

Forthcoming in *Information, Communication & Society*

Title

Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility

Author:

Dr Aristeia Fotopoulou, University of Brighton

Word count: 7,191

Author contact details:

a.fotopoulou@brighton.ac.uk

University of Brighton, School of Media
Watts Building, Lewes Road, Brighton BN2 4GJ
UNITED KINGDOM

Abstract

This article develops a sociotechnical conceptualisation of data literacies in relation to citizens' data practices: highlighting the agentic, contextual, critical, and social aspects of data skills and competencies, it frames data literacies as both discursive and material. In order for civil society organisations to make sense of big, small, open and other data they need multiple skills, beyond the technical; it is therefore unhelpful to talk about a single form of data literacy. It is more helpful to consider how such literacies in the plural develop within the material social contexts of civic cultures, and how they can progress in tandem with critical awareness about the power aspects of data, so they can become central tenets of data advocacy. The primary purpose of the article is to move forward the debate around how to conceptualise data literacy – and to question how far the concept is useful in the first place. The article draws on empirical research and starts from the premise that it is imperative to develop frameworks and training schemes that enable civil society actors and publics more generally to use open data, to make data more relevant to stakeholders, and to support their engagement in policy debates around datafication.

1. Introduction

This article develops a sociotechnical conceptualisation of data literacies in relation to citizen data practices: highlighting the agentic, contextual, critical, and social aspects of data skills and competencies, it frames data literacies as both discursive and material. In order for civil society organisations to make sense of big, small, open and other data they need multiple skills, beyond the technical (Bhargava et al., 2015); it is therefore unhelpful to talk about a single form of data literacy. It is more helpful to consider how such *literacies* in the plural develop within the material social contexts of civic cultures, and how they can progress in tandem with critical awareness about the power aspects of data, so they can become central tenets of “data activism” (Milan & Gutiérrez 2017).

Civil society organisations today, such as NGOs and charities, produce a huge volume of data. Making sense of data, communicating in ways that are relevant to broad audiences, and using data for the social good require those actors, and citizens more generally, to acquire the skills to access, analyse and interpret them. This is important especially because government policies and decisions that rely heavily on big data analytics are often not informed by the views of their citizens and don't always reflect their needs. Projects aiming to enhance data analytics skills and competences usually focus on administrative and technical competences, and are intended for professionals and service providers (Frank, Walker, & Tygel, 2016). With few exceptions the emphasis of data literacy frameworks is disproportionately placed on technical literacy (for instance, - advancing coding skills, engaging citizens in hackathons, and launching University degrees in computing), and still disregards the need to address deeper structural issues of inequality. Making sense and meaning of data and big datasets, such as electoral data or health data, is not only a technical but also a sociocultural process; it is done by humans and it involves processes of interpretation and categorising (boyd & Crawford, 2013; Crawford 2013; Bowker, 2013; Dourish, 2016). Likewise, policy- and decision-making relies hugely on the values and understandings that get embedded in the interpretation of data. There is thus a pressing need for data literacies that move beyond a focus on enhancing quantitative analysis and technical skills. There are several projects that take a creative approach to enhancing data analysis skill. For example, *City Digits: Local Lotto*, a project piloted in Bushwick's School for Social Justice in New York City, allowed high school students to analyse and visualise the impact of a state-sponsored lottery in their low-income neighbourhoods (Oram, 2017). These projects demonstrate how the combination of creative media, storytelling and analytics allows participants to generate debates around specific issues that affect their communities (D'Ignazio 2017).

Definitions of data literacy come from a variety of scholarly fields and address important concerns in relation to open government, transparency and accountability. Often the scholarship on how data technologies can be used for civic participation concentrates on emerging *individual* practices and personal data literacies: for example, practices of appropriation (e.g. Mollen & Dhaenens, 2018); individual citizen practices that resist top-down data collection (Treré, 2018); the negotiation of

content created by algorithms (Ytre-Arne & Das, 2019); or critical understandings of personal data (Pangrazio & Selwyn, 2018). Similarly, McCosker (2017) understands analytics as commercial techniques that potentially furnish personal tactics in managing the social media self, and hence attempts to build a *personal* data literacy framework by drawing together common self-oriented metrics across dominant platforms. Other researchers concentrate on advancing data literacy from engagement with personal data but produced through social interactions and practices within hackathon settings, and offer insights on the human aspects of engaging with the technical objects of datafication (See Pybus, Coté, & Blanke, 2015). However, my approach here is different. My interest in this article is threefold: first, I am concerned with the skills and competences that allow civil society organisations and citizens to analyse and use open data from governmental and other databases about issues important for their communities, rather than personal data collection practices or personal data collected by social media platforms; second, I am primarily focusing on collective actors rather than individuals, and the ways in which the social aspects of citizens' data practices may shape critical data literacies; and third, I am interested in the role of data literacies in raising awareness about the ideological, ethical and power aspects of data. In this sense this article (and the associated research project) is better aligned with the aims of earlier work on social analytics, whereby organisations use “analytics not for the sake of measurement itself (or to make profit from measurement) but in order to fulfil better their social ends through an enhancement of their digital presence” (Couldry, Fotopoulou & Dickens, 2016, p. 19). It is also aligned with research that advances a data justice framework (Dencik, Hintz & Cable, 2016); in that it seeks to provide the resources for citizens to ultimately engage with the social and political debates of datafication in an informed way.

