

Special Issue on Data Literacy: Editorial

# Data Literacy - What is it and how can we make it happen?

---

**Mark Frank**

University of Southampton, United Kingdom  
Corresponding Author.  
[mark.t.frank@gmail.com](mailto:mark.t.frank@gmail.com)

**Johanna Walker**

University of Southampton, United Kingdom  
[J.C.Walker@soton.ac.uk](mailto:J.C.Walker@soton.ac.uk)

**Julie Attard**

University of Bonn, Germany  
[attard@iai.uni-bonn.de](mailto:attard@iai.uni-bonn.de)

**Alan Tygel**

Federal University of Rio de Janeiro, Brazil  
[alantygel@ppgi.ufrj.br](mailto:alantygel@ppgi.ufrj.br)

---

The preceding special issue of this journal highlighted the enormous potential of the growing open data movement for social change and sustainable development (Sharif and Van Schalkwyk 2016). Others have emphasised its potential for economic development (Iemma 2012; Stott 2014; Vickery 2011), transparency and accountability (Geiger and von Lucke 2012; Janssen 2011, 2012; Meijer, Curtin, and Hillebrandt 2012), and political engagement (Baack 2015; Meng 2014; Noveck 2009). However, in many respects the movement has so far failed to fulfil its initial promise. Take-up has been disappointing (Peled and Nahon 2015; Worthy 2015), open data portals have not been maintained after promising beginnings (World Wide Web Foundation 2015), and there are increasing concerns about inequality arising from unequal access to data (Davies and Bawa 2012; Gurstein 2011). Key to these problems is the difficulty that the majority of people have in finding, understanding, manipulating and using data. In its early stages the open data movement was driven by the objective of ‘getting the data out there’ in the right technical format; little attention was paid to who would use the data, how they would use it and what support they would need to do so. The key issue was perceived to be overcoming resistance to publishing data; it was widely assumed that once the data was available people would respond as they did to the World Wide Web, creating new forms and concepts in unpredictable but productive ways. In practice this has only happened to a limited extent. The concept of the ‘armchair auditor’, a citizen who browses government

---

Frank, M., Walker, J., Attard, J., Tygel, A. (2016). Data Literacy: what is it and how can we make it happen? Editorial. *The Journal of Community Informatics*, 12(3), 4-8.

Date submitted: 2015-09-16. Date accepted: 2015-11-19.

Copyright (C), 2016 (the authors as stated). Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5. Available at: [www.ci-journal.net/index.php/ciej/article/view/1347](http://www.ci-journal.net/index.php/ciej/article/view/1347)

data to monitor its activities and hold it to account, has not materialised. Most people still rely on a small number of intermediaries such as specialist applications, data journalists, pressure groups and political parties to select and interpret data on their behalf. As a result, there is a growing focus on the user and what they require to be able to take advantage of data. This ability is increasingly identified as data literacy. As such this issue can be seen as a natural partner to its predecessor.

Data literacy is a recent addition to a growing band of literacies such as numerical literacy, statistical literacy and IT literacy. All of them refer to the ability to make use of a widely available medium or technology that is considered to be of fundamental importance. And they all, of course, draw an analogy with literacy as the ability to read, i.e. understand and use text. Data literacy refers to the ability to understand and use data, particularly in the context of the Internet. As a research topic it has, until recently, been largely confined to the skills that students and researchers need to use data. This is changing, largely because of the rapidly increasing profile of open data. However, the significance of data literacy is not confined to open data or indeed the Web. Data has played an expanding role in the lives of more and more people since the industrial revolution, and as a basis for many products and services data becomes increasingly a commodity in our information society. Businesses and governments have adopted scientific approaches to decision making based on data (Porter 1996) and democratic governments have accepted that citizens have a right to be informed about matters that affect them – including central and local government but also other public and private institutions. Census results, company accounts and trade statistics are all examples of data which are intended for public use and which existed for many decades before the internet. Other examples of this ‘datafication’ (Cukier and Mayer-Schoenberger 2013) include government data portals, reviews, feedback, and product suggestion on e-commerce websites, weather emergencies forecast, patient monitoring, citizen participation and decision-making, etc. The ability to understand and use such data is important for personal decisions such as choosing schools and investing in companies, and is also a plank of effective democracy – widely regarded as providing the transparency which is a prerequisite for accountability (Heald 2006).

