



‘Personal data literacies’: A critical literacies approach to enhancing understandings of personal digital data

new media & society

1–19

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1461444818799523

journals.sagepub.com/home/nms**Luci Pangrazio** 

Deakin University, Australia

Neil Selwyn

Monash University, Australia

Abstract

The capacity to understand and control one's personal data is now a crucial part of living in contemporary society. In this sense, traditional concerns over supporting the development of 'digital literacy' are now being usurped by concerns over citizens' 'data literacies'. In contrast to recent data safety and data science approaches, this article argues for a more critical form of 'personal data literacies' where digital data are understood as socially situated and context dependent. Drawing on the critical literacies tradition, the article outlines a range of salient socio-technical understandings of personal data generation and processing. Specifically, the article proposes a framework of 'Personal Data Literacies' that distinguishes five significant domains: (1) Data Identification, (2) Data Understandings, (3) Data Reflexivity, (4) Data Uses, and (5) Data Tactics. The article concludes by outlining the implications of this framework for future education and research around the area of individuals' understandings of personal data.

Keywords

Critical literacies, digital platforms, personal data, Personal Data Literacies, personal data literacies framework

Corresponding author:

Luci Pangrazio, Faculty of Arts and Education, Deakin University, 221 Burwood Highway, Melbourne, VIC 3125, Australia.

Email: luci.pangrazio@deakin.edu.au

Introduction

The proliferation of personal digital devices and mass user platforms such as Google, Apple, Facebook and Amazon has given rise to unprecedented rates of data generation, collection and reuse. While digital technology users consciously volunteer masses of data on a daily basis, in many other instances data are collected without an individual's knowledge. This is particularly the case when it comes to personal data, which individuals often generate unconsciously and with little understanding of where, how or why the data are being collected. Thus, as digital data become more ubiquitous to everyday life, it is also becoming increasingly difficult for non-specialists to define and understand. Brunton and Nissenbaum (2015: 3) describe this as 'information asymmetry', where the 'data about us are collected in circumstances we may not understand, for purposes we may not understand, and are used in ways we may not understand'. In light of this tension, recognition is growing that individuals need to adopt more informed and critical stances towards how and why their data are being used. More specifically in terms of this article, it seems necessary to argue for the importance of 'Personal Data Literacies'.

The epistemological starting point of this article is that digital data are neither neutral nor independent of the thought systems that create, collect and aggregate them. As Kitchin (2014: 20) explains, 'no data are pre-analytic, or objective and independent'. Instead, it is acknowledged that personal data are socially constructed – meaning claims to their representativeness should be seen as partial and highly contestable. Moreover, neither does any piece of digital data persist in one fixed permanent form. Instead, any piece of digital data is subject to ongoing (re)interpretation, (re)use and (re)application. For example, digital data are routinely recombined with other data (such as Facebook's combining of user data with offline demographic databases before selling the composite data to third parties), meaning they have a recursivity that 'shapes as well as captures culture' (Beer and Burrows, 2013: 56).

This article responds to growing calls for a deliberately hermeneutic approach to understanding the various qualities and capabilities of digital data (Couldry, 2014). In particular, it sets out the basis for an appropriately nuanced 'Personal Data Literacies' framework. While digital data have already been defined and categorised in many ways, we argue that a 'critical literacies' framework provides a useful means of foregrounding the interests and needs of individual users – in particular, how individuals might better engage with and make use of the 'personal data' generated by their own digital practices. Applying a critical literacies approach leads to personal data being distinguished in terms of type and context – thereby providing a framework through which to understand and interpret their personal data. Developing understandings that are not only more technically accurate but also more attuned to the complex and evolving socio-technical processes and systems underpinning contemporary digital society is fundamental to practical efforts to support personal data education, as well as research into everyday uses of data-based technologies.

What are personal data?

Personal data are any piece of information that can identify or be identifiable to an individual. This is often referred to in legal terms as 'personally identifiable information'. Of

specific interest is digital forms of personal data which can be drawn from a wide range of software and hardware sources and take a variety of modes, including numbers, characters, symbols, images, electromagnetic waves, sensor information and sounds (Kitchin, 2014). Personal data are generated and collated for a wide range of reasons, from improving individual performance to safety and security purposes. In particular, a thriving data knowledge economy has developed in which data, often of a personal nature, have acquired considerable commercial value (Bodle, 2016). However, defining personal data is by no means straightforward. As Golumbia (2018: para 2) reasons, ‘personal data must be understood as a much larger and even more invasive class of information than the straightforward items we might think’. In developing a framework for personal data literacies, then, we contend that there are at least three distinct types of personal data that individuals should be aware of.

Data that users give to devices/systems

First are data that users give to devices and systems. This might include self-tracking information, social media data (including videos, pictures, texts and tweets), emails and videos. The proliferation of social media platforms such as Facebook, Instagram and Twitter has increased the potential to collect personal data that individuals consciously give to devices and systems. The practices enacted through these platforms are often expressive and emotive, resulting in what are seen as rich and detailed sources of personal information. Personal data can also be generated voluntarily through activities that take place in work and educational settings. Using a ‘learning management system’ in university, for example, involves students accessing and uploading learning resources, contributing to discussion forums and completing quizzes and assessment tasks. While these activities might facilitate learning they also generate personal data on each individual, which can be processed and analysed to predict and optimise future use of the system (Pardo and Siemens, 2014). Crucially, these forms of consciously volunteered data are linked to other forms of data that users are less likely to be aware of. For example, all the data points just described will have attached technical details and various other forms of metadata that represent ‘unstructured and finely granular information’ (Peacock, 2014: 2).