The primary purpose of this article is to move forward the debate around how to conceptualise data literacy – and to question how far the concept is useful in the first place – and it draws from empirical work to do so. It starts from the premise that it is imperative to develop frameworks and training schemes that enable civil society actors and publics more generally to use open data for advocacy, to make data more relevant and useful to stakeholders, and to support their engagement in policy debates around datafication. The empirical research entailed the development of a training instrument for critical data literacies, which was piloted in training workshops with seven civil society organisations in the South East of England. The term “civil society organisations” is used throughout the article to refer to progressive non-governmental, third sector non-for-profit organisations, charities and community groups. The study involved eight representatives of seven organisations: an organisation that provides older and young people with support in improving their quality of life, such as housing and advocacy; a mental health charity for LGBTQ people; a charity that supports survivors of child sexual abuse, sexual abuse and domestic violence; an organisation that supports young people and families across Sussex, providing accommodation, family work, health, advice, and education

services; a music charity specialising in grassroots music and talent development, social change and innovation in education; a foundation supporting people with learning disabilities by finding accommodation, organising activities and providing community support; and a charity that provides suicide prevention training¹.

Following the mapping of existing data literacy projects², the development of the critical data literacy instrument comprised two training workshops that used a combination of tool-based capacity-building and critical skills through storytelling, visualisation and analytics. It aimed to address the knowledge needs and skills of these actors, in order to make data accessible and to equip them with the resources necessary for addressing the critical and ethical questions that relate to datafication³.

Drawing from insights that emerged from this fieldwork and the response to data analytics training by participants, the article offers a framing of data literacies as agentic, contextual, critical, multiple, and inherently social. Skills and competencies that relate to the use of big data and open data, alongside symbolic meanings and interpretative acts, materials, technologies, interfaces, and users, are understood here to comprise “citizens’ data practices” (Fotopoulou, 2019). It is argued that *literacies* in the plural, rather than “literacy”, allows us to understand the multiplicity and interconnection of data literacy practices with other literacies, while the emphasis on criticality stresses the importance of raising awareness about the ideological and power aspects of data. At the same time, it is proposed that data literacies should be understood as “social literacies”, because of the real life material conditions within which learning takes place, and the social contexts within which data practices acquire their meaning. Thinking of data literacies as contextual allows us to take into account the prior experiences and institutional contexts within which actors operate, as well as the contexts of data collection and processing. Finally, it is suggested that data literacies should be understood as agentic, in that they require actors to exercise agency when selecting domain-specific datasets, and to reflect on the ethics of their own data practices and their duty of care.

In the next section the article begins by engaging with recent theoretical conceptualisations and debates around data literacy but also some older frameworks around critical media literacy. It moves on to outline how the critical data literacy training instrument was developed with citizens’ advocacy in mind, and proceeds to examine the main findings of the empirical research and the lessons from the implementation of the training with civil society organisations in the South East of England. Finally, it suggests the key components of a framework for developing data literacies and community use of civic data.

2. Conceptual and methodological background: Conceptualising critical data literacies as agentic, contextual, intersecting and social

In order to build up the case for understanding data literacies as plural, social, contextual and intersecting, this section problematizes the concept of “literacy” while also discussing some key issues emerging in current definitions of data literacy.

Literacy or expertise?

“Literacy” is a contested term and its use here merits clarification. It has been widely employed in the field of education to frame both critical media literacy and critical digital literacy (see Hammer, 2011; Hinrichsen & Coombs, 2013). These strands focus on the means necessary for individuals to understand how our everyday use of diverse media helps to construct our knowledge of the world and our position within it (Alvermann, Moon, Hagwood, & Hagood, 2018). In media literacy frameworks, the key practice in which audiences are engaged in is “reading”, understood as a sociocultural practice that is linked to membership in certain interpretative communities, with particular symbolic and material resources (Livingstone, 2004; Luke & Freebody, 1997). However, data systems lend themselves to a wide range of practices that cannot be limited to “reading”. What people do with data, the meanings that they make, and the competences that are necessary in order for them to engage with data in ways that empower them, are all aspects of contemporary everyday data practices (Fotopoulou, 2019). So, although we could extend concepts of critical media literacy to encompass algorithmic- and data-based media forms, data literacy adds distinctive dimensions to both media and information literacy.

This is also the case because data systems such as databases and algorithms are dynamic. Data have been understood to be part of assemblages (Kitchin & Lauriault, 2014), while the subject matter of data is very diverse; these factors constitute data science as fundamentally multidimensional (Iliadis & Russo, 2016). This aspect of diverse and multiple domains, alongside an understanding of data as being composed of multiple relations to context-specific assemblages (Jassanoff, 2017; Pickren, 2016; Kitchin & Lauriault, 2014), make it clear that a single definition of critical data literacy may not be possible; in fact, it is not necessarily desirable.

It should be noted that the term “literacy” has been deemed too basic and indeed counter-productive. In earlier work, we have found that “literacy” is the preferred term in policy framings from a variety of EU and UK actors; opting for the term “expertise”, we highlighted how ICT skills need to be developed in conjunction with other forms of literacy, such as health literacy (Bassett, Fotopoulou & Howland, 2015). Rethinking Winifred Nowotny’s term “socially distributed expertise”, Caroline Bassett (2015) poses important questions about “technological expertise in relation to Big Data, everyday life and critical practice” (548) rather than “literacy”. Bassett conceptualises big data expertise as a *redistribution of skills* (and hence power) in the contemporary trajectory towards automation.

