Approaches to Building Big Data Literacy

Catherine D'Ignazio
Emerson Engagement Lab
160 Boylston St. (4th floor)
Boston, MA 02116, USA
catherine_dignazio@emerson.edu

Rahul Bhargava
MIT Center for Civic Media
20 Ames St.
Cambridge, MA 02142, USA
rahulb@mit.edu

ABSTRACT

Big Data projects are being rapidly embraced by organizations working in the social good sector. This has led to a proliferation of new projects, with strong criticisms in response focusing on the disempowering aspects of these projects. This paper identifies four main problematic aspects of Big Data projects in the social good sector to focus on: *lack of transparency*, *extractive collection*, *technological complexity*, and *control of impact*. Leveraging Paulo Freire's concept of "Popular Education", we identify an opportunity to work on these issues in empowering ways through literacy education. We discuss existing definitions of data literacy and find a need to create an extended definition of Big Data literacy. Surveying existing approaches to building data literacy, we identify the need for new approaches and technologies to address the problematic aspects of Big Data projects. To flesh out this concept of "Popular Big Data", we close by offering seven ideas for how the field can ensure that Big Data projects are in line with the values of organizations working in the social good sector.

1. INTRODUCTION

Big Data is increasingly important in the fields that deal with social good, including government, humanitarian aid, education and social change contexts. These sectors have adopted Big Data approaches in a variety of ways. Large multinationals are actively exploring applications [35]. Storage and analysis software authors are introducing their tools to these audiences [24]. We have seen well documented case studies in public health, food scarcity, human migration, and other classic "social good" fields [17]. Most of these studies use the language of "promise" and "hope" that Big Data will be able to solve problems at scale; on the Gartner hype cycle, we are well into the peak of inflated expectations [8]. Big Data analysis is being looked to as a new necessary skill in governmental, humanitarian and social change contexts.

The explosion of interest in Big Data has led to thoughtful critiques from a variety of perspectives. boyd and Crawford [14] raise concerns about epistemology (the "numbers cannot speak for themselves"), ethics ("just because it is accessible doesn't make it ethical") and accessibility (Big Data is the property of corporations and government who can selectively share and withhold from the public). Welles calls for Big Data to pay attention to the outliers and minorities which are often missed in aggregate analysis [42]. D'Ignazio, Warren and Blair emphasize the asymmetry of ownership, access and know-how in relation to Big Data and propose a citizen-led, participatory approach to collecting environ-

mental data [26]. And the recent Snowden Affair [2] led to widespread discussion of the perils of Big Data in regard to compromising privacy and surveillance [5].

Policy-makers and researchers are working on addressing these critiques. Argentina and the European Union, for example, have passed "Right to be Forgotten" laws which grant digital users more control over the retention of their data [9]. Lawmakers in California and the UK have special provisions for minors regarding personal data [33]. Technical solutions are being built to provide better anonymization and finer-grained control of personal data [18]. And visualization projects such as Mozilla Lightbeam [3], Immersion [31], and "Me and My Shadow" [16] raise awareness about personal metadata for the general online public.

These attempts are admirable, however in the context of sectors that work for the social good, we see a larger need for education about a number of problematic issues related to Big Data projects. Without basic understanding of data, how to use data and how to protect one's data, projects aimed at empowering users and citizens will fail. In this paper we highlight four specific issues related to many Big Data projects that need to be addressed: *lack of transparency*, *extractive collection*, *technological complexity*, and the *control of impacts*. We go on to define a notion of Big Data Literacy necessary to work on these issues. Next, we introduce some existing data literacy work, and conclude with a number of provisional and mostly untested ideas to tackle these problems.

2. BIG DATA HAS AN EMPOWERMENT PROBLEM

The hype surrounding Big Data has led to a widespread lack of awareness of what the term means. In discussing Big Data, we are using the three-part definition proposed by boyd and Crawford [14]. Quoting them, Big Data is comprised of:

1. **Technology**: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets.

2. **Analysis**: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.

3. **Mythology**: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.

A precise understanding of the definition, possible benefits and clear limitations of Big Data is especially critical in sectors concerned with social good. Non-profit advocacy groups, government, and others embracing Big Data projects in service of their constituents and citizens are especially focused on results that impact people in positive ways. One might argue that having Big Data be in service of the subjects needs' is sufficient to argue it is beneficial, but empowerment is not handed from those in power to those without, it bubbles from the bottom up.

