

Special Issue on Data Literacy: Points of View

Some Key Challenges for Data Literacy

Mark Frank

University of Southampton, United Kingdom

Corresponding Author.

mark.t.frank@gmail.com**Johanna Walker**

University of Southampton, United Kingdom

J.C.Walker@soton.ac.uk

We discuss the motivation for holding the Data Literacy workshop at ACM Web Science 2015 and the key challenges for advancing the praxis and research of data literacy that emerged: defining data literacy; establishing its importance; identifying beneficial projects and building a community.

Our first encounter with data literacy was while researching how to measure the value of open data to users. As part of this we convened groups of people working in third sector organisations addressing homelessness in both the UK and India. In these workshops we asked participants to identify problems they currently faced in dealing with homelessness and then we presented them with open data that could theoretically assist with those problems. We noted that “Our users [in the study] were as competent as anyone could reasonably expect; technically and in their knowledge of subject matter and also in their understanding the significance of the data.” (Frank and Walker, 2016). They came armed with queries that have become institutionalised as ‘known unknowns’ when working with data, such as “Why was this data collected? Who collected it? When?”. “Yet” we wrote, “they struggled to interpret aspects of every dataset that was presented to them.” In the highly distributed, non-standardised world of open data, people who were highly literate, had domain expertise and were conventionally familiar with IT still experienced problems with open data severe enough to prevent them from attempting to use it.

Our experience of the workshops on both continents suggested we needed to investigate research into what constitutes data literacy - the skills, knowledge, attitudes, and social structures required for different populations to use data on the Web. It had become apparent to us that without data literacy, the impact of open data would be substantially reduced.

Frank, M., Walker, J. (2016). Some key challenges for Data Literacy. *The Journal of Community Informatics*, 12(3), 232–235.

Date submitted: 2016-09-22. Date accepted: 2016-09-23.

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/1349 .

Our first step was to identify who was producing work on this area. Given the emerging nature of this domain, there were no discipline-specific journals, departments or even academics that we were aware of, so we set about discovering the community in a more hands-on way. At the Open Data Camp Unconference in Winchester, UK we met with around 30 people interested in this area from practical, research and educational points of view. The conversation was stimulating but challenging.

Although it was relatively easy to identify why a lack of data literacy would be a problem, determining exactly what it was and who it was for was far more challenging. “The concept of data literacy is touted as the be-all-and-end-all solution to all information issues, but it’s pretty loosely defined, and may not be entirely viable for the wider public.” (from the Live Blog <http://odcamp.org.uk/data-literacy-what-is-it-and-how-to-promote-it/>).

There was a substantial divide that was characterised as the ‘English/plumbing’ issue: those who felt data literacy was like, “learning the English language – an absolutely necessity” and those who felt it was more akin to “learning the trade of plumbing – useful but something you’re likely to outsource”.

Our next step was to build on this experience by issuing a Call for Papers for a workshop for ACM Web Science 2015 (<http://www.dataliteracy.eita.org.br/1st-dl-workshop/>). Somewhat unusually for an academic workshop we not only sought academics to present peer-reviewed work (which is published in this special edition) but also invited practitioners and supporters of data literacy from the UK government, the Open Data Institute, the Open Knowledge Foundation and the School of Data to offer context.

The presentations of speakers and academics together provoked these key questions amongst attendees.

- (1) What is data literacy?
- (2) Why does data literacy matter?
- (3) What data literacy projects should we work on?
- (4) How do we build a data literacy community of practice?

The following were the key thoughts in response to these questions.

1. What is Data Literacy?

As discussed in Winchester, it is more than simply a prescribed set of skills. To meet the challenges data literacy needs to encompass enablers, attitudes and awareness. These are perhaps relevant to increasing numbers of people, where a few need concrete skills but everyone has a need for awareness.

It is an attribute of a community. In specifying that ‘Data Literacy is...enabling a community to take advantage of data’ we were guided particularly by the contribution from the School of Data (Rogers 2015), whose research suggests that data literacy is inseparable from its function within a community - whether defined in the traditional sense or in the more transitive sense of a group of colleagues or activists approaching an issue. That function is to enable a

community to take advantage of data sets, but it can clearly be imagined that people can play different roles in different groups, according to the relevant skills, enablers, attitudes and awareness within each group.