Approaching data literacies

The question that this project addressed related to how lessons from both theoretical conceptualizations and empirical projects can be operationalized in relation to civil society organisations' use of data and data skills training. An important aspect of such operationalization is access to open data. This can be through open access to governmental databases, or data requests for specific data, and is a basic precondition of the right to know and hold governments accountable (alongside the right to privacy and the right to be forgotten). It is widely accepted that the more visible the workings of government are, the more accountable and legitimate it seems, and transparency is largely framed as a self-evident good (Fenster, 2015). However, as Fenster (2015) notes, in practice transparency often falls short of that promise because of the complexity of the state, government information, and the public. Other more hopeful approaches have underlined the potential of information visualisation to enhance parliamentary transparency and contribute to the empowerment of citizens, since it allows stakeholders to comprehend large amounts of data and to identify patterns in information (Papaloi & Gouscos, 2013). But, despite the active promotion of Open Government Data (OGD), access to these data has been limited (Lnenicka, 2015). What is more, it could be that the proliferation of open governance data, including the publishing of data and an increase in visualizations of government data, inhibits its effective and efficient use, and in fact causes information overload (Koltay, 2017). As discussed later, the material and social contingencies of governmental open data, manifested in the questions of how far transparency is possible and how far such data are used effectively to enhance citizen participation, have implications for developing a framework of critical data literacies.

An interesting agenda comes exactly from an effort to mitigate against information overload. Koltay (2017) instigates a social approach to information and data literacy, which is situated within the socio-cultural practices of a particular discipline. Focusing specifically on the need to improve data literacy in higher education, Van de Geel and colleagues (2017) examined the changes in educators' data literacy and data-based decision-making in schools during a two-year project. Their definition of data literacy highlights the importance of being able to interpret data and make it actionable in a way that is beneficial to students. Matthews (2016) also analysed data literacies from a community informatics point of view, and identified the need to shape these according to four different areas of implementation: research (academic focus), classroom (secondary educational focus), carpentry (practical training focus) and inclusion (community development focus). Following the data inclusion paradigm of *Data-Pop Alliance* (Bhargava et al., 2015), Matthews (2016) draws attention to the direct involvement of community members, and the empowerment effects that working with data may have for them. These approaches take into account the *contextual* elements of data-related literacies, which is something that this project also highlighted.

At its outset, the project adopted D'Ignazio and Bhargava's (2015) conceptualisation of critical data literacy and big data literacy. D'Ignazio and Bhargava (2015) propose the adoption of elements of Paulo Freire's literacy method (namely, his emancipatory pedagogy of Popular Education) in a pedagogical path towards developing data literacy, which they envision as a set of capabilities that enables people to produce and use data in a critical way. This set of capabilities entails:

- a) First, the ability to read data in a way that takes into account how this data was generated. This *data reading* dimension of critical data literacy includes an appreciation of the granularity of a dataset being read, as well as the methodological basis for the capture, measuring, purpose, subjective or contested nature, and ownership of the data.
- b) Second, an ability to use statistical tools, and associated technologies, to *process* data effectively. A data processing capability includes the linking of data to relevant other sources, as well as the capacity to process data with a specific and clearly defined objective.
- c) The third dimension of critical data literacy is defined as the capacity to effectively *communicate* data. This encompasses the ability to interrogate a dataset in order to find matches between different types of data, including time series, distributions, comparisons or networks, and to find relevant ways of displaying them, including text, maps, charts, or infographics. An important additional dimension of data communication is the ability to understand target audiences and to tailor communications so that the information being transmitted is in fact understood. This presupposes ethical considerations, including the accuracy of what is being communicated.
- d) The fourth dimension of critical data literacy identified is the capacity to effectively *produce* data. This capability includes an understanding of the data formats and data publishing technologies involved in producing data which is useful, open and accessible to non-experts.

It is clear, then, that D'Ignazio and Bhargava's (2015) framework involves both the advancement of technical skills and the transformative project that results from literacy. Although the project aims supported this hands-on approach, D'Ignazio and Bhargava's focus on capabilities is arguably limiting. As discussed in Section 4 of this article, it is rather helpful to take into account the social aspects of data literacies: the real life, institutional and communicative contexts that inform the data practices of civil society organisations, and, by extension, other social actors.

Similarly, another useful approach suggests "data infrastructure literacy" as an alternative term that pays attention to relations, rather than "things", and experimental public practices (Gray, Gerlitz & Bounegru 2018). Bringing science and technology studies terminology and foci into the debate, Gray and colleagues are concerned with the "shifting relations of databases, software, standards and classification systems, procedures, committees, processes, coordinates, user interface components and many other elements which are involved in the making

and use of data” (2018, p.2). Their suggestion of incorporating data infrastructure literacy (e.g. with complementary field trips and infrastructure ethnography) into learning statistical and technical skills is invaluable for academic programmes. However, digital literacy initiatives aimed at civil society organisations may depend upon the pragmatics of limited resources, and the time commitment that such complementary training requires away from case work. As is shown here, taking into consideration the institutional, economical and social contexts in which data skills training initiatives will apply, as well as the needs and priorities of the participants, is key to designing data literacy programmes.