But by itself data is not information. For data to be useful people must be able to extract information from it. The ability to do this is rapidly becoming a requirement to participate in modern life – as fundamental as the ability to use a telephone or money. Those who do not have this ability are in an important sense disadvantaged. This used to be the subject of introductions to statistics but the Internet has changed this. It has transformed both the opportunities and the challenges. Prior to the Internet there were limited sources of data available – usually from credible sources, but hard to access and understand. Most people lacked the resources to use such sources directly and relied on intermediaries such as the press to access and interpret data for them. The biggest challenges for most people were to understand and critically assess the ways that intermediaries presented data such as tables, graphs and charts. This was commonly understood as statistical literacy (see Woolff et al.’s comparison of data literacy and statistical literacy in this issue) and it is still a core concern of data literacy (see Zubiaga et al in this issue). The Internet has fundamentally changed the game by potentially allowing anyone with Internet access to access a vast range of data sources. To take advantage of this, in addition to statistical literacy, people have to find data, select it from a mass of alternative sources, evaluate its quality and trustworthiness and manipulate it to extract the information they need.

These challenges raises fundamental questions about the definition of data literacy. While authors have frequently offered their own definitions, typically this has not been based on any kind of systematic analysis. Three of the papers in this special issue address this. Woolf et al. , in their paper on Urban Data in the primary classroom, examine different ways of teaching data literacy in schools and discuss what this implies for a definition of data literacy, compare data literacy with the “more coherently defined statistical literacy”, and produce a comprehensive and rigorous definition of their own. David Crusoe takes a different approach, starting with the population of data users and what data literacy means for them. His definition is intentionally broad and less detailed than Woolf et al’s with its emphasis on teaching. Both Woolf and Crusoe define data literacy in terms of cognitive skills such as collecting, selecting, cleaning, analysing, interpreting, critiquing, visualising and sharing. It is possible to take an even broader approach. Paul Matthews teases out four different concepts of data literacy found in the literature and argues for a capabilities approach to all four concepts including social capabilities. Competence includes affective as well as cognitive considerations (Bloom et al. 1956). To be able to take advantage of data a user needs attributes such as confidence and belief in the value of data as a source of information, possibly tempered with an appropriate level of scepticism. In this issue Kayser-Bril describes the importance of the right incentives for data literacy amongst journalists. The scope of data literacy need not be limited to the personal attributes of the user. It may also include the way data is made available and the support provided to him or her. To pursue the analogy with textual literacy - we are all illiterate when the text is in a language we don’t know and we have no dictionary. Data in a familiar format such as CSV may be more accessible and usable than a potentially more powerful but less familiar format such as RDF (Frank and Walker 2016). Metadata that clearly describes the provenance of the data allows a user to put it in context and know to what extent they can trust it. Extending the concept even further, it may be useful to think of data literacy as a property of a community as opposed to an individual, with members of the community making different contributions. So that the presence of some people who can find data, some people who can manipulate it, and some who can present the result might constitute data literacy for that community. In this issue Prado et al, Bhargava et al, and Tygel and Kirsch all explore the importance of learning to use data as a community and in a social context.

Although defining data literacy increases rigor and helps to clarify its scope it also important to be flexible and not let discussion of definitions inhibit or constrain research. Any attempt to define a term in its formative stage is as much prescriptive as descriptive – it is an implicit recommendation to include certain things and exclude others. Woolf et al.’s paper on creating an understanding of data literacy critiques the lack of comprehensiveness of each approach they study but a comprehensive approach may not be possible or desirable and it may be more productive to accept different definitions according to the context. Instead of assuming that data literacy is a coherent whole we may need to consider several different kinds of data literacy for different situations e.g. data producers, data specialists and non-specialist users. Crusoe and Woolf et al. also consider the ethical implications of data literacy; that we should all be aware of the responsibilities that come with the rights to use data. It is not enough to know how to combine data sets, there is also a requirement to be aware of the effects of combining or using them; no one wants to inadvertently breach the privacy of others.

However it is defined, there has been increasing activity aimed at raising data literacy in recent years. A key part of this activity is education and training. Some element of basic

statistics has been part of school curricula in many countries for decades. The challenge is to bridge the gap from statistical literacy to data literacy and to help children relate the skills to the wider context. This is a topic of continuing research at the Open University in the UK. Woolf et al. describe the success of narrative and inquiry based learning in the UK to help primary school age children become comfortable with using data. But education is not limited to children (Crusoe questions whether curricular constraints mean that schools are the right place for data literacy training at all). For example, the School of Data is an international network of individuals and organisations round the world specialising in increasing data literacy through education and support. It interprets data literacy quite broadly in the sense of including skills for specialist roles such as journalists as well as citizens. The Open Data Institute (ODI) has also included raising data literacy as part of its international programme of activities. In their notes from the field Argast et al. describe the data literacy activities of the Canadian branch of the ODI including open data jams, hackathons and workshops in public libraries. However, there is a limit to what can be achieved in the classroom or workshop. One way out of this is through the use of technology. D'Ignazio et al. describe three web-based tools for assisting data literacy. These approaches are technology based. Others are attempting to increase data literacy using broader, more contextual approaches. Tygel et al. propose an approach to data literacy that is theoretically underpinned by Paulo Freire's work on textual literacy. This stresses the importance of embedding learning in a social context that means something to the learners, and learning by doing. Bhargava et al. describe the use of data murals in communities in Brazil to help people understand and "buy-in" to the use of data. Their work suggests that the arts and more visual communication methods can be a vital entry point to developing data literacy. Prado et al. place data literacy in the context of digital inclusion. They examine how marginalized populations in Brazil perceive the practice of digital literacy that will allow us to better understand the factors that affect the sustainability of initiatives that promote universal access and digital inclusion. These three papers all come from Brazil, raising the exciting possibility that data literacy will be the first 'literacy' whose dominant model comes from the global south.

