instances of ‘evidence’ being cited as important to an organisation’s mission. Also, it shows how some CSOs operating in the migration or social welfare sector explicitly position themselves as campaigning or advocacy organisations, often with reference to generating some kind of change in attitude, policy, or economic outcome.

Table 1. Civil Society Organisations Participating in the Pilot Study

Name of CSO	Mission or Statement of Purpose	Based	Interviews
The Conversation UK	Give experts a greater voice in shaping scientific, cultural and intellectual agendas by providing a trusted platform that values and promotes new thinking and evidence-based research. Unlock the knowledge and expertise of researchers and academics to provide the public with clarity and insight into society's biggest problems. Provide a fact-based and editorially independent forum, free of commercial or political bias.	London	2
Eaves	Lead the way in exposing and addressing the overlapping issues of domestic abuse, sexual violence, and exploitation of women in the UK	London	1
Joseph Rowntree Foundation (JRF)	Want lasting change for people and places in poverty, communities where everyone can thrive and a more equal society. Now and for future generations.	York	1
Migrant Rights Network (MRN)	Working and campaigning in support of migrants in the UK. Our work brings together migrant activists and support organizations, think tanks, academics, faith groups and public sector representatives to advocate for a rights-based approach towards migration in the UK.	London	2
Migration Yorkshire	A local authority-led regional migration partnership. We work with national government, local government, and others to ensure that Yorkshire and Humber can deal with, and benefit from, migration. We work with agencies across the statutory, voluntary, community and private sectors to help support the delivery of high quality services to migrants in a way that benefits everyone living in local communities.	Leeds	1
National Council for Voluntary Organisations (NCVO)	Help voluntary organisations and volunteers make the biggest difference they can. We will use evidence (base what we say and do on the best research and our members' experiences), be creative (explore new ideas and approaches, looking for what will add real value), be collaborative (work with our members and partners to achieve the best results), be inclusive (value diversity and work to make sure that opportunities are open to all), and work with integrity (be open and honest and do what we believe is best for our members, volunteers and the voluntary sector)	London	2
Renaisi	Dedicated to making improvements to disadvantaged communities, putting local people at the centre of positive change through our extensive experience delivering neighbourhood-based regeneration programmes, local economic development initiatives and employment services	London	1
Scottish Refugee Council	A Scotland in which all people seeking refugee protection are welcome. It is a place where women, children and men are protected, find safety and support, have their human rights and dignity respected and are able to achieve their full potential.	Glasgow	1

Other important sources of CSO discourse about data and evidence include their published materials such as research reports, briefings, and press releases. These kinds of outputs are also valuable for this study because they represent a large part of the outward-facing profile of an organisation—particularly if other members of the public are accessing these resources. Study of these documents using a computer-assisted corpus linguistics approach can reveal patterns of language around mentions of 'data' and 'evidence' which may not be apparent from a surface, selective reading of a limited number of documents (McEnery and Hardie 2012). This study collected, as far as possible, all of the main documents published online by the eight organisations that participated in the study from 1 January 2007 to 15 August 2014. These were manually downloaded following a specification that captured all main publication types available on the respective CSOs' websites, excluding blogs. Table 2 displays the range of documents that these CSOs produce and publish.

Table 2. Types of Publications Retrieved for the CSO Corpus

Organisation	Publication Types Included in the Corpus
The Conversation UK	'Hard Evidence' column
Eaves	Consultations Current Research Past Reports Press Releases
Joseph Rowntree Foundation	Publications Policy Briefings Press Releases
Migrant Rights Network	Policy Reports Policy Briefings Newsletters (2007-2010) Latest News (2010-Present)
Migration Yorkshire	Publications Policy Briefings Consultation Responses
NCVO	Policy and Research Briefings Press Releases
Renaisi	Publications
Scottish Refugee Council	Research Reports Policy Briefs Policy Responses Annual Reviews Press Releases

Temporarily omitting The Conversation UK for this working paper, which featured a column titled 'Hard Evidence' and may have unintentionally skewed the resulting linguistic analysis by overstating the frequency of 'hard' as a way of describing 'evidence', the resulting corpus used for this working paper contained 2,704 items totalling 9,589,892 words. This corpus was then analysed using the Sketch Engine, a web-based piece of lexicography software that was designed for handling large corpora. The Sketch Engine generates 'word sketches', or snapshots of how a given word functions in a corpus, based on grammatical rules that are automatically applied to the textual data. This can reveal, for example, what kinds of adjectives most frequently describe a noun.