Developing a critical data literacy training instrument for civil society organisations: Workshop structure and learning aims

The tool-based capacity-building training workshops aimed to introduce and advance the data analytics skills of civil society organisations. The workshops also became an opportunity for participants to voice concerns about the challenges that data-intensive media environments present for the sector more generally, and created a space for them to consider the ideological and political dimensions of datafication. This section first presents the data literacy instrument design and its inspiration, and then outlines the workshop structure and learning outcomes. Secondly, it examines how the processes of technical tool-based learning and data literacies acquisition intersected with the development of critical awareness during the workshops, by drawing on participant insights. These insights are instrumental in reshaping how we think about citizens’ data literacies, and inform the article’s discussion in Section 4 on how to conceptualise critical data literacies as social, plural, and context-specific.

The project ran two training workshops (called *Datahubs*), which were tool-based and aimed at capacity-building for civil society organisations. Seven representatives from seven organisations in the South East of England who participated in these workshops were introduced to various data analytics and data visualisation skills. Specifically:

Datahub Workshop 1. The first workshop enabled participants to: a) explore the relationship between data and knowledge; b) examine different types of data, including where they can be found and accessed (data portals and aggregators, e.g. data.gov.uk, Africaopendata.org); and c) understand how information can be used for advocacy and campaigning.

A subsequent session of the same workshop focused on telling stories with data. Participants explored how data has changed the way we tell stories, through leading examples from campaigning, journalism and the arts. Finally, the workshop included an introduction in creative data visualisation and data journalism.

Datahub Workshop 2. The second workshop was dedicated to equipping participants with the practical technical skills to work with data. The workshop covered a) the use of Google sheets to perform basic data cleansing operations; b) the exploration of a dataset using databasic.io's WTFcsv (a collection of basic web tools that introduce concepts of working with data); c) the manipulation of data in pivot tables in Google sheets; d) the development of narratives from data through the completion of creative briefs; and e) the visualisation of those narratives in the online infographics application infogr.am.

First participants were introduced to open data portals. These are catalogues of links to datasets and metadata run by a catalogue operator, which may be a citizen initiative or a government agency. They act as interfaces between government data and re-users of those data. As such their navigation, and an understanding of their scope and limitations, was important for the workshops. This took the form of a series of guided tutorials, with each participant using a computer to follow the facilitator.

Second, participants were guided through the process of manipulating a single dataset. The dataset that participants worked with contained open data on spending and expenses by Brighton and Hove City Council. This dataset was chosen because it crossed a number of domains, and so contained information that was likely to be of interest to a greater cross section of the participants.

3. Participant insights

For the majority of the seven organisations that took part in the project, the primary motivation for participating in the *Datahub* workshops was to learn how to use statistical data in order to raise (or justify) third-party funding. This focus on income acquisition is perhaps predictable given the current funding landscape in the UK. There are approximately 140,500 voluntary organisations in England and Wales, and nearly 165,800 in the UK (UK NCVO data for 2014/15); competition for funding in the South East is particularly harsh. What is more, in 2016, 39% of charities earned less than £10,000 per annum (Keen and Audickas 2017). Adding new skillsets to their organisations in order to survive in this context was thus essential for participants, who were primarily managers or communications officers.

Participants expected that basic literacy and competence in data analytics would allow them to understand and manipulate information as credible facts, giving them a better chance at obtaining funding. As one participant noted, "using data to get that funding or to get our message across is really important to us across a lot of diverse set of work" (Participant 1, 2018). Participants agreed that numerical data are widely understood as objective facts, and therefore reinstate legitimacy and trust in charities in relation to the cases of support for funding applications. However, as noted next, this instrumentalist understanding of data literacy shifted for participants during the

workshops, to eventually include the critical, ethical and political convolutions of selecting, analysing, interpreting and communicating data.

(a) Domain-specific data and data gaps

The introductory part of the data literacy training workshops was dedicated to themes of accessing databases and finding data, and to ways in which these data could be used for campaigning. Getting a sense of what datasets are available for them through weblinks to governmental and other data aggregators and portals was a first step for most participants. For others, however, this introductory approach proved too general, as their needs were quite distinct: because their primary audiences are policy- and decision-makers, civil society organisations require access to domain-specific data that are relevant and on which they can build a case for advocacy. As Participant 3 noted:

“There is big data out there, but I suppose it's a question of how useful is it, really? And how complete a picture we can ever have about the numbers?”

For example, a mental health charity working with LGBTQ people has in the past used statistical data to make a case about the need to focus on and mental health issues for people identifying as LGBTQ. As their spokesperson acknowledged:

“There's a lot more data now. However, there are still lots of gaps in them, and there are lots of local gaps. For instance, what about poverty in the LGBTQ community? And how does that affect people's lives and their mental health?” (Participant 3).

