Contributions of Paulo Freire for a critical data literacy

Alan Tygel
Graduate Program on Informatics – UFRJ, Brazil
alantygel@ppgi.ufrj.br

Rosana Kirsch
EITA Cooperative, Brazil
rosana@eita.org.br

ABSTRACT

Paulo Freire is the patron of education in Brazil. His main work - the Popular Education pedagogy - influenced many educators all over the world who believed in education as a way of liberating poor oppressed people. One of the most important outcomes of Freire's work was a literacy method, created in the 1960's. In this paper, we propose adapting elements of the Paulo Freire's Literacy Method for use in data literacy, i.e., the act of building capacities for working with data. After tracing some parallels between literacy education and data literacy, we suggest some data literacy strategies on generative themes, thematization, problematization and systematization stages, and derive from it a definition for critical data literacy. The preliminary contributions are mostly theoretical, even though they come from a previous concrete experience in teaching open data for social movements.

General Terms

data literacy, literacy education, popular education, open data, Paulo Freire

1. INTRODUCTION

The introduction of new technologies in the everyday life is an irrefutable reality. Even though in unequal ways, the information and communication technologies (ICTs) impact both those who have structure and education to enjoy the comfort brought by the ICTs, and those who do not. To analyse these impacts from a critical point of view, since the beginning of public Internet, in the 1990's, studies about *digital divide* – a term coined to define this social phenomenon – were developed, relying on a supposed *digital inclusion* as a way to overcome the inequalities[1].

One fundamental step of *digital inclusion* is *digital literacy*, a term which references a parallel between the act of learning how to read and write – literacy – and the act of learning how to deal with computers. And with the growing presence of ICTs in the society, specialized questions arise inside digital literacy.

From the mid-2000s onwards, governments in the whole world, including Brazil, started to publish on the Web big quantities of data [4]. It was the beginning of a worldwide movement, and also when, at the same time, the term *data literacy* started to be used, even without a formal definition. The promises brought by the open data initiatives relates to a more transparent society, a deeper participative democracy, and possibilities of generating value from data [13]. Meanwhile, with the severe social inequalities, directly reflected in the education level of the population, there is a strong potential of generating a mass of *data illiterates*.

In the Northeast Region of Brazil, on the 1960's, the illiteracy rate – percentage of adult people who could not read or write – reached 72,6% [6]. In that context arose the work of the philosopher Paulo Freire. He characterized the process of literacy education as both referring to learning technically how to read and to write, and to the emancipatory process of understanding and expressing itself in the world: "to learn how to read is to learn how to say the own word. And the own human word imitates the divine word: it creates." [7].

In this paper, we aim to trace parallels between the reflections of Freire about literacy education and the process of technical control and critical analysis over data, bringing elements to comprehend the new phenomenon of data literacy. We advert that this is an introductory paper, with a series of limits. The scarce literature about data literacy obliges us to bring inspiration from other sources, and is precisely in this sense that we seek the contributions of alphabet literacy methods to the field of data literacy. The ideas brought here are mostly in the theoretical field. Nevertheless, they came from concrete experiences in teaching open data [17] and developing information systems for social movements [2].

It must also be noted that Freire's development was driven in a specific context – teaching poor peasants how to read and write, with the intention of raising their consciences – and thus, any use in other contexts move must take this into account.

2. PAULO FREIRE AND THE LITERACY EDUCATION

In the whole Latin America, but especially in Brazil, the history of education cannot be told without the name of Paulo Freire. Born in Pernambuco, in 1921, he became worldwide famous for his critical

---

[1] There is a vast literature about digital divide, which is out of the scope of this paper. For a very recent debate on this topic, we recommend Gurstein's paper *Why I'm giving up on the digital divide* [12].

[2] The EITA Cooperative dedicates to developing ICTs for social movements. More on http://www.eita.org.br

pedagogy, mostly on the development of the philosophical principles of the Popular Education, whose product is a literacy method.

In the 1960's, the traditional literacy method was spread through primers[3] This was the central working tool for education, and the focus was on repeating loose words, and in creating decontextualized phrases to reinforce syllables and words. Some classic examples are shown in Table 1.

Table 1: Decontextualized phrases used in traditional literacy method, in Brazil.