Data that devices/systems extract from users

Second are personal data that are extracted from users by devices and systems on behalf of others. These personal data practices are involuntary and include ‘surveillance, and harvesting of people’s device use, online searches and transactions by policing and security agencies, the Internet empires and the data mining industry, and the development of tools and software to produce, analyse, represent, and store big data sets’ (Lupton, 2017: 340). Crucially, these data are brought into existence through their collection, meaning that the company, organisation or institution that generates such personal data can claim control (Hearn, 2010). The individual whose actions trigger the generation of such extracted personal data often has the least control as companies, governments, researchers and scientists seek to process and reuse any collected information. Control over the

Application Programming Interface (API) of digital platforms ensures that technology companies have the greatest access to these personal data. This leaves individual users always at a deficit, playing catch up with data specialists (Galloway, 2014). If an individual has agreed to the 'terms and conditions' of a platform, then this is considered to be legal use of their personal data. However, whether terms and conditions agreements can cover all the possible ways in which data might be reused is questionable (see Andrejevic, 2014).

Data that devices/systems process on behalf of users

Third are the many ways in which personal data are processed into the form of more socially meaningful 'data entities' which provide information of relevance to people, places and institutions. Often, individual users will have little or no exposure to these data, as it is used to inform system processes and institutional procedures, often in the form of 'data doubles' designed to render the individual identifiable, knowable but also interpassive (Ruppert, 2011). Yet some forms of these data are fed back to individuals, albeit in partial form. For example, many social and/or consumer digital platforms are designed to aggregate and process data generated through platform participation and (re) present it back to individual users via dashboards, analytics pages and similar. Individuals can then use this information to participate differently within the context of the platform. Despite efforts to present data in a way that can be readily understood by users (i.e. visualisations, dashboards and profiles), identifying and interpreting processed data can be challenging (see Roberts et al., 2016). In part, this reflects the fact that any instance of data processing is directed by the motivations of those who own and control the data – thus shaping any likely interpretations and usefulness.

Current approaches to data literacies

The increased prevalence of these various types of personal data across multiple domains has given rise to different perspectives on the skills and competencies required to be 'data literate'. For example, academic debates within the natural and computational sciences tend to focus on the need for highly technical skills (Carlson et al., 2011), whereas information science tends to focus on the ability to locate and manage data (Calzada-Prado and Marzal, 2013). The field of media and communication perhaps offers a more nuanced approach to data literacies, where the focus is upon developing critical understandings of the language, audience and representations of data. These specific skills and approaches can be applied to particular features of digital platforms in useful ways. Thus, current approaches to digital data literacies can be seen as taking one of four forms.

Data safety and data management

An increasingly popular approach to data literacy is that of 'data safety' – rooted in the idea of protecting and/or selectively preventing the dissemination of personal data (Office of the eSafety Commissioner, 2017). This extends 'cybersafety' and 'Internet safety' approaches by focusing on the development of people's skills to manage and

control the data trails and traces they leave when using digital media – what is sometimes referred to as one’s ‘digital footprint’ (Roddell, 2006). Data safety approaches tend to focus on personal data that individuals volunteer consciously to devices and systems. This can be thought of as social data generated through using the features of connective media platforms, such as creating a personal profile or using social buttons to ‘Like’, ‘upvote’, rate and so on. Data safety is most often directed at children, young people and their families, adopting a didactic approach that seeks to enhance privacy and security practices. As with cybersafety, the ‘data safety’ discourse tends to be promoted most vigorously by advocacy groups and educational institutions, often in the form of websites and classroom-based programmes. These typically support the development of normative strategies that can be enacted within a device or system, such as adjusting security settings within the platform, or reading the terms and conditions attached to a particular platform.

Another popular data safety approach is using software that lets the user choose which companies and brands they share their personal data with. Rather than altering security setting within platforms, apps like *Citizenme* and *People.io* aim to provide users with some form of ‘control’ over all the personal data they generate. Such software seeks to ‘liberate’ personal data by making the artificial intelligence that can be gathered from such information available to individuals. Users are therefore offered personal insight into their digital identities and interactions. In addition, users can be remunerated (i.e. through PayPal, iTunes) for sharing their data with companies and brands of their choosing.

Data science

Elsewhere, is support for forms of data literacy based around reading, comprehending and analysing sets of open data. This form of data literacy is focused on supporting individuals’ capacity to work with big ‘open’ data sets that have been collected by governments and organisations for various purposes. Examples of the data science approach can be found in the World Bank’s *Data Boot Camp*, Software Carpentry’s *Data Carpentry* and the School of Data’s *Data Expeditions*. These are all based around ‘hands on’ pedagogical approaches – ranging from basic skills of interpretation through to complex engagement with geographic information (GI) analysis and structured introductions to the principles of software development. While understanding and analysing this data is commonly conflated with ‘data literacy’, it is more accurately a form of ‘data science’ in that it involves ‘statistics, or the systematic study of the organization, properties, and analysis of data and its role in inference’ (Dhar, 2013: 64).

These data science approaches to data literacy are typically directed towards professionals (such as journalists, civil society and civic coders) with little or no prior experience of computer programming. For example, the World Bank *Data Boot Camp* programme is a part of a wider scheme to ‘inspire and empower citizens to use Open Data and maximise value to the public in practical ways’. As such, these data science approaches to data literacy are often underpinned by a discourse of productivity, centred on developing skills to improve society and supposedly empower citizens. These outcomes are seen to arise from the kinds of data that data science most commonly works

with, that is, large sets of (meta)data collected by governments and organisations, and therefore relevant to broad societal issues. Despite the rhetoric of transparency promised through ‘open’ data science, some scholars argue that there are inherent contradictions in the notion of governments and other vested interests supporting meaningful forms of openness, open-access and open data (Halford et al., 2013; Langlois et al., 2015).

Data hacking

A third type of data literacy is what can be described as ‘data hacking’ approaches to accessing and repurposing data. While hacking is often popularly understood to involve breaking illegally into government or corporate systems, the hacker ethos within programmer cultures is more abstract in its objectives (Boyd, 2017). In this sense, the hacker ethos is more accurately seen as actively engaging with the ways in which the world is made up of codes or systems that can be hacked, from ‘programming, language, poetic language, math, or music, curves or colourings’ (Wark, 2004: 13). In this way, hacking can be thought of as a deep form of literacy focused on understanding the various systems and codes associated with society.