Given the non-representative sample from which the textual corpus was derived, it was desirable to compare the CSO corpus with a larger, more comprehensive corpus to see whether the observed linguistic phenomena were particular to that dataset or were reflecting general features in English. The Feed Corpus, a reference corpus available through the Sketch Engine, contains over 645 million words of English automatically drawn from a broad cross-section of online sources (Minocha, Reddy, and Kilgarriff 2013). Due to its size, recent collection, and broad online content, this corpus provides a useful initial comparison to the smaller CSO corpus by showing how words tend to appear in general online English. However, any direct comparison between the two corpora is unadvisable given the different ways in which they were built as well as sources from which they were derived.

Defining the Concepts of Data and Evidence

Before presenting some key findings from this pilot study, it is relevant to compare and contrast some 'dictionary definitions' of the terms 'data' and 'evidence' which may inform the ways that CSOs use them. Table 3 presents these definitions alongside a definition provided by the Alliance for Useful Evidence, a major network of researchers, businesses, and charities that operates in the UK as well as internationally.

Table 3. Contrasting Definitions of Data and Evidence

	Oxford English Dictionary Online (2014)	**Macmillan Dictionary Online (2014)**	**Alliance for Useful Evidence (2014)**
data	An item of information; related items of (chiefly numerical) information considered collectively, typically obtained by scientific work and used for reference, analysis, or calculation Quantities, characters, or symbols on which operations are performed by a computer, considered collectively. Also (in non-technical contexts): information in digital form	Facts or information used for making calculations or decisions Information in a form that a computer can use	
evidence	Actually present; prominent, conspicuous; an appearance from which inferences may be drawn; an indication, mark, sign, token, trace Ground for belief; testimony or facts tending to prove or disprove any conclusion; information, whether in the form of personal testimony, the language of documents, or the production of material objects, that is given in a legal investigation, to establish the fact or point in question	facts or physical signs that help to prove something facts, statements, or objects that help to prove whether or not someone has committed a crime	encompassing anything from research and evaluation studies, to administrative data, expert knowledge and stakeholder consultations, and potentially even data from social media [including] any particular method or research discipline, as long as they are robust, unbiased, and appropriate for the question(s) being asked by decision makers

It is apparent from the first row of definitions that 'data' is associated with digital technology, numeracy, and calculation by computers. Data are typically sourced from scientific or other systematic enquiry, and usually used for reference or form the basis of subsequent analysis and decisions. Meanwhile, the definitions of evidence tend to centre on proof and presence—sometimes in legal or criminal investigation settings—and may come from many sources including oral or written testimony, documents, artefacts, or other measures. These contrasting definitions suggest that 'evidence' is more closely associated with contexts where there are competing conclusions at stake (guilty or not guilty, correct or incorrect, true or untrue), while the term 'data' is used in contexts where calculation and systematic analysis are more predominant. Interestingly, the Alliance's definition of evidence includes reference to different kinds of data but places three key requirements on them: that they are robust, unbiased, and appropriate for the questions being asked. Therefore, according to their usage, data could be considered as evidence under certain conditions.

If CSOs produce outputs and use terms like 'data' and 'evidence' according to these dictionary definitions, then it would be expected that 'data' would be used alongside references to different kinds of analytical techniques as well as the sources of those data. Meanwhile, 'evidence' would be used in contexts where statements about proof or the level of confidence with which information can be considered as conclusively supporting (or refuting) a claim are more important. Use of the term 'evidence' may also be accompanied by references to its sources, relevance for a given decision, or quantity with the goal of drawing attention to its strengths in an argument.

Key Findings

This section outlines some key points that emerged from initial analysis of the corpus datasets and qualitative interviews—both separately and comparatively. Again, it is important to note that the CSO corpus on which this working paper draws excludes The Conversation UK, although its representatives are included in the qualitative interviewing findings. Since this project was a pilot study, these descriptive findings serve as stepping stones to deeper analysis in the future. The breadth and depth of corpus data, for example, present opportunities for additional qualitative analysis, as well as more sophisticated statistical testing to determine what patterns are more significantly associated with either target word.[1] Furthermore, given the sampling limitations described above, these results should not be interpreted as generalisable to either UK civil society as a whole or to all organisations specifically working with migration and social welfare topics. However, as is argued in the subsequent Key Issues section, these initial findings do reveal some interesting linguistic and practical phenomena which merit discussion and may have implications for CSO self-presentation.