Phrase in Portuguese	Consonant Highlighted	Translation in English
Eva viu a uva	V	Eva saw the grape
O boi baba	B	The ox drool
A ave voa	V	The bird flies

Freire said once that "it is not enough knowing that Eva saw the grape. It is necessary to comprehend what is the position of Eva in the social context, who worked to produce that grape, and who profited with this work" [8]. Moreover, Eva is an extremely uncommon name in the Northeast Region of Brazil, as well as the grape, grown typically in the south of the country. The statement is therefore completely decontextualized, and do not encourage the students other than memorizing.

According to the Freirean philosophy, the education must be contextualized, i.e., it should arise from the concrete experience of the educands[4], and from what is familiar to them.

To Freire, the comprehension of the reality does not occur through a mechanical relation between a sign – the written word – and a thing, but by the dialectical interaction subject-reality-subject, where signs and things relate themselves in a political, cultural and economic context. Therefore, the concepts *Eva* and *grape* should not be treated abstractly, but inside a context and a reality.

In a very simplified way, we can say the Freire's Literacy Method has three stages [16]:

1. **Inquiry stage:** the themes and words that compose the reality of the educands are defined. These themes must be part of the everyday life of the space where the educands live, and be very familiar.

2. **Thematization stage:** the themes are coded and decoded, alongside the discussion about their social meaning in the world.

3. **Problematization stage:** questioning about the meanings previously discussed, in a perspective of transformation of the reality.

The first big experience of the application of the method happened in Angicos, a city in Rio Grande do Norte state, in the Northeast

---

[3]Primars, translation of *cartilhas*, are booklets containing the content to be taught.

[4]Some words used in this paper are specific from Freire's bibliography: educands (students), educators (teachers), thematization and problematization. Debating the origin of them is out of the scope of this work.

Region of Brazil. In 1963, 300 sugar cane cutters became literate in 45 days, with 40 class hours, without any primers. At that time, the president of Brazil, João Goulart, invited Paulo Freire to organize a National Literacy Plan, with the goal of teaching more than 2 million people to read and write. The plan began in January 1964, but was quickly aborted by the civil-military coup, in 1st April 1964. Paulo Freire's method was substituted by the Brazilian Literacy Method (MOBRAL, in Portuguese), where all the critical view was removed. Paulo Freire was arrested, and had to leave the country, returning only in 1980.

When tackling Paulo Freire's Literacy Method, the Popular Education perspective must also be mentioned. As a whole educational philosophy, it is inspired in the stages of the literacy method, going deeper in its reflections. In the 1970's, many experiences of Popular Education in the South Cone – Chile, Argentina, Uruguay and Brazil – generated the reflection of this pedagogy as a permanent process of theorization over the practice in the context of the organization of the workers class [14]. The process of collective construction of knowledge from generative themes and thematic axis, emerged from a lived reality, received than the name of *Experience Systematization*. This should also be included as a fourth stage in the literacy method:

4. **Systematization stage:** organization, interpretation and presentation, in a communicative sense, of the lived experience.

3. PARALLELS BETWEEN LITERACY EDUCATION AND DATA LITERACY

Being data literacy a new study domain, and thus under construction, there is no established definition for the term. According to the *Data Journalism Handbook*, "data literacy is the ability to consume for knowledge, produce coherently and think critically about data"[10]. The *Wikipedia* term states that "Data literacy is the ability to read, create and communicate data as information."[19]. Another work highlights the importance of understanding how to produce data [3]. After discussing the parallels between literacies, and the possible contributions of Paulo Freire, we derive our own definition of Data Literacy in the end of Section 4.

Before discussing what contributions from Freire can be brought to data literacy, it is necessary to trace some parallels between elements of popular education, in general, and Freire's Literacy Method, in particular, and data literacy. In the following, we present three of such parallels.

As stated above, literacy education has two complementary and indivisible aspects: the technical ability of reading and writing, and the social emancipatory process of understanding and expressing itself in the world. In data literacy, we can observe the technical capacities related to data manipulation, including general computer abilities and statistical-mathematical methods, and the capacity of critically analyzing data, understanding the context where they were generated, and the reality pictured with them.