With high-profile hacks like WikiLeaks and the Snowden files prompting continued media attention, hacking is increasingly seen as a way of empowering individuals and redressing perceived power imbalances. As such, hackathons and collaborative computer programming events are now becoming regular activities for political groups and activists alongside programmer communities. The IT industry and other businesses are also making increased use of hacking approaches. As Kruger (2012) writes, hacking is no longer just for hackers, but is becoming a way for technology companies to recruit new staff or simply outsource their research and development. However, as a deep form of data literacy, hacking and hackathons typically appeal to those who already have technical skills, such as computer enthusiasts, programmers and software designers.

Media literacy approaches to personal data

Other approaches to developing data literacies are offered by scholars in the field of media and communication. While it is beyond the scope of this article to comprehensively review these efforts, they tend to follow two broad approaches. The first stems from a design approach and focuses on data competence and using data to engage and empower individuals in civic life. The second seeks to help individuals understand and manipulate data representations on social media platforms.

The first approach follows the Beyond Data Literacy (Data-Pop Alliance and Internews, 2015) definition of data literacies as ‘the desire and ability to constructively engage in society through and about data’ (p. iv). This identified four underpinning aspects of data literacies, including: data education, data visualisation, data modelling and data participation. This approach positions data literacies as a way of identifying and solving real-world issues, which are increasingly mediated through data. As Deahl (2014: 49) reasons: ‘Since the goal of data literacy is to help individuals learn to illuminate real-world phenomena through data, learning should be project-based, problem-driven, and culturally relevant’. A key skill within this strand of data literacies is what the Oceans of

Data Institute (Education Development Centre Inc., 2016) call ‘analytical thinking’ – that is, ‘taking a (sometimes) difficult problem, breaking it into pieces, and building it back up again to gain interesting insights’ (p. 8). This approach shares similarities with that of data science in that data literacies are equated with improving civic life. Nevertheless, while media education provides a series of practical strategies for developing data competence in citizens, there is nothing specifically digital and/or personal about the data being focused upon.

The second strand focuses more specifically on the role of data on social media platforms. In this approach, data literacies involve guiding individuals to identify, understand and manipulate representations of data to suit their needs. While rarely explicitly stated, this approach seeks to support users to critically unpack the social media logic (Van Dijck and Poell, 2013), and the norms, strategies, economies and dynamics which underpin social media practices. This strand of data literacies is predicated on the idea that data are ‘multivalent’ (Gerlitz, 2016) – that is, having value within multiple, sometimes conflicting, value regimes. For example, data have personal value to the user, to the platform operators who have pre-structured particular communicative acts into data, and other stakeholders who process and repurpose these data according to new value regimes. McCosker’s (2017) approach to data literacies focuses on the value of data to the user. Metrics and analytics, he argues, are the building blocks to data literacies – understanding and manipulating these can help individuals to make sense of data and the data publics that encompass them. This suggests that data literacy interventions focus on specific targets such as ‘identity’, ‘activity’, ‘interactivity’ and ‘visibility’.

Limitations in the current approaches to data literacies

While these approaches address significant aspects of data use, we argue that there is still a need to address many of the most important issues arising from the growing significance of personal data in contemporary society – in particular, making sense of the place of individual users within the data economy. Even when these approaches do wrest some power back from second parties (e.g. platform providers) and third parties (e.g. data brokers, advertisers), the burden of time and responsibility is shifted to the individual. For example, while recent software developments aim to make individual control of personal data more nuanced, these apps remain grounded in the idea of data having economic value.

In contrast, while data science and data hacking approaches might be thought of as supporting ‘deeper’ forms of data literacy, they also remain limited – especially in terms of the degree to which individuals are able to manage their personal data when much of it is generated and used in ways that they are not aware of. These forms of data literacy are typically underpinned either by discourses of productivity and/or political change and are therefore promoted by actors with particular goals or agendas in mind. Successfully engaging in these approaches also demands technical skills that non-specialists typically do not possess.

Similarly, the emerging data literacy approaches from within media and communications studies are clearly useful, as an immanent form of critique. Through more sophisticated understandings of metrics and analytics individuals can, as McCosker (2017: para 4)

points out, 'take control of the intentional and unintentional performance of self through self-oriented data'. However, as the platform mediates participation, not all stakeholders have the same degree of agency when it comes to social media data (Gerlitz, 2016). Through data, communicative acts are 'doubly articulated' (Langlois et al., 2009) so that as they are enacted on the platform, they simultaneously create new articulations at other points within the digital economy. Once counted within the digital economy, data are ascribed a value that is determined by a host of other stakeholders. In this way, data literacies that focus primarily on social media platforms rather than the broader data assemblage are limited in scope.

Notwithstanding these existing approaches, there is clearly a need to better support individuals to engage critically with their personal data, so they have a sense of understanding, control and agency within the data assemblage. Beer and Burrows (2013: 68) argue '[w]e know little about how the performativity of data circulation, the social life of data, feeds into the performance of subjectivity and the constitution of everyday experiences'. We contend that these are issues that require critical thought – particularly in light of recent imperatives raised by high-profile personal data incidents, such as the Edward Snowden NSA revelations and concerns over the role of data-organisations such as Cambridge Analytica in influencing the 2016 US presidential elections. Against this background, data literacies need to include critical understandings of the reconstitutions and recirculation of data – that is, what we would term 'Personal Data Literacies'. This not only involves technical skills and understandings, but should also include conceptualisations of the inherently political nature of the broader data assemblage. Personal data literacies should aim to build awareness of the social, political, economic and cultural implications of data, as well as cultivating the metaphorical 'space' to reflect critically on these processes.

