BLOG

SEE ALL ARTICLES (../)

Should 'data literacy' be promoted?

Emmanuel Letouzé, Director and co-Founder, Data-Pop Alliance; Visiting Scholar, MIT Media Lab

November 7, 2016

> **Part 1 of 2.**
> This post is the first of two parts, and builds on the White Paper "Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data (http://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/)" published on September 2015. Part 2 will discuss guiding principles and priority actions to promote data literacy as literacy in the age of data.

'Data literacy' has become a mainstream term—a borderline buzzword—of the 'Data Revolution' lingua. In November 2014, the report "A World That Counts"[i] called for *"(a) proposal for a special investment to increase global data literacy"*, and efforts to develop "*an education program and promote new learning approaches to improve people's, infomediaries' and public servants' data literacy."* Many others have stressed the need for increasing 'data literacy', notably to address capacity gaps of National Statistical Offices.[ii] The Global Partnership for Sustainable Development Data has also made 'data literacy' one of its strategic pillars and priorities.

Yet, what is meant by and expected from 'data literacy' is not clear—even to most of its proponents. For all its qualities, the report "A World That Counts" did not define nor discuss data literacy, assuming, one would imagine, that its meaning was self-explanatory and value self-evident. Asked on their feet, most people would define data literacy as some version of "the ability to use (or

analyze) data". Others present data literacy as "a critical skill of the 21st Century"—the same way, the argument goes, that reading and writing were critical skills in the past couple centuries, in fostering democracy, development, *etc*.

Two common misconceptions about literacy

In my view this skills-based approach misses important points and caveats—both about data literacy and literacy. Specifically, promoting data literacy as a set of skills (such as the ability to perform analyses on data, read graphics, *etc*.) is grounded in two common misconceptions about literacy.

First, literacy cannot be equated with reading and writing skills. In fact, when defining literacy, the United Nations Educational, Scientific and Cultural Organization (UNESCO) makes no mention of reading and writing; rather: *"[l]iteracy involves a continuum of learning in enabling individuals to achieve their goals (…) and to participate fully in their community and wider society"*. Indeed, whereas *"at first glance,'literacy' would seem to be a term that everyone understands, (…), as a concept (it) has proved to be both complex and dynamic"*. Efforts to promote data literacy must be based on a much firmer understanding of what literacy is and does than is currently the case.

Second, to us educated professionals who read and maybe even write for a living, it seems literacy can only be a good thing. We often think of the Dark Ages, then writing, then Gutenberg, then the Enlightenment, and then Norway.[iii] But this narrative doesn't really stand the test of history. Reflecting on the historical role of writing, literacy and the fight against illiteracy, the French anthropologist Claude Lévi-Strauss wrote some 60 years ago in his masterpiece "Triste Tropiques":

"Writing is a strange thing. It would seem as if its appearance could not have failed to wreak profound changes in the living conditions of our race, and that these transformations must have been above all intellectual in character. (…) Yet nothing of what we know of writing, or of its role in evolution, can be said to justify this conception. If my hypothesis is correct, the primary function of writing, as a means of communication, is to facilitate the enslavement of other human beings.

The use of writing for disinterested ends, and with a view to satisfactions of the mind in the fields either of science or the arts, is a secondary result of its invention and may even be no more than a way of reinforcing, justifying, or dissimulating its primary function. If writing was not sufficient to spur knowledge, it may have been necessary to reaffirm domination structures. (…)".