Results: CSO Corpus Dataset

In their published materials, the sampled CSOs tended to mention different types or sources of data, whereas mentions of evidence were usually modified by different qualities or strengths. The phrase 'Big Data' does not appear at all in the CSO corpus. Table 4 shows the top twenty modifiers of 'data' in the corpus. Sources like the Census and local administrations are particularly present. Also, types of data including interviews,

[1] Examples of more sophisticated statistical work, including some using the capabilities of the Sketch Engine, include Blinder and Allen (2014) and Vicol and Allen (2014).

longitudinal, and general categories like quantitative or qualitative, are highly frequent. These characteristics (source and type) are highlighted in blue. 'Big' as a modifier for DATA does not appear in the corpus. However, this result should be cautiously interpreted because of the non-representative sample: other CSOs, particularly those outside the migration and social welfare realm, may indeed talk about Big Data in their published materials.

Meanwhile, the top twenty modifiers of 'evidence' include more references to subjective qualities or strengths as highlighted in orange. These are indicated by adjectives such as LITTLE, CLEAR, STRONG, and LIMITED. These findings seem to match the dictionary definitions of 'data' and 'evidence' as earlier discussed.

Table 4. Top 20 Modifiers of DATA and EVIDENCE in CSO Corpus By Frequency

Modifers of DATA	Frequency	Modifers of EVIDENCE	Frequency
survey	135	little	211
census	128	research	145
qualitative	105	anecdotal	84
quantitative	87	available	77
available	78	clear	73
administrative	59	strong	66
local	48	limited	64
more	41	further	55
recent	40	empirical	51
statistical	39	recent	50
interview	38	more	46
migration	35	other	44
longitudinal	33	robust	39
such	33	international	35
reliable	32	good	32
national	32	study	32
other	32	case	31
new	31	statistical	29
detailed	30	qualitative	29
late	30	new	29

When compared with the larger Feed Corpus, a similar phenomenon was observed, as seen in Table 5 which also displays the top twenty modifiers of these terms alongside each other. Here, 'big' is the most frequent modifier of DATA. This may be explained by the fact that the Feeds corpus is built automatically through Twitter links: topics related to technology and data, including Big Data, are likely to appear on this platform.

Table 5. Top 20 Modifiers of DATA and EVIDENCE in Feed Corpus By Frequency

Modifers of DATA	Frequency	Modifers of EVIDENCE	Frequency
big	3,820	scientific	1,434
personal	1,818	strong	840
economic	1,304	further	755
user	895	enough	735
late	874	anecdotal	678
raw	806	clear	659
sensitive	806	empirical	642
open	726	sufficient	434
recent	618	compelling	408
historical	596	DNA	369
real-time	524	physical	365
satellite	476	direct	352
actual	420	overwhelming	320
scientific	418	convincing	303
survey	389	solid	299
clinical	378	insufficient	283
relevant	318	forensic	282
unstructured	315	circumstantial	267
unlimited	315	conclusive	255
accurate	292	ample	237

General online language around the phrase 'Big Data' is distinctly more abstract, future-orientated, and emotive compared to 'data' by itself. An advantage of using the larger Feeds corpus is that it captures some linguistic phenomena which may not be present in the smaller, more selective CSO corpus. As mentioned earlier, 'big data' does not appear at all in the CSO corpus. Therefore, the study turned attention to the 3,820 instances of 'Big Data' as it appears in the Feed Corpus. This analysis reveals the term is used as part of more abstract and emotive phrases, compared to DATA by itself. One example involves the phrase 'of Big Data' (or 'of data'). This kind of grammatical construction can signal a modifier that expresses particular qualities or types—such as '*amounts* of data' or 'the *usefulness* of Big Data'. Table 6 compares the top twenty words associated with this kind of phrase as they appear with both DATA and BIG DATA. It shows that phrases involving 'of data' tend to express quantities (AMOUNT, VOLUME, PIECE, LOT) as shown in blue. In this instance, 'GB' and 'MB' refer to 'gigabyte' and 'megabyte', which are measures of data storage. Meanwhile, words indicating relative value, qualities, and statements of present or future developments (CHALLENGE, PROMISE, ADVANTAGE) tend to predominately appear with BIG DATA. These also include some metaphorical language (as in the '*rise of big data*').