A critical literacies approach towards personal data

Given the variety of existing perspectives on data, it is important to differentiate 'personal data literacies' from other approaches of working with and managing digital data. Here, then, we turn to what is often referred to as a New Literacy studies approach (Gee, 2000; Street, 1994) to theorise personal data as a part of the cultural and material practices of daily life. While more traditional approaches to literacy tend to focus on the measurement of skills and capabilities, New Literacy studies has grown in prominence over the past 30 years in offering a significant change to what 'counts' as literacy. From this perspective, then, everyday practices such as reading cereal boxes or deciphering train timetables, are meaning-making processes and therefore literacy events and practices (Heath, 1982). Thus literacies are understood as always socially and culturally situated, and are used for a range of vernacular activities, including enacting identities, achieving particular goals and facilitating social relations.

Informal uses of digital technologies have long been seen as a key element of the New Literacy studies approach. This was formally recognised by a slight change to the plural 'Literacies' in the title – that is, New Literacies Studies (Gee, 2015). Digital practices – or 'assemblages of actions involving tools associated with digital technologies' (Jones et al., 2015: 3) – are now acknowledged as an integral part of the everyday activities that

most people engage in. Although not widely discussed by New Literacies scholars, clearly these are practices that also generate personal data. In this sense, it can be helpful to see personal data as a 'text' in the New Literacies sense – that is, as 'a collection of semiotic elements that can function as a tool for people to take social action' (Jones et al., 2015: 5, emphasis added). As such, conceptualising personal data as a 'text' marks a point of departure from the models outlined earlier in which data were not seen as providing the opportunity for social action, but primarily as information to be managed. Rather than focusing on isolated skills for management and security, a literacies approach towards data as 'text' can be adapted to foreground 'personal data' as a social practice and tool for action. When considering the political economy of digital platforms, this approach becomes increasingly important. Many of the digital texts individuals engage with on a daily basis are commercially produced. In light of this, interpreting and decoding the role digital texts play in the digital economy should be a key element of what it means to be literate in contemporary society.

Following this line of thinking, our approach to personal data literacies seeks to cultivate a more critical disposition towards personal data. In this way, our model is aligned specially with the critical strands of New Literacies studies, as is evident in the work of Luke (2000, 2013), Green (1988) and Green and Beavis (2012). While critical literacy education has been criticised for its didactic approach (see Buckingham, 2003), like Luke, Green and Beavis we recognise the benefits of being able to unpack the politics of everyday texts and – most importantly – for these critical awarenesses and dispositions to be a part of what people do with digital media. In this respect, a useful guide is Luke's (2000) definition of critical literacy – highlighting three components relating to increasingly sophisticated and critical engagements with texts. First is 'metaknowledge' of 'meaning systems and the sociocultural contexts in which they are produced and embedded'. Second are the technical skills to negotiate these systems. Finally, Luke points to the 'capacity to understand how these systems and skills operate in the interests of power' (p. 72). Evident in this three-point definition is the need to understand the sociocultural context of everyday 'texts', as well as the ideologies that underpin them.

In a similar way, then, personal data literacies can be seen as having technical, social and ethical dimensions, as well as being both individually and collectively negotiated. In contrast to data safety and data science approaches, personal data literacies require critical reflexivity regarding the implications of data profiling and data recirculation. That said, this is not intended to constitute a normative ideal for personal data practices (see Mathieu, 2016). Instead, what we shall now go on to propose as a Personal Data Literacies framework is intended to support greater understanding of personal data – primarily to support the development of greater agency on the part of individuals so they might make informed decisions about their data practices. Following Couldry (2014: 891), our working definition of data agency is as an elongated process of action and reflection on the sociocultural context of a practice or text – allowing individuals to give 'an account of what one has done, even more basically, making sense of the world so as to act within it'.

Applying a critical New Literacies approach to personal data requires focusing on the ways in which data are constructed and interpreted by various specialists and social actors within data assemblages. These constructions influence the capacity that the (re) use of personal data have to not only produce knowledge, but also shape realities

(Bowker, 2013; Renzi and Langlois, 2015). Evelyn Ruppert describes these as ‘agencements’ or the ‘specific arrangements and technologies whose mediations and interactions not only enact populations but also produce subjects’ (Ruppert, 2011: 218). What makes digital data especially powerful is that these realities are often internalised by individuals in unconscious ways. Individuals are not always aware of the personal data they are giving to devices and systems, and the ways in which these data are then being used. The Personal Data Literacies framework aims to support the capacity of individuals to identify and analyse these processes and then devise uses and tactics in response.

As detailed below, the starting point of our Personal Data Literacies framework is that personal data need to be identified and interpreted in context. This marks an important difference from existing data science and data safety approaches to data literacy, which tend to focus on data management and production, rather than the generative and emergent meanings the individual brings to personal data. Indeed, a key characteristic of a critical literacies approach is an emphasis on context – that is, that personal data are collected under the assumption that they represent or indicate something about the wider phenomena from which they are drawn. This coincides with the view that context becomes a critical part of turning any form of personal data into information (Boyd and Crawford, 2012). Similarly, it aligns with Helen Nissenbaum’s (2009) argument that a concept such as data privacy needs to be framed in terms of ‘contextual integrity’. Here Nissenbaum contends that neither individuals nor platforms can expect an absolute ‘right’ to complete control or complete secrecy. Instead, what might be deemed appropriate is dependent on the context and norms within which the flows of data are situated. Following this approach, then, we also argue that individuals should be able to pursue different lines of inquiry and application with regard to their personal data – however, this can only be properly achieved with careful attention to context.

A critical framework of personal data literacies

In light of these previous discussions, we now sketch out the beginnings of what might be termed a critical framework of personal digital literacies. There are five domains to this framework: (1) Data Identification, (2) Data Understandings, (3) Data Reflexivity, (4) Data Uses, and (5) Data Tactics (see Table 1). Given the opaque nature of the data assemblage, it is important to stress that each domain of the framework is critical in orientation – identifying and understanding how and where personal data generated are in and of themselves critical acts. Nevertheless, the framework does represent increasingly advanced levels of critique, although these do not have to be engaged within a linear and sequential manner. For example, an individual trying to find a particular address might have to decide whether or not to turn on the geolocational settings on their mobile phone in order to find out where they are through the use of a ‘Maps’ application. Being data literate means understanding the implications of changing this setting, as well as having the technical skills to do so.