It is contextualised. Rogers also suggested that data is “over-analysed and under-contextualised”, which concurred with our conclusion that, “there is more work to be done on the best way to provide context to any given dataset, which would go some way to removing this onerous requirement from the user.” (Frank and Walker 2016).

2. Why does Data Literacy matter?

In exploring this question workshop participants reframed it as, “Why does Data Illiteracy matter?”. It was approached by breaking down assumptions about the challenges of data illiteracy. Various aspects emerged that were grouped under two spheres: power, and the ability to participate. As data, open, big, personal or in any other guise, becomes increasingly important, power will flow to those who are able to create, control and understand data. Those who cannot, will become powerless. Further, their ability to participate in society will be severely challenged as they lack the tools to engage with an important raw material of society.

3. What Data Literacy projects should we work on?

There was agreement that the scope of data literacy lends itself to collaborative projects. Specific areas of investigation that arose were:

- Methods for measuring the impact of data literacy in the long
- Defining the essential data literacy concepts, rather than focusing on skills that may be redundant within a couple of decades;
- Based on Question 1, focusing on data literacy as an attribute of community, investigating how this might play out in terms of Individual versus group competencies
- In terms of benchmarking; both what needs to be achieved and also whether there is progress, investigating how organisations might assess data literacy levels amongst their staff;
- Based on Question 2, clearly establishing the affordances of data literacy.

An interesting issue was raised regarding the possibility of improving data inputs via the process of those in charge of inputs actually working with the outputs and transferring the knowledge gained back into how the data is presented. This would take some of the onus of literacy from the data user to the data owner. If this was possible in one context, would it be transferable from one community to another?

4. How do we build a Data Literacy community of practice?

While our work on this workshop and subsequent work on this journal has identified many people doing vital data literacy work - both in academia and as practitioners - the community of practice is still vague. Both the building blocks of community (meetings, virtual and in person, social media groups, publications and more) and the stakeholders (in all sectors) are still to be identified. Doing this also requires that we define why we need a community of

practice in order that the right steps are taken and we can answer questions such as what is role of professional associations and events like this for convening groups.

A key question was whether the right approach was to build community and attract the right people to it or to identify the right people and form the community around them? The European Data Science Academy have experience in this, and created their community by approaching relevant LinkedIn groups, rather than building from the ground up.

Building on these questions constitutes both a roadmap and challenge for practitioners and researchers in the forum of data literacy. Meanwhile, a subsequent iteration of the Open Data Camp in Bristol in May, 2016, built on the first conversations we had held in Winchester. More people appeared to be identifying or experiencing data literacy challenges in their work. This led to questions arising around whether data literacy is primarily an issue for publishers or users. Given the representation from local government, it was inevitable that there would be some thought about whether policy makers need to be data literate.

In many ways the conversation had moved on but we also found we still lacked consensus on the same questions: What are the aspects of data literacy? Do they need categorising? Is there a basic standard everyone should meet? However, we also saw the emergence of two new directions of investigation. The first was the use of models of information literacy and traditional literacy that could be used to usefully conceptualise data literacy, and we see exploration of that in this Special Issue.

The second was the emergence of data literacy tools. Data Campfire, currently in beta, combines stories of how data is used with a ‘how we did it’ approach. The end result is a step-by-step guide to the creation of data projects. This is an exciting approach to improving capabilities with huge opportunities for development and many corollaries in other spheres. Much as many of us turn to YouTube if we need to learn how to replace a wheel on a bike or clean a washing machine filter without needing to be mechanics or engineers, this approach might allow us to turn to multiple formats of guides online to achieve data literate projects without ongoing expertise in this area. In the meantime, please do contribute any data stories you have to www.datacampfire.com.

Please join in the data literacy conversation. We invite you to feed back any thoughts you may have on this by tweeting using the tags #dataliteracy #jci.

References

- Frank, Mark Tyrrell, and Walker, J. (2016). User Centred Methods for Measuring the Value of Open Data” *The Journal of Community Informatics: Special Issue on Open Data for Social Change and Sustainable Development*, 12(2), 47–68.
- Rogers, K. (2015). Data Literacy Lessons from a Data Trainer. *Workshop on Data Literacy* (Web Sci 15), Oxford, UK.