Table 6. Top 20 Words Associated with the Phrase 'of data' or 'of Big Data' by Frequency

Words Before 'of'	Frequency	Words Before 'of Big'	Frequency
amount	1,916	use	48
analysis	822	era	36
lot	613	power	24
use	587	world	24
type	481	promise	16
set	453	rise	14
volume	449	value	12
year	404	analysis	11
kind	368	challenge	11
piece	359	impact	11
collection	353	age	11
GB	336	benefit	10
source	308	advantage	9
lack	242	definition	8
bit	176	importance	8
release	169	potential	8
value	160	aspect	8
wealth	157	sense	8
MB	155	emergence	5
quality	145	implication	5

Within the CSO corpus, actions associated with 'evidence' tend to relate to supporting a given claim, while actions associated with 'data' tend to be related to showing or revealing a particular insight. Focusing on the verbs associated with 'data' and 'evidence' shows what kinds of actions are ascribed to these objects—or, more crudely, helps to reveal what data and evidence 'do' or accomplish in the context of these CSO's published materials. As indicated by the words highlighted in blue in Table 7, verbs associated with DATA when it is the subject of a sentence tend to present information. Meanwhile, the verbs used with 'evidence' express more of a particular stance, as illustrated by words like SUGGEST (over five times more frequently used with 'evidence' rather than 'data'), SUPPORT (three times more frequent), CONFIRM, and HIGHLIGHT. However, some of these verbs indicating a degree of selection are also present with mentions of DATA, such as DEMONSTRATE and SUPPORT.

Table 7. Top 20 Verbs Associated with DATA and EVIDENCE as Subjects in CSO Corpus

Verbs & DATA as Subject	Frequency	Verbs & EVIDENCE as Subject	Frequency
be	489	suggest	393
show	103	be	331
suggest	64	show	107
have	53	indicate	56
relate	32	have	52
do	31	support	49
provide	29	demonstrate	35
set	25	do	30
indicate	22	present	29
regard	17	regard	28
support	14	relate	25
include	14	highlight	16
use	14	review	15
reveal	13	emerge	14
present	13	confirm	12
demonstrate	11	point	12
become	11	concern	11
cover	10	provide	10
share	9	appear	9
allow	8	exist	9

Comparing this to the Feed corpus in Table 8 shows some similarities, particularly in the predominance of verbs linked with mentions of ‘evidence’ that express a particular stance—for example, PROVE, SUPPORT, and CONFIRM. This finding may be explained by the fact that ‘evidence’ can be used in several contexts but particularly in criminal or legal settings, a sense that is illustrated particularly well by the presence of EXONERATE.

Table 8. Top 20 Verbs Associated with DATA and EVIDENCE as Subjects in Feed Corpus

Verbs & DATA as Subject	Frequency	Verbs & EVIDENCE as Subject	Frequency
show	2,752	suggest	1,438
suggest	685	show	1,076
indicate	428	support	640
centre	402	indicate	318
plan	295	link	215
reveal	280	point	206
share	249	exist	173
store	237	prove	161
process	219	regard	129
warehouse	198	emerge	128
feed	196	demonstrate	109
point	189	mount	100
file	175	back	97
transfer	171	present	91
regard	171	confirm	70
gather	160	contradict	46
capture	147	accumulate	29
center	141	implicate	27
mine	134	surface	25
bind	97	exonerate	19

Looking at the activities typically done to data in the CSO corpus shows a greater emphasis on creation or discovery. Mentions of evidence, meanwhile, appear to be more frequently used in final, presentational contexts. Turning attention to the instances where 'data' and 'evidence' were the objects of a sentence rather than the main subject can show the range of actions that were done to them. More simply, it can help reveal what users of data and evidence do with both kinds of information. Table 9 shows the top twenty verbs associated with mentions of each key term when it appears as the object of the sentence. Words highlighted in blue express a sense of collection or generation. Verbs like COLLECT and GENERATE are more likely to appear with 'data' in this corpus rather than 'evidence'. However, as seen by the words highlighted in orange, the actions done to both key terms express a sense of presentation or finality. In this case, the presence of SUBMIT is particularly unique to 'evidence'. Closer qualitative analysis revealed that these instances centred around CSOs that were responding to official Inquiries by the UK government by 'submitting evidence'.

Table 9. Top 20 Verbs Associated with DATA and EVIDENCE as Objects in CSO Corpus

Verbs & DATA as Object	Frequency	Verbs & EVIDENCE as Object	Frequency
use	337	be	1,403
collect	264	provide	298
provide	126	find	235
analyse	104	present	86
be	87	gather	72
have	66	have	60
gather	50	review	54
present	43	give	47
generate	38	collect	45
obtain	34	use	42
include	33	submit	35
publish	31	consider	34
show	27	see	31
exist	25	exist	27
produce	19	show	27
require	19	examine	26
draw	17	support	25
contain	16	need	22
examine	16	grow	21
link	16	require	19