The first domain of the framework – Data Identification – is fundamental to developing any other domain of personal data literacy. To explain this further, we return to the three types of data outlined earlier: data that users give, data that devices and systems extract and data that are processed about users. Some practical examples of how these ‘types’ of personal data might be identified in relation the framework are as follows:

Table 1. Five domains of ‘Personal Data Literacies’.

Domain	Description/key questions	Action
Data Identification	What are personal data?	• Identification of personal data and their type (materialisation)
Data Understandings	What are the origins, circulations and uses of these different types of personal data?	• Identifying how and where personal data are generated and processed (data trails and traces) • Interpreting the information that are represented by processed data (data visualisations, charts and graphs)
Data Reflexivity	What are the implications of these different types of personal data for myself and others?	• Analysing and evaluating the profiling and predictions that are made from processed personal data (i.e. sentiment analysis, natural language processing) • Understanding the implications of managing, controlling and applying personal data (individual and collective critique)
Data Uses	How can I manage and make use of these different types of personal data?	• Applying, managing and controlling data • Building technical skills and interpretive competencies (reading the terms and conditions, adjusting privacy settings, blocking technologies, developing a shared language) • Applying the information that are represented by processed data (personal insights into digital self and performance)
Data Tactics	How can I do personal data differently?	• Employing tactics of resistance and obfuscation (tactics) • Repurposing data for personal and social reasons (creative applications)

- An individual gives data to a system when they upload a photo to Facebook or post on a discussion forum;
- A mobile phone extracts geolocational data from a user when their geolocational settings are switched on;
- Data are processed and fed back to users when they are presented with a calculation or evaluation, such as a social media user being informed about the ‘impact’ of their posts.

While it is important to be aware of each of these types of data, identifying or revealing the ways in which personal data are surreptitiously extracted from users without their knowledge is a critical act. So too, is identifying the ways in which analytics, dashboards and evaluations have origins in particular forms of personal data (while excluding others).

Once identified as a particular ‘type’ of personal data, individuals can begin to think about how and where this personal data was generated and the ways in which it is likely to be processed and used by other parties. This constitutes the second domain of the framework – Data Understandings. While it might be relatively straightforward to analyse the generation of personal data, it is difficult to accurately anticipate the ways in which data might be processed and used in the future. This part of the framework therefore involves the capacity for informed speculation based on an awareness of possible (re)uses of one’s data. There are some practical tools that can help with this. For example, data trails and traces can be examined by using tools such as Mozilla’s browser add on Lightbeam and the website IXmaps, which helps individuals to see how their data travel across the Internet. In regard to data that is fed back to users, this domain would involve understanding and interpreting the visual representations (e.g. dashboards, graphs, indicators) of the data, as well as understanding the source data underlying these representations.

Understanding how personal data are being used provides an impetus for the next domain of the framework, Data Reflexivity – that is, where individuals begin to analyse the implications of processing and reuse of their personal data. For example, with regard to social data that users ‘give’ to connective media platforms, individuals might think about how these data are used to profile or predict their behaviours or interactions. McCosker’s (2017) focus on metrics and analytics could be usefully employed in this domain of the framework. For example, a Facebook user might consider that their commenting on a ‘friend’s post influences what they will see on their Newsfeed and the kinds of data publics they create on the platform. They may also consider the implications of their actions on Facebook in terms of their subsequent experiences on other platforms (see Mathieu and Pavlíčková, 2017). Critique might occur individually or collectively. In regard to social data, it is increasingly argued that privacy is now best conceptualised as ‘networked’ (Marwick and Boyd, 2014), meaning collective negotiation of privacy settings and standards might be the best way to approach this domain.

Once an individual can critique the implications of their personal data, they are in a position to act. The fourth and fifth domains of the framework build a series of uses and tactics (Selwyn and Pangrazio, 2018) for working with data. Here, we draw on the ‘four resources’ model of literacy (Freebody and Luke, 1990; Luke and Freebody, 1999) to describe the stage at which users act upon the structures and social relations associated with personal data. More specifically, this involves ‘knowing about and acting on the different cultural and social functions that various texts perform’ in ways that reaffirm the ‘constructed order’ of a system (Luke and Freebody, 1999: 5). In terms of Data Uses, then, these might involve reading the ‘Terms of Service’ agreements, adjusting privacy settings, implementing ad block technologies or setting performance targets in order to influence the feed-back of data. Engaging in these strategies requires time and ongoing maintenance, however, understanding the implications that data processing can have on profiling and prediction might provide the impetus for such a commitment.

In contrast, many people’s everyday lives involve working around official structures by not relying on these official strategies, and resorting instead to political practices and forms of knowledge that subvert, usurp and undermine official norms and ‘ways of doing’. The notion of personal Data Tactics fits with a number of oppositional approaches

towards digital media that have taken various forms during the past 30 years. For example, oppositional behaviour, subversion and resistance have long underpinned the rise of 'tactical media' during the 1990s and 2000s (Garcia and Lovink, 1997), and more recent calls for 'data obfuscation'. As Brunton and Nissenbaum (2015) detail, these are tactics of 'data disobedience' intended to mitigate, evade or perhaps sabotage dominant structures of data reuse and recirculation.

Such data tactics might involve the deliberate use of false information to disrupt the connection between the personal data generated and the individual (i.e. entering an erroneous birthdate or gender identity) or setting up an independently assessed virtual private network (VPN) to protect privacy. Tactics could also involve more creative appropriations of personal data, designed to achieve more niche or specialised insights and benefits – for example, changing appearance to counter surveillance and facial recognition technology or repurposing personal data to create visualisations and representations for particular purposes. While these are somewhat uncommon practices, having the critical knowledge to understand why one might do this and the effect it would have, are indicative of a more sophisticated level of data literacies.