By highlighting the gaps in datasets, Participant 3 raised an important issue about the intersectional aspect of social inequalities, and a key challenge in using big data for advocacy and for social justice. Researchers have of course noted how, when it comes to vulnerable social groups, absences in data should be respected, as they indicate non-response; this is also a known limitation of big data (Dalton, Taylor & Thatcher 2016). However, beyond questioning that big data will ever be able to provide a comprehensive picture of complex social phenomena where social groups are disadvantaged in more than one way, the practical issue remains that data that are potentially of use for policy-oriented advocacy are highly domain-specific. Thus, data literacy framings need to take into account the distinct needs for access to domain-specific data according to the aims of various stakeholders.

(b) Data literacy as a new form of media and communication literacy

Reflecting on the training they received overall during the capacity building workshops, participants valued how data analytics and data visualisation could be

used in order to better communicate their organisation's aims and actions. Participants reflected particularly on the hands-on training of the storytelling sessions (where we talked about the process of making creative briefs and developing narratives from data), and the data visualisation sessions (where we used the online infographics application infogr.am to create simple data graphs). Many felt that learning how to do these tasks and use data effectively was significant, because it would allow them to communicate the work that they did, to educate audiences about the mission and priorities of their organisation, and potentially even get a specific campaign off the ground.

Although during the training sessions we mainly worked with open datasets of governmental organisations (focusing on Brighton and Hove Council expenses), participants were keen to explore citizen data generated internally by *their own* organisations at a future data literacy session, such as statistical data about the number of people contributing to specific campaigns, or evaluation data concerning the impact of certain actions. This is because they felt that having the skills and competences to present the benefits of their actions through the analysis and visualisation of data would enhance their relationship with their audiences and relevant stakeholders. However, participants recognised that there are limitations to what skills acquisition in data analytics can offer in terms of prevention and, subsequently, intervention in this context. For example, presenting data and specifics about their organisation's activity to the public is complex for those working with vulnerable groups. As a participant working in a suicide prevention organisation put it:

“We deal in a lot of preventative work, but how, again, how to measure from that? How do you come up with big data that is convincing about what you do now that may prevent [something from happening] later on?” (Participant 3).

Nonetheless, gaining the skills to use statistics and big data analytics for communication was particularly valuable for the civil society organisations that participated in the project. According to a participant:

“I think it's often about creating an emotional attachment to the statistical data so it's about saying, this is the bit you can relate to, this is the bit you identify with, but actually that is represented by a hundred different people or a thousand different people, and if you can understand their plight, you can understand the plight of everyone in that statistical data” (Participant 5).

The acquisition of data literacies was therefore seen to provide organisations with novel resources to empower the vulnerable communities they work for. Particularly, the importance of learning how to use data was understood to go beyond communicating complex numerical data to stakeholders; as is noted

next, it allowed civil society organisations to link the personal experiences of their service users to larger scale phenomena and social/political structures.

(c) Turning data into stories

Reflecting on the lessons from the data-based storytelling Datahub session, workshop participants were interested to discover more about advancing the emotional engagement of audiences with data. It was clear to them that such emotional engagement is a key dimension of effective communication for social change, but, as they acknowledged, data on their own don't inspire affective responses. As one participant put it:

“Engagement isn't going to be just rational. I mean, I can intellectualize information. I can intellectually engage with lots of things and walk away. If I'm not emotionally engaged to care, I'm not going to fight to change.”

Finding stories in data was therefore a process that participants were most interested in, as data stories become the link between the personal and the public, between an individual experience and a larger scale issue that affects a whole community. In the words of Participant 5:

“The personal story becomes about us all. This could be any of us, this is any of us, this is all of us, this is what they're telling us about, it involves all of us.” (Participant 5).

Learning how to tell stories with data was understood to be central in making sense of common experiences and connecting people.

Since participants already had the media literacy skills and competences required in order to create campaign material for their organisations through storytelling, the data literacy offered in the sessions added a new layer to their existing skills. Participants valued how learning to turn data into stories could potentially allow them to engage audiences emotionally by highlighting the scale of an issue, and, this way, to make a story more powerful.

(d) Learning about the fallacy of “raw data” and the importance of interpretation

During the course of the Datahub workshops participants learned from examples of both best and bad practice; they analysed several existing representations of data, including artistic visualisations and misleading graphs in the press from the new trend of data journalism. From this exercise participants became acutely aware that the same data may be used to tell different stories, and from diverse perspectives. In other words, they became alert to the fact that, although big data and smaller datasets and statistics are generally perceived as objective and factual, their representation is based on selection and is in fact subjective, sometimes even

manipulative. One participant posed some key questions in relation to the politics of representation:

“People are particularly keen to omit things that are liable to draw attention. How do you notice that? How do you produce decent infographics that aren't dishonest, but using limited data?” (Participant 2, 2018)

Participants reflected on their own storytelling practices for engaging audiences, which are currently non-data based, and contemplated how they would put their new skills of data-based storytelling and data visualisation into practice. They focused on the ethics of data visualisation and suggested that, despite now having access to a wealth of big data, and the skills to analyse them, not all stories need to be told. Some stories are simply too difficult or too painful for audiences to engage with (for example, suicide stories, as participants suggested). We discussed how important it therefore becomes to take into consideration the social, cultural and local contexts within which the interpretation of a story will take place, and participants learned the importance of tailoring data-based narratives according to their targeted audiences (for example the wider public, policy-makers, or industry partners).