In order to bring Big Data projects in line with the goals of the social good sector, we look to Paulo Freire's work on empowerment through literacy education. Freire was an educator in Brazil who used novel literacy learning approaches. His concept of Popular Education involves both the acquisition of technical skills and the emancipation achieved through the literacy process [25]. The latter is achieved through learner guided explorations, facilitation over teaching, accessibility to a diverse set of learners and a focus on real problems in the community [11]. This framing brings us back in line with the goals of the social good sector.

Bringing Freire's focus on empowerment back to the criticisms of Big Data projects, we find four problematic issues to focus on:

- **Lack of Transparency**: The data about people's interactions with the world is generally collected with only token approval, if any at all, from the user. This denies the subject awareness that their actions are being recorded at the time the actions occur.

- **Extractive Collection**: The data is collected by third parties and is not meant for observation or consumption by the people it is collected from (or about). This denies the subject agency in the data collection mechanism and interaction opportunities with the collector.

- **Technological Complexity**: The data is analyzed with a variety of advanced algorithmic techniques, and discussed with highly technical jargon. This denies the subject an understanding of how any results were achieved, and how they might be critiqued.

- **Control of Impact**: The data is used by the collector to used to make decisions that have consequences for the subject(s). This denies the subject participation in decisions that affect them.

These aspects, combined with the general confusion around what Big Data even means, have lead to a situation where subjects are routinely put in situations where they are not being served by the outputs of Big Data projects. As educators, we naturally approach this as a learning problem. Our key question is how educational approaches can be used to address these issues. How can we empower the subjects of the Big Data revolution? How can we ensure that the social good sector adopts Big Data in a manner that is consistent with its values? These are not simply theoretical distinctions – numerous reports have looked at the negative real world impacts created by these problematic aspects [27].

3. BIG DATA LITERACY

Within the space of education and with Freire's approach in mind, we define these problems as ones of "data literacy". We must work to address disempowering dynamics to contextualize and make Big Data relevant to learners. We must support tools and approaches that work on Big Data literacy in order empower the subjects of Big Data projects.

3.1 Defining Data Literacy

The emerging field of data literacy has many actors working on various topics and approaches. Some are building networks of trainers to support the growth of the open data field [4]. Others have focused on integrating into official school curriculum [30, 43, 19]. Previous work of our own has focused on using the arts as a bridge to building data literacy [12]. At the same time, many are trying to pick apart what data literacy means for a variety of audiences and what pedagogical frames to use [40]. These efforts represent strong attempts to define "data literacy" and put it into practice in a variety of ways.

The existing definitions of data literacy, including our own [12], are not quite adequate for bringing empowerment to the subjects of Big Data. Many of the cited definitions of data literacy have built on traditions in information and statistical literacy [30, 39]. More recent work has been about being able to put data into action [19]. For our purposes, data literacy includes the ability to read, work with, analyze and argue with data [13]. **Reading data** involves understanding what data is, and what aspects of the world it represents. **Working with data** involves creating, acquiring, cleaning, and managing it. **Analyzing data** involves filtering, sorting, aggregating, comparing, and performing other such analytic operations on it. **Arguing with data** involves using data to support a larger narrative intended to communicate some message to a particular audience.

This definition of data literacy is already connected to our set of problematic issues with Big Data projects. For instance, *arguing with data* is related to *control of impact*. However, it doesn't quite capture all of the issues. For example, *analyzing data* isn't possible for many due to the *technical complexity* of Big Data approaches. You can't *work with data* if it has been *extractively collected* and isn't available to you. You can't *read data* if there is a *lack of transparency* about when it is even being collected.

3.2 Defining Big Data Literacy

The gap between our existing definition of data literacy and our desire to address the problematic issues listed requires an extended definition of Big Data Literacy. These issues stem from the desire to ensure that Big Data use in the social good sector is in concert with its history and in service of its goals. To work on them we must have a solid understanding of what defines Big Data Literacy. We argue for including three more points for an extended definition of Big Data literacy:

- **Identifying** when and where data is being passively collected about your actions and interactions.

- **Understanding** the algorithmic manipulations performed on large sets of data to identify patterns.

- **Weighing** the real and potential ethical impacts of data-driven decisions for individuals and for society.