Further, and *"[t]o bring the matter nearer to our own time"* (then the late 1950s), and our topic, according to Lévi-Strauss:

"*The Europeanwide movement towards compulsory education in the nineteenth century went hand in hand with the extension of military service and the systematization of the proletariat".* During those decades, he concluded, *"[t]he struggle against illiteracy is indistinguishable from the increased control exerted over the individual citizen by the holders of power."* [iv]

In other words, Lévi-Strauss argues that literacy campaigns during the Industrial Revolution were not meant to enlighten or empower people, as was and remains often assumed, but to entrench and expand emerging power structures and systems; those of young Nation States and the nascent capitalist economy—by creating soldiers, workers, foremen, tax-payers, and law-abiding citizens. In

his view, rather than aimed at enabling individuals *"to achieve their goals and to participate fully in their community and wider society"*, these literacy campaigns were designed to allow the powerful to exert greater control over the masses who had to know the codes just enough to follow them.

Could data literacy efforts reinforce prevailing power structures and dynamics?

An essential question, if we believe, as I do, that Lévi-Strauss' hypotheses weren't entirely incorrect, is whether the same may or will be happening with current and future 'data literacy' efforts. In other words: is there a risk that data literacy efforts may reinforce and perpetuate, rather than challenge and change, prevailing power structures and dynamics?

This bleak perspective of the nature and role of literacy throughout history must be taken with a grain of salt. It was the New Testament in the hands of local revolutionaries that ended the era of God-Kings; it was the Italian and Scottish enlightenments, printing, and widespread literacy that eventually broke the power of Kings and the Pope. There is no denying that mass literacy, even measured through reading and writing, played an important part in the spread of voting rights for example. But these facts aren't entirely at odds with Lévi-Strauss' arguments either. By and large, it is only when literacy reached the masses during the 20th Century, and, *while it did*, that its meaning and metric expanded beyond the ability to sign one's name, as it was initially measured, that it became a lever and force of fundamental positive socio-political change.

Why and how this matters for data literacy is obvious but worth spelling out. To make this point clearer, I call upon Godwin's law of Big Data: a reference to Edward Snowden. If data literacy were some version of "the ability to use (or analyze) data", then a human society entirely composed of NSA analysts would be as data literate as it gets (the same would go with Amazon analysts). The argument is a bit of a stretch but it sticks and triggers a sort of light-bulb moment: intuitively, we feel it would be a terrible society to live in. We sense that data literacy is or should be different—broader, thicker—than data crunching. In short: a narrow, shallow, skilled-based, technical, conceptualization of data literacy fails to recognize the complexities and ultimately deeply political dimensions and implications of data in and for our lives—and does not adequately capture the features and functions of data literate citizens and societies.

Why is data literacy *generally* promoted, both as a concept and objective, in such a way, as a set of technical skills to perform data tasks? My personal hypothesis is that those in positions of power—chief of whom in corporations and governments—who currently advocate for data literacy initiatives, *either* consciously constrain the concept to 'the ability to use data' to make it as politically innocuous and economically profitable as possible, *or* do not realize that 'truly' data literate citizens and societies will challenge and threaten their power, the same way it eroded the power of Kings and the Pope. So what would 'truly' data literate citizens and societies look like and do? I can only offer a few pointers.

'Truly' data literate citizens and societies would consider that facts matter and that those making false or unsubstantiated claims are not worthy of their votes or trust. They would critically assess graphics and articles before sharing them on social media. They would read and argue with the terms and conditions of use of social media and other services before accepting them. They would actively engage in ethical and legal debates on data collection and control to demand and obtain greater direct control over data about themselves. They would distinguish what can vs. should be done with

data. They would question the origin of data and the objective of their analysis. They would ask that algorithms governing growing portions of their lives be transparent and accountable so as to be able to be scrutinized, assessed, and redressed. They would also realize that it is preferable to be governed with open algorithms than by secluded autocrats. Facts would almost always trump fears in their everyday decisions. They would demand that new data and new methods be leveraged to better capture and change socio-economic processes and outcomes.

A world characterized and fashioned by such a breed of data literacy would put the position of current 'leaders' in jeopardy, for their power largely hinges on opacity, secrecy, credulity, insecurity, and so on.