This initial descriptive analysis of the CSO corpus, and subsequent comparison to the larger Feed corpus, shows some interesting differences between the sets of words used around mentions of 'data' and 'evidence'. In the context of civil society publications, data tend to be described in terms of their sources or types, with more attention given to the manners in which they were generated, obtained, or collected. This closely matches the predictions generated by comparison of the dictionary definitions. Meanwhile, when CSOs use the term 'evidence', they tend to use it to support, indicate, or suggest particular points—and use appropriate adjectives to highlight the relative quality or strength of that information. These linguistic phenomena are also observed to a certain extent within the larger Feeds corpus,

suggesting that CSO discourse displays some similarities to the ways that the terms 'evidence' and 'data' are used in general online discourse.[2]

Results: Qualitative Interview Data

In the context of migration or social welfare, 'Big Data' and 'data-driven' do not prominently appear in CSO awareness or activities. Evidence and 'evidence-based research' are much more commonly used. Given that defining 'Big Data' is conceptually difficult, there were three broad kinds of reactions to the term 'Big Data' across most of the organisations that participated in interviews: (1) a concept with which participants were unfamiliar or did not come across in their everyday professional work; (2) a concept that was familiar to participants but did not immediately present any significant advantages over and above existing datasets, such as the ability to improve the CSO's ability to achieve its strategic goals; or (3) a familiar concept that had potential for improving the CSO's ability to achieve its goals but was still yet to be developed:

> They're not phrases that I really hear at all in our work…I suppose because we're not involved in discussions where people are making decisions about services. We are involved in discussions about broader migration strategy or what's new in terms of policy or number trends. (Policy and Research Manager)

> I think it's definitely on people's horizons as something that they know they need to think about, and I get the feeling that most people don't necessarily know what it means or certainly what it means for them. (Senior Research Officer)

> It's fascinating that all that data, that the big data revolution is happening but also totally worrying that anybody and everybody can access the data and do some research on it no matter what their skills are. (Programme Manager)

Two possible reasons for this relative lack of awareness of Big Data emerged. First, the issue area of migration and social welfare, especially as experienced by these organisations, does not immediately lend itself to large, complex datasets. Commonly used numerical data tended to include public opinion polling or official government statistics on population. Second, the scope of data on which some organisations tended to rely seemed to be more focused on smaller groups of service recipients or stakeholders:

> The only kind of data driven immigration policy I've come across is the net migration target. So…the data shows that X million people have come in in the last ten to fifteen years and X hundred thousand people have left…That's the kind of major data that is happening at the moment. But…the current Government data tends to be driven towards how can we reduce numbers?...It's all kind of geared towards reducing numbers. And irregular migration is another battle and they actually don't have data on that. (Public Affairs Officer)

> I think part of it is the scale that we work at. So I think of some of the services that I evaluate or the services that I deliver, I could add up all the data. It doesn't create all that many data points. So you don't need a big data mindset. If you're delivering on a neighbourhood scale, there's only going to be at most 1,000 people that might be coming through that one system. There's no drive to say, "actually we're working

[2] It is also possible that CSOs are also tapping into a vocabulary that is more 'academic' in order to bring a degree of authority to their published work. Comparison to corpora built from academic sources, rather than general English, would help to see if this is the case.

> with so many millions [of datapoints] that we can't begin to think about them." So I think that's one factor for us: we tend to work at smaller levels with smaller projects. (Director)

CSO representatives talked about 'evidence' in at least three senses: systematically-collected information that can be used to inform decision-making by others; research that is used to demonstrate or emphasise a pre-determined position; or information which promotes authoritativeness, impact, or importance of an issue. These different constructions subsequently help CSOs achieve their goals. When asked how they perceived 'evidence' as it is used in their work, participants tended to reference this term in different ways depending on the intended use. If the purpose was to identify important aspects of an issue which were either previously undetected or potentially impacting upon a decision, then evidence was perceived as a key resource for improving future practice. Processes of informing, testing, and explaining are all important parts of using evidence in this way. 'Evidence' in this sense is more closely linked to the Alliance for Useful Evidence's own conceptualisation, where qualities of rigour and 'robustness' are important considerations:

> Seeing what works, by finding out what works on the ground and what doesn't work and what relates, what kind of things cause some of the things to happen. Just to…be able to advise people properly on policies and things like that based on solid, robust research. (Programme Manager)

> We consider the real life experience of our service users as one of our evidence bases because when you support hundreds of women or thousands of women, what they tell you and the patterns and the trends that you can pick up of their experience could be considered as evidence. (Lead Researcher)

> To frontline practitioners or people that are service managers in their local authority, evidence based is much more, "has somebody else done it before?" To them, evidence is often some sort of proof of concept. (Director)

> It means not just making decisions without being well informed, without understanding the evidence...For us, it's something that we try and do fairly systematically. (Research Manager)