Here we additionally assert that Big Data Literacy is desirable for both non-technical and technical populations. Data scientists may be very sophisticated at the **Identifying** and **Understanding** pieces but lack the contextual

and domain knowledge to **Weigh the ethical impacts of their work**. The addition of these three aspects gives us the scaffolding we need to start to think about Popular Education-style activities and technologies that can address the disempowering aspects of Big Data we have singled out.

3.3 Applying Prior Work in Data Literacy

How do we create activities that are informed by this extended definition of Big Data literacy to address the disempowering aspects we have highlighted? To begin, let us look at our existing approaches to building data literacy in general. There is a variety of existing work on leveraging the power of small data for empowerment [41]. We have developed approaches in workshops and classrooms for small data sets. For example, Bhargava has worked with community groups to use their organizational data to create data murals [10]. D'Ignazio leads "walking data visualizations" that help citizens understand the rising seas of climate change [23] or the implications of a city's crisis response [22]. Both of us teach courses in data storytelling for undergraduates and graduates in which students work with government and community data sets to create data-driven journalism, media or art projects. We are also developing new data examination and exploration tools and activities to support data literacy learners [13].

As mentioned, all of these educational approaches work with small data sets - e.g. data that students can browse in Excel that relate to a single community group, a corpus of musical lyrics, or a geographic outline for a single city. We do not have an approach to do this with Big Data, nor do we have many examples of data literacy work that tackle Big Data for non-technical learners. There are sound reasons for this. First, there are no widely available, visual, populist tools for browsing large data sets, e.g. Excel for Big Data[1]. Popular WYSIWIG data exploration tools such as OpenRefine crash on over 100MB of data. Second, definitionally, the primary way of exploring Big Data is not through visual browsing but rather through complex algorithmic transformation. The prior work cited on data literacy has only sparingly tackled algorithmic literacy [28]. While there is growing research interest in the transparency of algorithms [38] [21], this is still nascent and the field has yet to articulate a well-formed pedagogical approach. In addition, the outputs of these algorithmic manipulations often require quite nuanced analysis and understanding to apply them well. Finally, because there are significant technical challenges in working with Big Data, there is the risk that Big Data Literacy might focus exclusively on the technical elements rather than the social processes and ethical questions that surround them.

4. POPULAR BIG DATA

Can we build a new model for "Popular Big Data" that empowers the non-technical subjects of Big Data? How might we go about this? What follows is a list of ideas to address the disempowering aspects of Big Data we have highlighted, based on the extended definition of Big Data Literacy we have provided, and built through the lens of the pedagogy we have discussed. We offer these as preliminary ideas. Some have been tested in educational settings and others are simply sketches to start conversations.

4.1 Idea 1: Participatory Algorithmic Simulations

We see an opportunity to leverage the history of participatory simulations to teach algorithms. Building on the long tradition of participatory simulations for understanding complex systems [15], we offer the idea that we could create participatory activities that let people *be* the data being operated on by a Big Data-related algorithm. What might this look like? Imagine TF-IDF[2] being explained by breaking a room of learners into groups by geography and talking about how each subgroup differs in their geographic diversity from the whole room. Linear search[3] could be illustrated by lining students up and then seeking the first student who fulfills some criteria, say, "Has eaten macaroni and cheese for breakfast". Drawing on the history of psychogeography, students could develop their own "walking algorithms" for how to explore outdoor spaces and test them out with each other [29]. We have not tried these yet, but see potential for some fun, explanatory activities that build up from an understanding of what algorithms are towards how some of the more complex algorithms operate. Simulating an algorithm with your body builds on the tradition of using body-syntonicity as a way to introduce computational concepts [37] at human-scale. These embodied simulations would directly address the Big Data empowerment problem of *Technological Complexity* by helping people *Understand Algorithms.*

4.2 Idea 2: Activities to Reverse Engineer Algorithms

Building on existing work in reverse engineering algorithms [21, 20], new learning activities might include very simple tracking or reverse-engineering experiments for the algorithms that the students experience on a daily basis (e.g. Google search, Amazon.com recommendations or Facebook feed). These might be particularly interesting in a comparative context. For example, if you ask learners to search for the same thing and compare their Google results you can spark a conversation about personalization. Another comparative activity might be to read about different search algorithms, compare Google results with Duck Duck Go and speculate on the reasons for the different results. Tracking and reverse-engineering algorithms with students directly addresses the Big Data empowerment problems of *Lack of Transparency*, *Extractive Collection* and *Control of Impact*. Through these activities students start from the outcome of an algorithmic process (Facebook feed) and trace it back to the source data points (Facebook is tracking who I speak with most often, what I like, keywords in my posts, etc) in order to reflect on a process that is hidden from public view. Looking back at our definition of Big Data Literacy, these activities address the questions of *Identifying when and where data is being collected* and *Understanding algorithms.*

---

[1]The maximum number of rows for an Excel spreadsheet is 1,048,576

[2]TF-IDF is short for Term Frequency-Inverse Document Frequency. It is an algorithm that surfaces the frequency and uniqueness of a single word in relation to a corpus or collection of words.