Future approaches to developing personal data literacies

A framework such as this is only of use if it supports and connects with other actions and interventions. There are two areas of obvious application – public education and academic research. First, while the approach to personal data literacies laid out in our framework is primarily focused on supporting individuals, any practical change in citizen understanding will often involve the support of others – most commonly, educators and advocacy groups, but also perhaps IT providers and government agencies. This raises an inevitable tension between focusing on individuals' literacy with regard to personal data and the need to respond to information asymmetries as a collective. Initiatives need to provide opportunities for both individuals and collectives to understand and respond to the challenges and opportunities presented through personal data. Developing these forms of meta-knowledge, technical skills and understanding of how power operates through personal data is a complex process, likely to require collective attention and the support of others. Second, this framework also has significant implications for academic research in the area of digital media use – pointing to a number of different questions and approaches of inquiry. As such, there are a number of ways that we envisage this framework being of use.

In terms of education, this framework has clear application for both formal and informal approaches to developing data literacies. To date, there have been few models to guide a formal curriculum based around the cultivation of data literacies. In fact, most approaches to developing general digital literacies have been based upon older literacy models, which have limited usefulness in current digital contexts (see Pangrazio, 2016). In light of this, the Personal Data Literacies framework would not only build specific data literacies, but also more general critical digital literacies, as understanding data flows inevitably requires knowledge of digital infrastructure. Such curricula are sorely missing from education, with the upper primary and secondary years marking an opportune time to begin cultivating these more critical understandings of the digital infrastructure (Selwyn and Pangrazio, 2018; Yan, 2005).

A formal curriculum approach to data literacies would also include a basic knowledge of the political economy of digital platforms, as the possibilities for commodification are a large part of the appeal of personal data. As such, the pedagogical approaches to domains 1–3 of the framework might be based around building knowledge and using online tools to track, trace and critique – all of which would be well suited to a school-based curricula and other non-formal educational providers such as libraries and youth groups. The latter parts of the framework, particularly ‘Data Uses’, could include opportunities to build collective understandings, especially when considering the challenges to privacy in a networked context.

The Personal Data Literacies framework might also articulate with informal educational initiatives more appropriate for adults. The popularity of public campaigns on personal data and privacy, such as ProPublica’s *Breaking the Blackbox* and Note to Self’s *The Privacy Paradox*,¹ suggest there is increasing interest in developing data literacies amongst the general population. These are also good examples of individual users collectively coming together to make sense of personal data and speaking back of the information asymmetries that currently exist. The Personal Data Literacies framework would complement such initiatives by providing a set of guidelines for people who seek to protect and work with their personal data. To this end, there are many online tools and functions within the settings of platforms that could be used as part of an informal education, such as working with Google’s *My Activity* section to identify what constitutes personal data (domain 1 of the framework), and *Ghostery*, which could be used as part of the Data Uses domain. The Personal Data Literacies framework might be used as a way to organise free online tools and resources and strengthen the learning that takes place through them.

In terms of academic research into data literacies, this framework also suggests a range of research questions that move beyond simply asking people what they do with personal data. Instead, Data Reflexivity, Data Uses and Data Tactics raise a range of significant questions to frame empirical research. These include:

- At what age can and should individuals start to learn about their personal data?
- In what ways do the practices encouraged by the framework limit individual’s experiences of the Internet?
- What forms of personal data tactics do people see as appropriate and/or attainable in terms of enhancing their engagement with digital media?
- To what extent (and with what outcomes) are people able to develop new data uses and/or tactics?

Research questions such as these fit well with the burgeoning trend of what Rogers (2013) terms ‘natively digital’ research – that is, social research based around the design, development and implementation of software and coded artefacts. This is illustrated in a handful of existing studies that have used software and coded artefacts to make ‘personal data’ and the machinations of the commercial data-brokering industry more accessible to ordinary users. For example, research by Skeggs and Yuill (2016) pursued this through combining information gleaned from Facebook’s own official API with their own bespoke developed ‘plug-in’ tool that was designed to extract data relating to participants’

Facebook activity that was otherwise not available from the API. This plug-in revealed advertising shown to users on Facebook, as well as indicating how Facebook was tracking users as they browsed elsewhere. This information was then used by participants to reflect on their engagements with Facebook – allowing the researchers to ‘open up the black box’ of the platform, if only temporarily.

Other studies have worked with young social media users to co-create software tools to gain a similar sense of their personal data generation and its reuse by third parties such as advertisers and data brokers. Pybus et al. (2015) augmented the development of their ‘MobileMiner’ plug-in with a series of hackathons to support participants’ understandings and critique of the personal data that was being generated and reused. Similarly, Pangrazio and Selwyn’s (2018) ‘PDQ’ smartphone app gave the teenage participants insight into how their personal data was processed by industry-standard data analytics and profiling tools (in the form of three commercial APIs concerned with object recognition, sentiment analysis and geolocation data). These insights were then used to test the appropriateness of various data uses and tactics. All of these existing studies point to the ways in which the domains of personal data literacy set out in our framework might be explored, provoked and supported by software-based empirical research.

Conclusion

Recent events, such as the 2018 Cambridge Analytica revelations, have confirmed what many critical data scholars have long described – that digital media users are tracked, traced and profiled through their everyday online (and offline) practices. There is now increased public awareness that many digital texts are commercially produced, and that most digital practices are datafied and commodified. Now more than ever, individuals need tools and resources to help them to continue to develop critical understandings of personal data and the data assemblage. In this article, we have argued that existing approaches to data literacy have been either too narrow or overblown in their ambition – therefore raising the likelihood of ordinary individuals missing out on the opportunity to develop the necessarily practical and interpretive skills. Instead, we have made a case for a hermeneutic approach to personal data, where the individual (rather than the corporation, institution or government) is foregrounded in the process of making sense of personal data. Given that it is the individual who generates personal data, their capacity to identify, interpret and act upon these should be the focus of efforts to develop data literacies.