Meanwhile, evidence can also be used to promote and illustrate particular viewpoints held by an organisation. The second quotation below illustrates how CSOs might employ evidence to highlight the importance of an issue as well as generate new ideas. Here, processes of linking activities to organisational missions or values—such as advocating for particular world views or policy stances—are important to consider. This sense seems to be more reflective of the traditional dictionary definitions of 'evidence', where it fulfils a role of informing decisions among competing conclusions. In the context of migration or social welfare, these conclusions might be policies or Government positions:

> Evidence, to me, is a fairly political term sometimes. Like I have a problem or I have an assertion or I have an opinion and I want evidence for it. (Assistant Commissioning Editor)

> A lot of the time if it's NGO research it tends to be 'here is an issue that we know is an issue' and let's look at gathering data, or an evidence-base to quantify that issue, exemplify it and show that there's a real issue here and provide recommendations to

> Government. So I think there is a tendency…to fall into that field of 'we already know that this is an issue and let's do some research and write up a report to show that it is'. (Head of Policy and Communications)

> So evidence from key people tends to help our work because what that does is provide the key information needed to make sure that parliamentarians fully understand the key issues of the day. So gathering evidence, gathering information from key organisations: we tend to be the experts and it is very helpful to our work. (Public Affairs Officer)

> For example, on this research we are working on, we are speaking to about 80 women who are Third Country national migrant women. We are going through a set of questions that we ask them and the issues that we are discussing. We are gathering that data to have an evidence-base to advocate for policy change or you know, services to be provided for these women. (Lead Researcher)

Finally, evidence can be used in a third sense that involves showing authority and impact to external partners. This may involve conscious reflection on the part of staffmembers to present organisational activities in particular ways:

> The Holy Grail of that stuff is can you evidence the impact that you've had on your beneficiaries? And that's really difficult to do. So, yes, the kind of other [evidence] that would be around 'what's the need for the service that you want to provide', or something about how you've provided it in the past. And if you can't get to that impac[t] then maybe [evidence] about the services or the high quality or you're able to do it at scale and that kind of thing. (Senior Research Officer)

> I think if you use the word evidence it will give some sort of authority to do with that word that there isn't just in data. Because I think data does suggest that you need to do something with it, whereas evidence almost is like you're presented with a *fait accompli* and they don't need to do anything with it then, "The evidence is that we need X." (Policy and Research Manager)

Here, it is suggested that data form part of a larger category of 'evidence', or at least serve a different function. This is also indicated by the dictionary definitions given earlier, where data were presented as discrete items of information.

Uptake of data and evidence is influenced by several factors, including audience, available skills, and demands of the CSO environment. As also seen in some of the quotations above, participants emphasised that different circles—policy, service provider, ground-level practitioners—view the usefulness of evidence differently:

> [Y]ou have to remember that service provision, particularly for public services, is about people who have vulnerabilities or are victims of crime or victims of prejudice. And it's about meeting those needs, it's not about the bigger population all of the time. (Policy and Research Manager)

> …[S]ome organisations just don't [use evidence and data]. And don't even see the case for it and maybe for certain very precise services they might not need much evidence. You know for some organisations, it would clearly improve their services, but for others depending on what they do, the case is less obvious and maybe it's fine not use to evidence. It might not be appropriate in all cases. (Research Manager)

Stakeholders' capacity to access and understand data is also sometimes constrained by a lack of skills, particularly statistical knowledge. This possibly maps onto more generalised anxieties about numbers and visual presentations of data:

> Their [local service managers] skillset isn't massively numerate, I suppose, and they are often problem solvers and programme managers. They'll work out how to take a budget and do as many things as they can with it, but they're not thinking about data in that sort of way at all. (Director)

> There is the fear of either too many charts for example, it's not just too many numbers all rammed in the text but too many charts that can be overwhelming. (Senior Commissioning Editor)

Finally, external demands from the economic, political, funding, or media environments in which CSOs operate can also influence how these organisations view evidence:

> …For campaigning purposes to highlight how awful the situation is and the impact it has on the individual [and] other organisations, etc. So sometimes it [evidence-based research] is things that are done quickly and fast just to keep the momentum of an issue, both in terms of the decision-maker, which is the Home Office and the UK Government, or it is also to create media attention or some public attention around the issue. (Head of Policy and Communications)

> People find numbers more convincing. So if they are making a funding application, they feel more desperate to have some numbers in it than to have qualitative research findings in it as a basis of funding them. So I suppose some of it might lie in what funders' expectations are as well. (Policy and Research Manager)

> Also commonly it's about what evidence can we use to put in a funding proposal or to make a grant application to boost our work a bit. (Senior Research Officer)

Reflecting on the linguistic and interview datasets, it is apparent that the concept of 'Big Data' does not appear to have much presence in either CSOs' public facing materials or their perceptions as they go about their research, managerial, advocacy, or policy-informing work. Also, CSO representatives' perceptions of how 'evidence' can be used, particularly as a resource for advancing strategic campaign or funding goals, matches the conclusion of the linguistic analysis which suggests that 'evidence' is more typically associated with expressing and justifying a particular stance.