Looking further into the technical aspect, we can trace another parallel: data literacy entails a higher technological complexity comparing with alphabetization. While the literacy education process demands only the necessary instruments for reading and writing – a book, a pencil and a paper – the data literacy education normally demands computers, mobile devices, and internet connection. Mathematical reasoning skills are also fundamental to this

process. So, we can affirm that *data literacy is a technically more complex process than literacy education*.

Relating to the absence of literacy, we can say that the social exclusions caused by both kinds of 'illiteracy' have deep differences, as a third parallel. According to the Brazilian statistical agency, in 2013 8.5% of the population older than 15 years is illiterate. A closer look reveals a high correlation with poverty and regional inequality. In the Northeast Region, the poorest of the country, the index almost doubles: 16.6%. The rural slice reveals an even higher index: 18.6% of countryside residents are illiterate. Therefore, *a correlation between illiteracy, socio-economic standing, and geographical location can be observed*.

In turn, "data illiteracy", if we can already refer to this term, covers a much larger slice of the population and results in more subtle disadvantages, which however tend to get stronger as far as the open data policies advance. Gurstein [11] cites two examples where "data illiterates" were severely affected by the release of land ownership records as open data, one in Nova Scotia, Canada and another in Bangalore, India. By not having access to data, in both cases, small farmers lost their land for other landowners who checked inconsistencies in the land records and judicially claimed the ownership. The small farmers were elderly and illiterate, and thus also data illiterate.

4. ESSAYS ABOUT A FREIREAN INSPIRED CRITICAL DATA LITERACY

In the following, an exercise of adapting key-concepts of Freire's Literacy Method to what we are going to call, on a preliminary basis, *critical data literacy*, will be presented. At the end of this section, we derive our own definition for the term.

4.1 Technical Literacy and Emancipation

As already discussed, the literacy concept can be analyzed in two dimensions: the technical ability and the emancipation achieved through the literacy process. Given the high technical complexity of data manipulation, it seems to be a natural tendency that this dimension suppresses the other. Immerse in studies involving the use of computers, specialized software, various data sources and statistical methods, there might be a tendency of the educands to leave behind the critical reflection about the social meanings of data in the world, and therefore the emancipatory perspective can be lost.

The emancipatory perspective resulted from data literacy can be materialized in certain abilities acquired be the educands, for example:

- **Comprehension of the reality:** critical analysis of a specific reality, based on benchmarking and statistics. As an example, we can cite the Brazilian Agricultural Census[5], which proves with data what peasants already know from their own experience: 45% of the arable land is controlled by 1% of landowners, making Brazil one of the countries with the most concentrate land possession in the world.

- **Questioning of hegemonic (or common sense) concepts:** based on hegemonic concepts sustained by data, comprehension of the mechanics behind the generation of these data,

_____

[5]This census is driven by the Brazilian Statistical Agency - IBGE

and thus questioning it. One example is the concept of Gross Domestic Product (GDP). Although considered to be the most important measure of a country's economy, it does not consider the income distribution or the environmental consequences of economic development.

- **Development of new well-founded concepts:** to enlighten invisible realities through the consistent generation of data. In 2007, a mapping revealed that almost 2 million people in Brazil work in self-managed cooperatives, within the Solidarity Economy context [18]. This concept sheds light on other forms of work organization, which normally are hidden or considered small experiments.

4.2 The Generative Themes

The primordial idea behind the generative themes is that the educational process shall start from the educands reality. Just like the grape is not a fruit typical from the Northeast Region of Brazil, a database is also probably not something that explicitly makes part of the everyday life of data educands. (Their personal data, however, are almost sure registered in one or more databases.) Thus, there is a challenge for educators to dialogue with educands about themes that have to do with the concrete aspects of their lives [5]. On data literacy it is necessary to search in the reality of each educand elements that relate to how the use of data can contribute to understanding and modifying the world.

The generative theme is related to "the universe of speech, culture and place, which must be inquired, surveyed, researched, unveiled" [2]. The research of the universe of vocabulary and the identification of keywords of the group or community are the base for developing generative theme, and thus, for literacy education. They express limit situations, which for Freire are mostly oppressive situations [5]. On the data literacy context, the reality diagnosis of the involved group and the elaboration of a generative theme may point to the use of specific data sets, whose analysis will be organized in thematic axis.

4.3 Thematization

At this stage, the main goal is to motivate the understanding of the world through data. Either for a local or global reality, about specific or generic themes, data allows an understanding of reality commonly understood as "objective".