[3]Linear Search is one of the most basic algorithms that searches sequentially for a value in a collection and stops when it finds that value.

4.3 Idea 3: Explanatory Graphics About Algorithms (Descriptive and Speculative)

Other activities, particularly for visual thinkers like artists and designers, might be to develop flow charts and information graphics to explain how particular algorithms work. Activities for understanding algorithms should focus on describing how they work based on research and experimentation (reverse-engineering). Additionally, students could produce speculative infographics for an algorithm *should* work. After they have read about and experimented with how a particular algorithm works, what it prioritizes and what might be left out, students might produce a flow chart or infographic for a "better" algorithm based on certain criteria that emerge from their particular community or concerns. This kind of speculative thinking situates the learner as an author of reality and activates what Henry Jenkins calls "the civic imagination" - the capacity to imagine alternatives to current social, political or economic conditions as well as to see oneself as an active political agent [32]. Creating descriptive and speculative infographics address the Big Data problems of *Lack of Transparency* and *Technological Complexity* to make the inner workings of common algorithms visually accessible. Placing the student in a speculative position activates the *Weighing the real and potential ethical impacts* aspect of our definition of Big Data literacy. How should the algorithm work? And for whom?

4.4 Idea 4: Learner-focused Tools

While there is a proliferation of tools created to assist novices in gathering, working with, and visualizing data [6, 36] most of these tools focus on small data sets. There is a significant lack of popular, visual tools for working with Big Data. In addition, existing tools currently focus on outputs (spreadsheets, visualizations, etc), and not on the ability to help novices learn. Visualizations, which garner so much popular media and social media attention, are the outputs of a process. These flashy pictures attract the bulk of the attention, which has led tool designers to prioritize features that quickly create strong visuals, at the expense of tools that scaffold a process for learners. There has been little discussion of why and when to use these tools in appropriate ways for the learners that do not yet "speak data". We assert that designing for learners is fundamentally different than designing for users. Data literacy tools for learners need to be introduced with activities that are inclusive, use data that are relevant to the learner, and be open to creating unexpected outputs. In prior work we discussed four design principles for learner-centered data tools: focused, guided, inviting and expandable [13]. We are currently working on a suite of data literacy tools called DataBasic.io that enact these principles. Learner-focused tools could address the Big Data empowerment problems of *Technological complexity* and *Extractive collection* by making data collected about subjects available to them to explore and operate on. They could also help address the *Understanding algorithms* aspect of our definition of Big Data literacy.

4.5 Idea 5: Partnerships to Leverage Results

Typically Big Data analysis is performed by people with advanced technical skills, but outside of the context of the data. One way to build Big Data literacy is to create settings for more engaging partnerships that bridge between those who own the issue and context and those who have the technical expertise. DataKind [1] is an example of an organization in this space. They offer up a network of volunteer data scientists to partner with NGOs in order to help them analyze their data. The NGO owns the mission and goals, and through their close collaboration with the data scientists, guides the driving questions of the Big Data exploration. This type of relationship between advocacy groups and data scientists is fundamentally more empowering than simply throwing the data at the proverbial wall for analysis and results. Bridging partnerships give subjects of the data more *Control over the impact* of the findings and mitigate other Big Data empowerment issues such as *Extractive collection* and *Technological complexity* by creating dialogue around each stage of the process. In the process, the NGO or community achieves greater *Understanding of algorithms* and the data scientists have deeper and more meaningful conversations to *Weigh the real and potential ethical impacts* of their work. This idea also highlights that it is not only the subjects of Big Data that stand to benefit from Big Data Literacy but also the data scientists and statisticians who may have technical skills but lack contextual knowledge to assess the ethical impact of their work.