For centuries textual literacy has been a requirement for an individual to take a full part in society. The opportunities in life are far greater for the literate than they are for the illiterate. A high level of literacy is a necessity for any society to develop. With the advent of the Internet data literacy promises to take on a similar significance. Although the concept is at a very early stage our experience is that whenever it is discussed there a strong interest and a wide acceptance of its importance. When it appears as a topic in events such as open data camps or academic conferences it has always been among the best attended (see Frank and Walker in this issue). The papers in this issue demonstrate there are major challenges in both defining what it should be and raising the level of data literacy round the world. But it also presents an opportunity for enabling the Internet to fulfill its potential as an instrument of constructive social change.

## References

- Baack, S. 2015. "Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism." *Big Data & Society* 2(2).
- Bloom, B.S. et al. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York, New York, USA: David McKay.

- Cukier, Neil, and Viktor Mayer-Schoenberger. 2013. "The Rise of Big Data." *Foreign Affairs*.
- Davies, Tim G., and Zainab Ashraf Bawa. 2012. "The Promises and Perils of Open Government Data (OGD)." *The Journal of Community Informatics* 8(2).
- Frank, Mark, and Johanna Walker. 2016. "User Centred Methods for Measuring the Value of Open Data." *The Journal of Community Informatics* 12(2).
- Geiger, Christian P, and Jörn von Lucke. 2012. "Open Government and (Linked)(Open)(Government) (Data)." *eJournal of eDemocracy & Open Government* 4(2).
- Gurstein, M. 2011. "Open Data: Empowering the Empowered or Effective Data Use for Everyone?" *First Monday* 16(2): 7.
- Heald, David. 2006. "Transparency as an Instrumental Value." In *Transparency : The Key to Better Governance*, eds. Christopher Hood and David Heald. Oxford: OUP/British Academy. incollection.
- Iemma, Raimondo. 2012. *Open Government Data: A Focus on Key Economic and Organizational Drivers*. Rochester, NY. <http://papers.ssrn.com/abstract=2262943> (November 9, 2014).
- Janssen, Katleen. 2011. "The Influence of the PSI Directive on Open Government Data: An Overview of Recent Developments." *Government Information Quarterly* 28(4): 446–56.
- . 2012. *Open Government Data: Right to Information 2.0 or Its Rollback Version?* Rochester, NY. <http://papers.ssrn.com/abstract=2152566> (November 9, 2014).
- Meijer, A J., D. Curtin, and M. Hillebrandt. 2012. "Open Government: Connecting Vision and Voice." *International Review of Administrative Sciences* 78(1): 10–29.
- Meng, Amanda. 2014. "Investigating the Roots of Open Data's Social Impact." *JeDEM - eJournal of eDemocracy and Open Government* 6(1): 1–13.
- Noveck, Beth Simone. 2009. *Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful*. Washington, D.C.: Brookings Institution Press.
- Peled, Alon, and Karine Nahon. 2015. "Data Ships : An Empirical Examination of Open ( Closed ) Government Data." In *Proceedings of the 48th Annual Hawaii International Conference on System Sciences*, Waikoloa, HI.
- Porter, Theodore M. 1996. *Trust in Numbers : The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J.: Princeton University Press.
- Sharif, Raed M, and Francois Van Schalkwyk. 2016. "Special Issue on Open Data for Social Change and Sustainable Development." *The Journal of Community Informatics* 12(2).
- Stott, Andrew. 2014. *Open Data for Economic Growth*. Washington DC. <http://www.worldbank.org/content/dam/Worldbank/document/Open-Data-for-Economic-Growth.pdf>.
- Vickery, Graham. 2011. *Review of Recent Studies on PSI Reuse and Related Market Developments*. Paris. <https://ec.europa.eu/digital-single-market/news/review-recent-studies-psi-reuse-and-related-market-developments>.
- World Wide Web Foundation. 2015. *Open Data Barometer 3rd Edition*. <http://opendatabarometer.org/>.
- Worthy, Ben. 2015. "THE IMPACT OF OPEN DATA IN THE UK: COMPLEX, UNPREDICTABLE, AND POLITICAL." *Public Administration* 93(3): 788–805.