A conceptualization of data literacy where data is used by and for people

In many ways, such data literate citizens and societies would and hopefully will soon *use* data—"*to achieve their goals (…) and to participate fully in their community and wider society"*. But they would *use* data as a socio-political anchor and lever—considering here data as a singular term. Doing so does and will often require some abilities to *use* data the way it is explicitly or implicitly meant in a skill-based conceptualization of data literacy—where data are considered plural, and valuable. But being able to use data in that latter sense (running a regression) is neither a necessary nor sufficient condition for being able to use data in the former sense.

A purely skills-based conceptualization of data literacy ("being able to use or analyze data") requires caring neither about where the data come from nor what their analysis is for. In contrast a conceptualization of data literacy where data is used by and for people "*to achieve their goals (…) and to participate fully in their community and wider society"* demands assessing means and specifying ends in light of agreed-upon societal norms—primarily ethical and philosophical. In some cases, a data literate individual would opt *not* to use data, if collected unethically or that may cause harm. Neither "data" nor "use" means the same thing in these two cases: in the former, data are resources to be exploited to fuel or grease existing power systems. In the latter data is an ecosystem to be shaped to challenge these very power systems.

The repeated mention of the phrase "*to achieve their goals (…) and to participate fully in their community and wider society"* of UNESCO's definition of literacy to describe the purpose of a broadened and thickened conceptualization of data literacy is of course not incidental. To be literate in the 21st Century, where being literate means being able to be an agent of change, our children will need to be data literate, defined as having "*the desire and ability to constructively engage in societies through and about data*".[v] Reciprocally, anyone who is data literate will be literate. There will be various levels of data literacy along a continuum of learning, the same ways there have been and remain different levels of literacy. And so I argue that for the children of the first and future 'data generations' of humans, data literacy will not be dissimilar to literacy, this concept that *"would seem to be a term that everyone understands, (…), [but] (..) has proved to be both complex and dynamic"*—adapting to circumstances and contexts. In the near future literacy will evolve and expand to include and then become data literacy such that data literacy will be literacy—literacy in a world and age of data.

To be clear, I am not advocating for *not promoting* data literacy. Instead, I am arguing against a technical, skills-based, conceptualization of data literacy (as some version of *"the ability to use data"*). And I am arguing that the current focus on data literacy is an opportunity, reflecting back on the nature and role of literacy in history, to promote and foster a consequentialist, broader and thicker, conceptualization of data literacy as literacy in the age of data, one that will allow citizens and societies to challenge current power structures and dynamics to meet their goals, and perhaps the Sustainable Development Goals.

References:

[i] Published in November 2014 by the Independent Expert Advisory Group (IEAG) on the Data Revolution for Sustainable Development appointed by the UN Secretary General http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf (http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf).

[ii] Cf Statistical Conference of the Americas in November 2015, for example: http://repositorio.cepal.org/bitstream/handle/11362/40065/S1600207_en.pdf?sequence=1&isAllowed=y (http://repositorio.cepal.org/bitstream/handle/11362/40065/S1600207_en.pdf?sequence=1&isAllowed=y)

[iii] Norway has topped the human development rankings for the past decade. See http://hdr.undp.org/en/content/human-development-index-hdi (http://hdr.undp.org/en/content/human-development-index-hdi).

[iv] The rest of the extract is *"For it is only when everyone can read that Authority can decree that ignorance of the law is no defence. All this moved rapidly from the national to the international level, thanks to the mutual complicity that sprang up between newborn states confronted as these were with the problems that had been our own, a century or two ago and an international society of peoples long privileged. These latter recognize that their stability may well be endangered by nations whose knowledge of the written word has not, as yet, empowered them to think in formulae which can be modified at will. Such nations are not yet ready to be "edified"; and when they are first given the freedom of the library shelves they are perilously vulnerable to the ever more deliberately misleading effects of the printed word."* Source: Claude Lévi-Strauss, Tristes Tropiques, 1955.

[v] Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data (http://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/)