Key Issues

Explaining why CSOs present themselves to key stakeholder groups in certain ways partly rests on a fuller understanding of the values, motivations, and strategies which guide decision-making. Study of CSOs' public documents has shown how terms like 'evidence' often convey certain stances in addition to producing new knowledge for its own value. Meanwhile, references to 'data' are typically connected with different sources or quantities. These usages tend to fit with dictionary definitions of the terms. However, this only reveals part of the story about the roles of research in achieving an organisation's mission. Why an organisation would use terms like 'data' and 'evidence'—or indeed an alternative—could be related either to how its members perceive the usefulness of such language, or to the extent to which it matches their values and existing practices. This could be facilitated by closer study of stated values like those seen in Table 1. Equally, as the

earlier qualitative materials reveals, different audiences with whom CSOs engage may move an organisation to use evidence in different ways:

> They [Members of Parliament] like things bound up to look professional and they like them to be rigorous and they like them to be independent. (Policy Director)

> [O]ur business is to be politically neutral to provide strategic support and information for those organisations to then go and do their job with it… we see ourselves as providing some more objective information than they [local authorities] would otherwise have. Which I suppose leads to more effective services and efficient services and there's a chain of benefits that results from that I think. (Policy and Research Manager)

> We've always been not for profit, so we're asset locked. And that's one element which the board has really stuck to, but it was easy to because I think the staff stuck to that anyway. It's quite staff led as an organisation. So we're limited by guarantee and call ourselves a social enterprise because it's cool at the moment to call yourself a social enterprise… And in many ways we are, but in other ways we're just a business that has social aims. (Director)

> [I]t is primarily about increasing public attitudes towards refugees. Part of our theory of change is that political negativity towards refugees and legislation impacts on refugees. Part of the creation of that is because that is desirable amongst the public and there is a lack of knowledge about refugees and migration generally. So part of it is to try to increase public understanding so that people have better attitudes and behaviours towards refugees and hopefully that will support and feed into better Government attitudes towards them. (Head of Policy and Communications)

The focus of this study was on documenting the ways in which 'evidence' and 'data' were used, rather than explaining the rationale for these usages. However, the qualitative data opens up interesting avenues for future exploration about the motivations that guide decisions about research update.

Understanding the opportunities and limitations associated with 'open data', particularly from official sources like the UK government, as well as 'Big Data' is an important objective for CSOs. When asked to identify other key issues or debates that are ongoing in UK civil society, some interviewees stressed that the changing nature and means of accessing data was particularly important. 'Open' data, or datasets that are publicly available, has the potential to make information available to CSOs that need it most, but it presents additional problems related to the available skillsets of researchers who are accessing them:

> Previously we've been limited to the UK data archive for example: you have to say where you are from. You have to say your affiliated institution and everything…But people will be able to get their hands on all kinds of things…It is something that they should be aware of and something that if we use it in the right way people can benefit from obviously, which will be brilliant but it is also defining the worries about it. (Programme Manager)

> I actually think you could have all the data open, it wouldn't mean necessarily that the data would be used. I think it's good to have open data and you'd have to think through what people need to actually use the data so, the two conversations, it's linked. What I was saying before about skills. I think actually not many people can use

> the open data that's available. You do have to have specialist skills… (Senior Research Officer)

How to develop skills to find, interpret, and use data—when it is seen as valuable for achieving an organisation's mission—appears to be an area that needs further attention.