To conclude the report of the empirical findings, it is noteworthy to observe the interesting course that the data literacy training took in this case: starting with the modest expectation of acquiring basic technical skills in data analytics, participants went on to make critical reflections on data visualisation and data journalism applications, and concluded with considerations about the shaping of ethically and socially responsible data practices. The themes that emerged during workshop discussions (finding a voice, linking the personal to a large-scale public issue, the emotional engagement of audiences through data stories, the ethics of data visualisation, data gaps and social justice) show how critical awareness was developed progressively and in parallel with developing certain technical competences. This critical awareness included a reflection on the discourses of data objectivity that circulate in the media, a realisation that data are more than just numbers, and a productive thinking process about how data literacy can shape citizen data practices within their respective institutional, material and cultural contexts.

4. Agency, context awareness and socially responsible data practices

As the previous section manifests, workshop participants were eager to gain expertise in using data and advance their technical capabilities. But the fact that workshop discussions kept returning to the themes of a) interpretative processes, b) the ethics of data visualisation and c) the fallacy of “raw” data, indicates something important. It supports the idea that data literacy initiatives for civil society organisations and citizens more generally need to also advance other kinds of

literacies, such as data ethics, critical thinking, media literacy and communication skills.

From the discussions that took place with participants of the workshops it becomes evident that civil society organisations are willing to take up training that includes data analytics, algorithmic interfaces and data visualisation in order to remain active in civic participation within changing sociotechnical environments that favour automation. However, there are three issues that constitute data literacy practices as intrinsically socio-technical.

First, the data practices of civil society organisations are contextual. The social contexts within which data practices acquire their meaning vary substantially. These contexts determine the material and symbolic conditions within which the production of a data visualisation or other data-based story takes place, as well as the different environments within which a data-story will be interpreted. Thinking about how civil society organisations are often already highly skilled in communication makes it important to note that, when thinking about data-specific literacies, we need to account for the complex institutional contexts within which such actors operate. As Van Geel and colleagues (2017) have also noted, it is important to take into account an individual's prior experience with digital media and public communication, their attitudes, and the way in which they make ethical decisions (all of which vary widely across social groups and generations) when targeting strategies towards specific groups of learners. Using Brian Street's (1995) term, we could say that data literacies are indeed "social literacies", because of the communicative practices and real social contexts upon which people draw.

Second, beyond basic access to online networks, actors need to exercise a certain degree of *agency* in deciding what data to pursue for specific advocacy campaigns or actions, and this involves data competencies and skills. In order for these actors to participate fully and to harness the benefits of big data (and smaller datasets) for their organisations and communities, they need access to domain-specific content that is tailored to their needs. But access to relevant data is not always straightforward and easy, as governmental datasets are open but often hard to access for non-specialists. As Gray and colleagues (2018) note, the conventions, norms and standards of open data infrastructures may offer possibilities to new, unintended publics. Hence "access", a term often used to refer to basic online access in schemes of digital inclusion, has a different meaning here. It presupposes the skills for accessing, searching and finding relevant data, and then cleaning data, all of which are highly laborious exercises which demand that actors exercise agency. Thus when developing instruments that address citizens' data skills, competencies and practices, we need to enable actors to exercise such agency. These aspects of agency complement other elements of big data literacy (D'Ignazio & Bhargava, 2015), namely awareness of when personal data are being collected, understanding the algorithmic processes that are applied to big data to identify

patterns in them, and appreciating the ethical implications of data-driven decision-making.

And, finally, civil society organisations that acquire critical data literacies and are able to use data analytics and data visualisations become the producers of communication in their own right, based on the active selection and manipulation of data, which is intended to effectively benefit their communities. Hence some complex, *meta* questions emerge concerning the critical literacy of target audiences (although in the more traditional understanding of “reading” texts). One question that arises, for example, is: how can the production of texts (such as stories and visualisations) that are based on the analysis of data be transparent, and should this always be a requirement? Researchers have noted that it is necessary to develop learner-centred tools and data biographies that contextualise data, by including descriptions of data collection methodologies and details about how the data have been managed and processed (D’Ignazio, 2017). Such processes add to the invisible labour of underfunded community organisations.

What is more, usually when thinking about audiences and producers of media, we tend to think of a top-down power imbalance between media industries and ordinary people (Livingstone, 2004; Das, 2017). But communications officers of civil society organisations who already have media and communication skills need to produce socially responsible media texts from data for audiences that are sometimes *more powerful* than them (e.g. funders, policy-makers).

Taking into account such institutional contexts (lack of resources) and power imbalances is important when instigating data literacy programmes and requests for transparency in the collection and analysis of data for civil society organisations, such as NGOs and charities. Working with community organisations towards learning data practices (such as data collection, cleaning, and management) therefore needs to be guided by an ethos of care: this involves paying attention to the experiences of vulnerable groups and principal ethical consideration in regards to handling private information (Fotopoulou 2019, Corple & Linabary 2018). An ethos of care applied to initiatives that aim at building data skills capacity means posing questions such as: How do stakeholders (and those who are closest to the social issues to be addressed with data science) benefit from the acquisition of data analysis skills? How may staff development activities that aim at data skills acquisition impact on the regular service provision offered by under-resourced organisations? And how does the proposed data literacy initiative help address problematic power relations? A care ethos in learning data practices can help organisations to be “attentive to both scientific rigour and the situated knowledge of those most affected by data-based policy decisions” (Zegura et al., 2018, p.2)

To conclude this section, a set of normative principles are proposed, which can be operationalised when developing data skills training instruments for civil society/

community organisations. These principles highlight the sociotechnical aspects of data literacies and foreground the centrality of agency, context awareness, and social responsibility in learning data practices.