4.6 Idea 6: Creating a Data Journal

Big Data is often built on the passive collection of data created from daily interactions with infrastructural or personal technologies. In Bhargava's graduate and undergraduate-level data storytelling course, he requires students to create data journals that ask them to write down every time an interaction creates data over the course of one day [7]. The resulting lists are quite illuminating and students in class discussions demonstrated an increased awareness of how often they are creating data. This activity builds awareness of the problematic *Lack of Transparency* and *Extractive Collection* involved in many Big Data efforts. A follow up discussion about the most benign and nefarious of these data owners starts to build an ability for *Weighing the impacts and ethics* of these efforts.

4.7 Idea 7: Build Tools for Distributed Ownership

In order to truly realize "Popular Big Data", we must not lose sight of the powerful asymmetries that exist in the Big Data ecosystem. Data flows from the everyday actions of agents in the system into the databases of central corporate and governmental authorities and data brokers. These authorities own the data, own the algorithms and own the infrastructure. They choose what is open and what is classified or proprietary. Responding to the extractive and opaque aspects of Big Data, we suggest building infrastructures that allow for returning ownership of the data to the subjects. While users can choose to opt-out of large scale data collection from large companies like cellular phone providers, they do not allow users to own their actual data [34]. The previously mentioned OpenPDS system is attempt to model what this type of system would look like [18]. Empowering more people with data might also mean fundamentally rethinking the Big Data ecosystem and developing more peer-to-peer methods for owning and sharing data. Some of these imbalances might additionally be corrected by developing more legal and regulatory tools (e.g. a FOIA for algorithms) for citizens to access data and algorithms. This idea builds across all of the Big Data empowerment problems from *Lack*

of Transparency to *Control of Impact* and acknowledges that Big Data Literacy goes beyond the education sector to include efforts from law, policy, science and technology.

5. CONCLUSION

The rise of Big Data is an important trend in the social good sector. Its ascendance makes it necessary to consider whether projects are in concert with producing positive outcomes for the data subjects. We make the case that Big Data has an empowerment problem due to its lack of transparency, extractive collection, technological complexity and lack of control over impact. These problems can be partially addressed through our field: education. We propose a definition for Big Data Literacy that borrows from Freire's assertion that literacy is not just about the acquisition of technical skills but the emancipation achieved through the literacy process. We outline why Big Data Literacy is a challenging concept and propose a number of provisional ideas for learning activities that can start to build literacy amongst non-profits, technologists, and the regular folks the data is about. We assert that it is not only the non-technical people but also the technical people who can stand to benefit from Big Data Literacy. In order to truly leverage data for social good the field must grapple with these questions of participation, empowerment and literacy.