Greater awareness of the needs, practices, and motivating factors facing academics and civil society or voluntary organisations has the potential to improve the ways researchers design and execute their work. At a broader level, this project has touched on the ways in which researchers interact with CSOs. One theme that was particularly salient was a perception that academics did not always consider either the time constraints upon CSOs or how the two sets of objectives did not always match:

> I would say you need to be a bit more fleet of foot and you need to understand the sort of pressures, both in terms of resources and time on NGOs. Academic papers are not our primary objective within doing this work. They are a helpful add on and add to the credibility of what we are doing but it's not the main focus and yeah, I think maybe also the experience of NGOs in terms of their ability to use the data and influence decision-makers with that data and so they can be strong partners and they can have access to particular client groups as well. (Head of Policy and Communications)

> Certainly for NGOs where you are, as soon as you put something out it is almost over and they are looking for the next thing and the next thing. So, you know, it is hard to produce quickly enough. (Policy Director)

Another theme centred on the ways in which academics collaborate with CSOs and voluntary orgnisations. Consciously building these organisations into the research design from the beginning of a project was crucial:

> …[I]t is important to think about what you want to get out of it at the end at the beginning and to think more about the dissemination and the impact of the dissemination when you are actually designing it. (Programme Manager)

> Community organisations want to see change; like they just want – if they are going to put time into something they want to – they say they want to see change but in reality they are often more satisfied with less because they are realistic. One paper isn't going to change the world but they do want to see things going out kind of publicly. (Policy Director)

> It would be about building a relationship, something should come out of that research, which also supports the organisation in terms of providing service or of giving them credit in their involvement in that research. (Lead Researcher)

Next Steps

This seed project was intended both as a standalone piece of research and as a means of building a platform for future questions. The main findings from this project will inform the next stages of the Principal Investigator's ongoing project funded by the Toyota Foundation, titled 'Big Data, Big Visions: Challenges and Opportunities for British Civil Society Engagement with Data-Driven Research'. This working paper has begun to identify some of these challenges and opportunities already, such as capacity, skills, and context. However,

more work remains to flesh these concepts out—as well as see how CSO issue area impacts the uptake of research data and evidence.

Impact

Since this seed project has just finished, it is anticipated that impact from the research will occur as results are shared during Autumn and Winter 2014. This research may also inform other knowledge exchange work that is already ongoing or under development at COMPAS.

Dissemination

Easton-Calabria, Evan and William Allen (Under Review). Developing Ethical Approaches to Data and Civil Society: From Availability to Accessibility. Submitted to *Innovation: The European Journal of Social Science Research.*

Allen, William (October 2014). Evidence and Data in British Civil Society: Results from a Pilot Study. (Centre on Migration, Policy, and Society Works-In-Progress Seminar Series: University of Oxford).

A more formal academic article from the corpus linguistic and qualitative interviewing is planned for submission in Autumn or Winter 2014.

Funding

Following the funding of this seed project, the author was successful as a named Co-Investigator on an Arts and Humanities Research Council (AHRC) project titled 'Seeing Data: Are Good Big Data Visualisations Possible?' along with Principal Investigator Dr Helen Kennedy of the University of Leeds, and Consultant Researcher Mr Andy Kirk of Visualising Data, Ltd. The 'Seeing Data' project begins from the premise that data of all sizes and types, but particularly large datasets, are increasingly being communicated through visualisations. However, little is known about how members of the public interact with, perceive, and make sense of these visualisations. That project is scheduled to run for 15 months from 1 January 2014 to 31 March 2015. More information can be found on the project website: www.seeingdata.org.

Acknowledgements

I would like to thank the civil society and voluntary organisation staff who generously shared their time to participate in this study. Their insight into the experiences, opportunities, challenges, and satisfaction derived from working with their respective teams to achieve their goals was invaluable. Also, many thanks to Evan Easton-Calabria who collected the corpus data. Finally, I would like to recognise the guidance and support of staff at the Communities and Culture Network+ throughout this seed project.

Works Cited

Bastow, Simon, Patrick Dunleavy, and Jane Tinkler (2014). The Impact of the Social Sciences: How Academics and Their Research Make a Difference, (London: SAGE).

Blinder, Scott and William Allen (2014). Constructing Immigrants: Portrayals of Migrant Groups in British Newspapers, 2010-2012. Centre on Migration, Policy, and Society: University of Oxford, WP-14-117.

Boswell, Christina (2009). The Political Uses of Expert Knowledge: Immigration Policy and Social Research, (Cambridge: Cambridge University Press).

Minocha, Akshay, Siva Reddy, and Adam Kilgarriff (2013). Feed Corpus: An Ever Growing Up-To-Date Corpus. Available at: http://sivareddy.in/papers/FeedCorpus_sigwac.pdf, Accessed 1 September 2014.

McEnery, Tony and Andrew Hardie (2012). Corpus Linguistics: Method, Theory and Practice, (Cambridge: Cambridge University Press).

Ross, Duncan. (2013). Big Data and Social Organisations: A State of the Art Review (Oxford: Nominet Trust).

Vicol, Dora-Olivia and William Allen (2014). Bulgarians and Romanians in the British National Press, 1 December 2012-1 December 2013. Migration Observatory Report, COMPAS: University of Oxford.