1. Data literacies complement other types of literacies that are significant for civic engagement, including basic literacy, media literacy, critical thinking, data ethics, information and statistical skills.
2. The process of learning data practices needs to take into account the institutional, social, political and economic contexts and structures which inform the needs and priorities of each organisation (such as the state of funding, data governance and management), and the use of online monitoring system platforms.
3. Data skills learning schemes need to enable actors to exercise agency in relation to the collection, cleaning, and management of data, and to do so with care, especially when the data pertain to socially vulnerable groups or culturally sensitive experiences.

Conclusion

The focus of this article has been on advancing the literacies, competencies and skills of civil society organisations in relation to emerging dynamic data-based communicative environments. Starting from the premise that data systems are sociotechnical, the article examined some of the complexities of advancing citizens' data literacies beyond technical competencies, to enable agency, context awareness and social responsibility.

Piloting the data literacy instrument in workshops with seven organisations in the South East of England, it became evident that participants were inclined to go beyond the acquisition of basic technical skills in data analytics. As noted in the article, a variety of critical themes emerged during the workshops, such as: finding a voice and linking personal experience to large scale socio-political issues; emotional engagement of audiences through data stories; the implication of data gaps for social justice; and ethically and socially responsible data practices. The themes that emerged show how critical awareness about the ideological and power aspects of data can be developed gradually and in parallel to developing certain technical competences.

Stressing the real-life material conditions within which learning takes place, and the social contexts within which data practices acquire their meaning, the article argued that data literacies should be understood as “social literacies”. *Literacies* was here used to underline that data literacy practices are multiple and intersecting, linking critical media, information literacy and digital literacy as essential components for

citizen engagement. Moreover, framing data literacies as contextual, the article showed how data literacies can shape the data practices of civil society organisations within their respective institutional, material and symbolic contexts, and in turn should be informed by these contexts. This is particularly evident when we take into account the prior experiences and existing skills, as well as the underfunded institutional contexts within which actors operate, and the contexts of data collection and processing. Noting how there is a need for access to domain-specific content that is of relevance to communities and their needs, it was suggested that data literacies should be understood as highly agentic, in that they require actors to exercise agency when selecting such datasets, and to reflect on the ethics of their own data practices.

By conceptualising data literacies as agentic, contextual, critical, multiple, and inherently social, the article builds on but takes a different direction from projects that focus on capabilities, and suggests that schemes that attempt to implement citizens' data literacies should focus on participatory, real-life contexts of learning and doing.

Table (1)

Acknowledgements

The author would like to thank the research officer Peter Eyres for his help with the DataHub workshops, and the anonymous reviewers for their helpful comments. This project was undertaken with support from the University of Brighton's Research and Enterprise Initiatives.

References

- Alvermann, D. E., Moon, J. S., Hagwood, M. C., & Hagood, M. C. (2018). *Popular culture in the classroom: Teaching and researching critical media literacy*. Routledge.
- Bassett, C. (2015). Plenty as a response to austerity? Big data expertise, cultures and communities. *European Journal of Cultural Studies*, 18(4-5), pp.548-563.
- Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D. & Shoup, N. (2015). Beyond data literacy: reinventing community engagement and empowerment in the age of data. *Data-Pop Alliance White Paper Series. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Lad and Overseas Development Institute) and Internews*.
- boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), pp.662-679.
- Bowker, G. C. (2013). "Data Flakes: An Afterword to 'Raw Data' Is an Oxymoron." In Gitelman, L. *'Raw Data' Is an Oxymoron*, 167-171. Cambridge, MA: MIT Press.
- Corple, D. J., & Linabary, J. R. (2019). From data points to people: feminist situated ethics in online big data research. *International Journal of Social Research Methodology*, 1–14. <https://doi.org/10.1080/13645579.2019.1649832>
- Couldry, N., Fotopoulou, A., & Dickens, L. (2016). Real social analytics: A contribution towards a phenomenology of a digital world. *The British Journal of Sociology*, 67(1), pp.118-137.
- Crawford, K. (2013). "The hidden biases in big data," *HBR Blog Network* (1 April). Retrieved from <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/>
- Dalton, C.M., Taylor, L. & Thatcher, J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society*, 3(1), p.2053951716648346.
- Das, R. (2017). Audiences: A decade of transformations—Reflections from the CEDAR network on emerging directions in audience analysis. *Media, Culture & Society*, 39(8), pp.1257-1267.
- Dencik, L., Hintz, A. & Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*, 3(2), p.2053951716679678.
- D'Ignazio, C. & Bhargava, R. (2015, September). Approaches to Building Big Data Literacy. *Proceedings of the Bloomberg Data for Good Exchange Conference*.
- D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1) 6-18. DOI: 10.1075/idj.23.1.03dig
- Fenster, M. (2015). Transparency in search of a theory. *European Journal of Social Theory*, 18(2), pp.150-167.
- Fotopoulou, A. (2019). Understanding citizen data practices from a feminist perspective: embodiment and the ethics of care. In Stephansen, H., & Trere, E. (eds.) *Citizen Media and Practice*. Taylor & Francis/Routledge: Oxford.
- Frank, M., Walker, J., Attard, J., & Tygel, A. (2016). Data Literacy-What is it and how can we make it happen?. *The Journal of Community Informatics*, 12(3).