As a starting point, we have laid out a framework for developing personal data literacies. In addressing such issues, we do not consider these ideas definitive or complete, and welcome them being tested, critiqued, improved and reworked by others. One of the obvious limitations to these domains of data literacy is the varied capacity of individuals to act on any understandings and reflections. At a basic level, then, we need to acknowledge the unequal agency that individuals and social groups have when engaging with digital data. On one hand, these different ‘data classes’ are ordered along lines of technical and statistical expertise – what Lev Manovich (2011) has described as a ‘data analysis divide’ between data experts and those without specialist computational training and skills. Yet, data inequalities also map onto existing power differentials and unequal social

relations. Indeed, as Ruppert (2013) points out, it is notable how the dominant ‘data-scapes’ of contemporary society tend to be tied closely with dominant ‘theories of social order’.

Thus, the extent to which it is possible to support the more widespread development of personal data understanding and agency across general populations requires further attention. The issues raised in this article certainly require continued thought, refinement and testing. In so doing, it makes sense to look back to pre-digital forms of critical literacy development, and use these established notions of ‘literacy’ as a basis for working out realistic ways of supporting the capacity of individual users to engage with seemingly imperceptible personal data infrastructures and data economies. While these are not easy issues to address, we hope that this article begins a productive line of dialogue, counter-commentary and action.

Acknowledgements

We would to thank the anonymous reviewers for their feedback on earlier drafts of the paper.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Note

1. A number of public awareness campaigns around digital data have been run in recent years by non-profit and civil society groups. The US non-profit investigative journalism organisation Propublica ran a high-profile project (*Breaking the Black Box*) aiming to support public understanding of Facebook personal data sharing. This involved the production of video and text reportage, alongside the use of custom-built online tools that allowed public members to analyse and share Facebook’s collection and profiling of their personal data. Similarly, the National Public Radio (NPR) programme ‘Note to Self’ ran a project (*The Privacy Paradox*) which also elicited public participation through a series of personal data ‘challenges’ designed to develop awareness and counter-actions in the sharing of personal data through social media platforms.

ORCID iD

Luci Pangrazio  <https://orcid.org/0000-0002-7346-1313>

References

- Andrejevic M (2014) The big data divide. *International Journal of Communication* 8: 1673–1689.
- Beer D and Burrows R (2013) Popular culture, digital archives and the new social life of data. *Theory, Culture & Society* 30(4): 47–71.
- Bodle R (2016) A critical theory of advertising as surveillance. In: Hamilton J, Bodle R and Korin E (eds) *Explorations in Critical Studies of Advertising*. Abingdon: Routledge, pp. 138–152.
- Bowler G (2013) Data flakes: an afterword to ‘raw data’ is an oxymoron. In: Gitelman L (ed.) *‘Raw Data’ is an Oxymoron*. Cambridge, MA: The MIT Press, pp. 167–173.
- Boyd D (2017) Hacking the attention economy. In: *Data and Society: Points*. Available at: <https://points.datasociety.net/hacking-the-attention-economy-9fa1daca7a37>

- Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.
- Brunton F and Nissenbaum H (2015) *Obfuscation: A User's Guide for Privacy and Protest*. Cambridge, MA: The MIT Press.
- Buckingham D (2003) *Media Education: Literacy, Learning and Contemporary Culture*. Cambridge, MA; Malden, MA: Polity Press.
- Calzada-Prado J and Marzal M (2013) Incorporating data literacy into information literacy programs: core competencies and contents. *Libri: International Journal of Libraries and Information Services* 63(2): 123–134.
- Carlson J, Michael FCC, Miller M, et al. (2011) Determining data information literacy needs: a study of students and research faculty. *Portal: Libraries and the Academy* 11(2): 629.
- Couldry N (2014) A necessary disenchantment: myth, agency and injustice in a digital world. *The Sociological Review* 62: 880–897.
- Data-Pop Alliance and Internews (2015) Beyond data literacy: reinventing community engagement and empowerment in the age of data (Data-Pop alliance white paper series). Working paper, Data-Pop Alliance and Internews, New York, September.
- Deahl E (2014) *Better the data you know*. MSc Thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Dhar V (2013) Data science and prediction. *Communications of the ACM* 56(12): 64–73.
- Education Development Centre Inc (2016) *Building global interest in data literacy*. Workshop report. Waltham, MA: Oceans of Data Institute. Available at: <http://oceansofdata.org/our-work/building-global-interest-data-literacy-dialogue-workshop-report>
- Office of the eSafety Commissioner (2017) *Protecting personal information*. Report. Canberra, ACT, Australia: Australian Government. Available at: <https://www.esafety.gov.au/esafety-information/esafety-issues/protecting-personal-information>
- Freebody P and Luke A (1990) Literacies programs: debates and demands in cultural context. *Prospect: Australian Journal of TESOL* 5(3): 7–16.
- Galloway A (2014) The cybernetic hypothesis. *Differences* 25(1): 107–131.
- Garcia D and Lovink G (1997) The ABC of tactical media. Nettime. Available at: <http://www.nettime.org/Lists-Archives/nettime-l-9705/msg00096.html>
- Gee J (2000) The new literacy studies: from 'socially situated' to the work of the social. In: Barton D (ed.) *Situated Literacies: Reading and Writing in Context*. London: Routledge, pp. 180–197.
- Gee J (2015) The New Literacy Studies. In: Rowsell J and Pahl K *The Routledge Handbook of Literacy Studies*. London: Routledge, pp. 35–48.
- Gerlitz C (2016) What counts? Reflections on the multivalence of social media data. *Digital Culture & Society* 2(2): 19–38.
- Golumbia D (2018) We don't know what 'personal data' means. Available at: <https://www.uncomputing.org/?p=1983>
- Green B (1988) Subject-specific literacy and school learning: a focus on writing. *Australian Journal of Education* 32(2): 156–179.
- Green B and Beavis C (2012) *Literacy in 3D: An Integrated Perspective in Theory and Practice*. Melbourne, VIC, Australia: ACER Press.
- Halford S, Pope C and Weal M (2013) Digital futures? Sociological challenges and opportunities in the emergent semantic web. *Sociology* 47(1): 173–189.
- Hearn A (2010) Structuring feeling: web 2.0, online ranking and rating, and the digital 'reputation' economy. *Ephemera: Theory & Politics in Organization* 10(3–4): 412–438.
- Heath SB (1982) What no bedtime story means: narrative skills at home and school. *Language in Society* 11(1): 49–76.