6. REFERENCES

[1] DataKind. http://datakind.org/. Accessed: 2015-07-1.
[2] Edward Snowden. https://en.wikipedia.org/w/index.php?title=Edward_Snowden&oldid=673506052. Accessed: 2015-07-28.
[3] Lightbeam for Firefox. https://www.mozilla.org/en-US/lightbeam/. Accessed: 2015-07-28.
[4] School of Data. http://schoolofdata.org. Accessed: 2015-07-25.
[5] The NSA files. http://www.theguardian.com/us-news/the-nsa-files, 2013. Accessed: 2015-07-28.
[6] *Visualizing Information for Advocacy*. Tactical Technology Collective, The Netherlands, second edition, 2014.
[7] data logs | Data Storytelling Studio @ MIT. http://cms631.datatherapy.org/category/assignments/data-logs/, Feb. 2015. Accessed: 2015-07-21.
[8] Hype cycle. https://en.wikipedia.org/w/index.php?title=Hype_cycle&oldid=670485416, July 2015. Accessed: 2015-07-23.
[9] Right to be forgotten. https://en.wikipedia.org/w/index.php?title=Right_to_be_forgotten&oldid=672390411, July 2015. Accessed: 2015-07-28.
[10] R. Bhargava. Data Murals. http://datatherapy.org/data-mural-gallery/, note = Accessed: 2015-07-28.
[11] R. Bhargava. Towards a Concept of "Popular Data" | MIT Center for Civic Media. https://civic.mit.edu/blog/rahulb/towards-a-concept-of-popular-data, Nov. 2013. Accessed: 2015-07-28.
[12] R. Bhargava. Data Therapy: Activities. http://datatherapy.org/activities, 2015.
[13] R. Bhargava and C. D'Ignazio. Designing Tools and Activities for Data Literacy Learners. June 2015.
[14] d. boyd and K. Crawford. Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5):662–679, June 2012.
[15] V. S. Colella. Participatory Simulations: Building Collaborative Understanding through Immersive Dynamic Modeling. *The Journal of the Learning Sciences*, Vol. 9, No. 4:471–500, 2000.
[16] T. T. Collective. Me and my shadow. https://myshadow.org/trace-my-shadow. Accessed: 2015-07-23.
[17] A. Coppola, O. Calvo-Gonzalez, E. Sabet, N. Arjomand, R. Siegel, C. Freeman, and N. Massarrat. *Big Data in Action for Development Report*. World Bank Group, 2014.
[18] Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland. openPDS: Protecting the Privacy of Metadata through SafeAnswers. *PLoS ONE*, 9(7):e98790, July 2014.
[19] E. S. Deahl. *Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments*. PhD thesis, 2014.
[20] N. Diakopoulos. Algorithmic accountability reporting : on the investigation of black boxes, Feb. 2014.
[21] N. Diakopoulos. Algorithmic Accountability. *Digital Journalism*, 3(3):398, May 2015.
[22] C. D'Ignazio. It takes 154,000 breaths to evacuate Boston. http://www.kanarinka.com/project/it-takes-154000-breaths-to-evacuate-boston/, 2008. Accessed: 2015-07-28.
[23] C. D'Ignazio. Boston Coastline: Future Past | kanarinka. http://www.kanarinka.com/project/boston-coastline-future-past/, 2015. Accessed: 2015-07-28.
[24] S. O. Fadiya, S. Saydam, and V. V. Zira. Advancing Big Data for Humanitarian Needs. *Procedia Engineering*, 78:88–95, 2014.
[25] P. Freire. *Pedagogy of the Oppressed*. 1968.
[26] O. Grandstrand. Less is more: The role of small data for governance in the 21st century. In W.-J. B. D. D'Ignazio, C., editor, *Digital Governance*. 2014.
[27] M. B. Gurstein. Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2), Jan. 2011.
[28] K. Hamilton, K. Karahalios, C. Sandvig, and M. Eslami. A path to understanding the effects of algorithm awareness. *Conference on Human Factors in Computing Systems Proceedings*, page 631, Apr. 2014.
[29] J. Hart. A new way of walking. *UTNE-MINNEAPOLIS*, pages 40–43, 2004.
[30] K. Hunt. The Challenges of Integrating Data Literacy into the Curriculum in an Undergraduate Institution, 2004.
[31] D. Jagdish. *IMMERSION : a platform for visualization and temporal analysis of email data*. Thesis, Massachusetts Institute of Technology, 2014.
[32] H. Jenkins, S. Shresthova, L. Gamber-Thompson, and

N. Kligler-Vilenchik. Superpowers to the people!: How young activists are tapping the civic imagination. In E. Gordon and P. Mihailidis, editors, *Civic Media: Technology, Design, Practice*. MIT Press, Cambridge, MA, 2016.

[33] T. John. Teenagers Could Have the Right to Delete Their Online History in the U.K. `http://time.com/3974435/uk-teenagers-online/`, note = Accessed: 2015-07-27, July 2015.

[34] J. Leber. How Verizon and Other Wireless Carriers Are Mining Customer Data. `http://www.technologyreview.com/news/513016/how-wireless-carriers-are-monetizing-your-movements/`, note = Accessed: 2015-07-30, Apr. 2013.

[35] E. Letouzé. *Big Data For Development*. UN Global Pulse, May 2012.

[36] D. Othman, H. Craig, and A. Debigare. NetStories. `http://www.netstories.org/`. Accessed: 2015-05-14.

[37] S. Papert. *Mindstorms: children, computers, and powerful ideas*. Basic Books, New York, 1980.

[38] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, 2014.

[39] M. Schield. Information literacy, statistical literacy and data literacy. *IASSIST Quarterly*, 28(2/3):6–11, 2004.

[40] A. Tygel and R. Kirsch. Contributions of Paulo Freire for a critical data literacy. June 2015.

[41] J. Warren. The Promise of Small Data. `http://techpresident.com/news/24176/backchannel-promise-small-data`, July 2013. Accessed: 2015-07-23.

[42] B. F. Welles. On minorities and outliers: The case for making Big Data small. *Big Data & Society*, 1(1):2053951714540613, Apr. 2014.

[43] S. Williams, E. Deahl, L. Rubel, and V. Lim. City Digits: Local Lotto: Developing Youth Data Literacy by Investigating the Lottery | Journal of Digital and Media Literacy. 2015.