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*. <https://doi.org/10.1177/2053951718786316>

Hammer, R. (2011). Critical Media Literacy as Engaged Pedagogy. *E-Learning and Digital Media* 8(4), pp. 357–363. doi: 10.2304/elea.2011.8.4.357.

Hinrichsen, J., & Coombs, A. (2013). The five resources of critical digital literacy: a framework for curriculum integration. *Research in Learning Technology* 21. Available at: <https://journal.alt.ac.uk/index.php/rlt/article/view/1433>

Iliadis, A. & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2), p.2053951716674238.

Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society*, 4(2), p.2053951717724477.

Keen, R., & Audickas, L. (2017). Briefing Paper No SN05428. Charities and the voluntary sector. 16 August 2017. House of Commons Library

Kellner, D., & Share, J. (2005). Toward critical media literacy: Core concepts, debates, organizations, and policy. *Discourse: studies in the cultural politics of education*, 26(3), pp.369-386.

Kitchin, R. & Lauriault, T.P. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work.

Koltay, T. (2017). The bright side of information: ways of mitigating information overload. *Journal of Documentation*, 73(4).

Lnenicka, M. (2015). An in-depth analysis of open data portals as an emerging public e-service. *International Journal of Social, Behavioral, Educational, Economic and Management Engineering*, 9(2), pp.589-599.

Luke, A., & Freebody, P. (1997). Shaping the social practices of reading. *Constructing critical literacies: Teaching and learning textual practice*, 6, pp.460-475.

Matthews, P. (2016). Data literacy conceptions, community capabilities. *The Journal of Community Informatics*, 12(3), 47-56.

Milan, S., & Gutiérrez, M. (2017). Citizens' media meets big data: the emergence of data activism. *MEDIACIONES*.

Mollen, A., & Dhaenens, F. (2018). Audiences' Coping Practices with Intrusive Interfaces: Researching Audiences in Algorithmic, Datafied, Platform Societies. In *The Future of Audiences* (pp. 43-60). Palgrave Macmillan, Cham.

Oram, S. (2017). What can we learn from lotto spending?. *MIT News*, <http://news.mit.edu/2017/what-can-we-learn-from-lottery-spending-0126>

Pangrazio, L., & Selwyn, N. (2018). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, p.1461444818799523.

Papaloi, A., & Gouscos, D. (2013). Parliamentary Information Visualization as a Means for Legislative Transparency and Citizen Empowerment?. *JeDEM-eJournal of eDemocracy and Open Government*, 5(2), pp.174-186.

Pickren, G. (2016). 'The global assemblage of digital flow' Critical data studies and the infrastructures of computing. *Progress in Human Geography*, p.0309132516673241.

Pybus, J., Coté, M. & Blanke, T. (2015). Hacking the social life of big data. *Big Data & Society*, 2(2), p.2053951715616649.

- Street, B.V. (2014). *Social literacies: Critical approaches to literacy in development, ethnography and education*. Routledge.
- Treré, E. (2018). From digital activism to algorithmic resistance. In *The Routledge Companion to Media and Activism* (pp. 367-375). Routledge.
- Tygel, A.F., & Kirsch, R. (2016). Contributions of Paulo Freire for a Critical Data Literacy: a Popular Education Approach. *The Journal of Community Informatics*, 12(3).
- Van Geel, M., Keuning, T., Visscher, A., & Fox, J.P. (2017). Changes in educators' data literacy during a data-based decision making intervention. *Teaching and teacher education*, 64, pp.187-198.
- Ytre-Arne, B., & Das, R. (2019). An agenda in the interest of audiences: Facing the challenges of intrusive media technologies. *Television & New Media*, 20(2), 184-198.
- Zegura, E., DiSalvo, C., & Meng, A. (2018). Care and the Practice of Data Science for Social Good (pp. 1–9). Association for Computing Machinery (ACM).
<https://doi.org/10.1145/3209811.3209877>

¹ Participants were identified from a network of over 2,300 organisations and community groups with the aid of *Community Works*, a platform that connects the charity and voluntary sector. A total of 12 community groups were recruited, and 8 people took part in the workshops.

² There is a wealth of projects promoting data literacy, and they cover a variety of content while they address different target audiences. Although it is beyond the scope of this article to consider in detail what these projects do, they were categorised in one or more of three categories: a) educational tools and projects aimed at increasing critical data capabilities of real-world participants and/or online audiences; b) projects which connect stakeholders together to engender mutual learning and knowledge exchange; and c) repositories of information relevant to data literacy, signposting useful resources. Table (1) summarises such projects, including the data literacy topics that the projects address, the domains that are addressed by them (i.e. the empirical foci covered by the project components), and, where reflective literature exists, critical reflections on these projects.

³ The project deployed a combination of methods:

- a) Tool-based and capacity-building work consisting of the development of the data analytics training instrument and workshops with civil society organisations and community groups.
- b) Participant observation and a focus group interview with workshop participants.
- c) Desk research consisting of mapping existing projects in the field of data literacy.