- Jones RH, Chik A and Hafner CA (2015) Introduction. In: Jones RH, Chik A and Hafner CA (eds) *Discourse and Digital Practices: Doing Discourse Analysis in the Digital Age*. Abingdon: Routledge, pp. 1–17.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: SAGE.
- Kruger A (2012) Hackathons aren't just for hacking. *Wired*, 6 June. Available at: <https://www.wired.com/2012/06/hackathons-arent-just-for-hacking/>
- Langlois G, Elmer G, McKelvey F, et al. (2009) Networked publics: the double articulation of code and politics on Facebook. *Canadian Journal of Communication* 34(3): 415–434.
- Langlois G, Redden J and Elmer G (2015) Introduction: Compromised data. In: Langlois G, Redden J and Elmer G (eds) *Compromised Data: From Social Media to Big Data*. New York; London: Bloomsbury, pp. 1–24.
- Luke A (2000) Critical literacy in Australia: a matter of context and standpoint. *Journal of Adolescent & Adult Literacy* 43(5): 448–461.
- Luke A (2013) Regrounding critical literacy. In: Hawkins MR (ed.) *Framing Languages and Literacies: Socially Situated Views and Perspectives*. New York; London: Routledge, pp. 136–148.
- Luke A and Freebody P (1999) Further notes on the four resources model. *Reading Online*. Available at: <https://pdfs.semanticscholar.org/a916/0ce3d5e75744de3d0ddacfaf6861fe928b9e.pdf>
- Lupton D (2017) Personal data practices in the age of lively data. In: Daniels J, Gregory K and McMillan Cottom T (eds) *Digital Sociologies*. Bristol: Policy Press, pp. 339–354.
- McCosker A (2017) Data literacies for the postdemographic social media self. *First Monday* 22(10). Available at: <http://firstmonday.org/ojs/index.php/fm/article/view/7307/6550> (accessed 2 October 2017).
- Manovich L (2011) Trending: the promises and the challenges of big social data. In: Gold M (ed.) *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press, pp. 460–475.
- Marwick A and Boyd D (2014) Networked privacy: how teenagers negotiate context in social media. *New Media & Society* 16(7): 1051–1067.
- Mathieu D (2016) Users' encounter with normative discourses on Facebook: a three-pronged analysis of user agency as power structure, nexus, and reception. *Social Media + Society* 2(4): 1–11.
- Mathieu D and Pavlíčková T (2017) Cross-media within the Facebook newsfeed: the role of the reader in cross-media uses. *Convergence: The International Journal of Research into New Media Technologies* 23(4): 425–438.
- Nissenbaum H (2009) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.
- Pangrazio L (2016) Reconceptualising critical digital literacy. *Discourse: Studies in the Cultural Politics of Education* 37(2): 163–174.
- Pangrazio L and Selwyn N (2018) 'It's not like it's life or death or whatever': young people's understandings of social media data. *Social Media + Society* 4(3): 1–9.
- Pardo A and Siemens G (2014) Ethical and privacy principles for learning analytics. *British Journal of Educational Technology* 45(3): 438–450.
- Peacock SE (2014) How web tracking changes user agency in the age of big data: the used user. *Big Data & Society* 1(2): 1–11.
- Pybus J, Coté M and Blanke T (2015) Hacking the social life of big data. *Big Data & Society* 2(2): 1–12.

- Renzi A and Langlois G (2015) Data activism. In: Langlois G, Redden J and Elmer G (eds) *Compromised Data: From Social Media to Big Data*. London: Bloomsbury, pp. 202–225.
- Roberts LR, Howell JA, Seaman K, et al. (2016) Student attitudes toward learning analytics in higher education. *Frontiers in Psychology* 7(1959). DOI: 10.3389/fpsyg.2016.01959.
- Roddel V (2006) *Internet Safety Parents' Guide*. Morrisville, NC: LuLu Press.
- Rogers R (2013) *Digital Methods*. Cambridge, MA: The MIT Press.
- Ruppert E (2011) Population objects: interpassive subjects. *Sociology* 45: 218–233.
- Ruppert E (2013) Rethinking empirical social sciences. *Dialogues in Human Geography* 3(3): 268–273.
- Selwyn N and Pangrazio L (2018) Doing data differently? Developing personal data tactics and strategies amongst young mobile media users *Big Data & Society* 5(1). DOI: 10.1177/2053951718765021.
- Skeggs B and Yuill S (2016) The methodology of a multi-model project examining how Facebook infrastructures social relations. *Information, Communication & Society* 19(10): 1356–1372.
- Street B (1994) The new literacy studies: implications for education and pedagogy. *Changing English* 1(1): 113–126.
- Van Dijck J and Poell T (2013) Understanding social media logic. *Media and Communication* 1(1): 2–14.
- Wark M (2004) *A Hacker Manifesto*. Harvard, MA: Harvard University Press.
- Yan Z (2005) Age differences in children's understanding of the complexity of the internet. *Journal of Applied Developmental Psychology* 26: 385–396.

Author biographies

Luci Pangrazio is a postdoctoral research fellow at the Centre for Research for Educational Impact (REDI) at Deakin University, Melbourne. Her research focuses on young people's practices and understandings of personal data. Her recent book is "Young People's Literacies in the Digital Age: Continuities, Conflicts and Contradictions" (Routledge, 2018).

Neil Selwyn is a professor in the Faculty of Education at Monash University, Melbourne. His current research focuses on issues of digital labor and data-based practices in education. Recent books include "What is Digital Sociology?" (Polity 2018) and "Everyday Schooling in the Digital Age" (Routledge 2017).