The elaboration of *thematic axis* relates the generative theme with aspects of a particular or conjunctural reality, and at the same time, organizes the learning process in an articulated sequence. The thematic axis seek for interweaving diagnostics and theoretical questions [15], fostering the dialectic sequence action-reflection-action from the group involved in the learning process.

As stated by [7], one way of dealing with thematic axis in the learning process is with the codification process, i.e., the representation of the world using symbols as language, drawing or images. Conversely, decodification is the process of interpreting the codifications. The decodification generates new information through the production of more abstract higher level codification, based on the knowledge of the world that each educand has [1].

In data literacy education, coding a certain thematic axis as aggregated information – statistics, graphics, tables – may lead to more complex decoding about the same theme. A reality can be coded

into data, which can be once more coded into aggregated information, and then can be further decoded, generating a modified view over the same reality. It is always important to notice that this process has an intrinsic bias, related to the design choices at data acquisition and processing.

4.4 Problematization

After the "enchantment" with the world of data, it is fundamental to problematize it, i.e., to unveil what is behind the scenes when talking about data. In order to use data with responsibility, it is necessary to know where they came from, how they were generated and to what purpose. Thus, it is possible to politicize the use of data, and deal with them not only from the point of view of a passive user, but from the perspective of someone who is also able to produce data, and with them, "say his word".

We list here, without any aspiration of completeness, two issues that can serve as a starting point for the problematization stage:

Non-neutrality of Data

Data are not neutral. The seducing precision and objectivity of data grounded statements almost always hide ideologies and intentions about anything one wants to prove. Thus, it is fundamental to problematize data origin. Are data from the government or from civil society organizations? What was the political position of that organization at that time when data were generated? If it is about scientific data, who funded the research? More complex, but also of great importance, is the knowledge of the methodology used to gather data. Unawareness of the methodological approach can lead to misunderstandings and flawed conclusions.

With that information – origin and method – it is possible to infer what was the objective of data generation, in the case it is not explicit. Producing data is a costly activity, which requires a considerable amount of resources, especially when dealing with big areas, as in the case of Brazil. Therefore, every research that generates data has a very well defined purpose, which must be unveiled and problematized.

The research is designed by specific actors, to reach strategical goals, as well as methodologies are designed in order to highlight some aspects, and not others. This is why we can affirm that data resulted from these researches are not neutral, and therefore its non-neutrality must be problematized in a critical perspective of data literacy education.

Transparency

In many cases, the critical use of data will come across the lack of other data. These missing data may not exist, be hidden or poorly organized, which is the case of most governmental data. In order to work critically with data, it is necessary to have conscience of one's rights to access information, which is directly related to *transparency policies*. Many countries are advancing in this field,

publishing their data online and creating laws to guarantee access to information, transparency and open data, with the valuable argument of enhancing democracy and fighting corruption. However, as stated by the Global Open Data Index[6], only 11% of assessed datasets in 97 countries are open. Thus, discussing transparency and access to information is a possibility of problematizing data literacy.

4.5 Systematizing

Systematizing, more than put data and information about a context together, is the exercise of theorizing about an experience, deeply analyzing it. Systems of though, information, management and action imposed by dominant powers promote a unique vision of the lived world, and this stage has the aim of elaborating an alternative view [9]. The act of systematizing implies in an evaluation of the advances and innovations generated inside a collective experience, which can inspire other groups in other realities.

Systematizing requires data and information about an experience. In the data literacy context, the ability of putting together data retrieved from various external sources and subjective qualitative information empirically obtained should be encouraged.

In data literacy, the systematizing stage should be the conclusion of the whole process – inquiry, thematization and problematization. Of crucial importance is the communication of the results. Data can be exposed in several forms, such as graphics, tables, maps, infographics, music, film or even text. Sensibility to choose the right way of systematizing and communicating data is certainly a point that should be stressed in data literacy.

4.6 Definition

Considering the arguments developed in this section, we derive our definition of *critical data literacy*:

DEFINITION 1. Critical Data Literacy *is the set of abilities which allows one to use and produce data in a critical way. This set is composed by:*

- **Data reading:** *The ability of reading data starts at understanding how data was generated, i.e., which methodologies were used in order to capture data from a context, which facts, measures and dimensions were considered, and at which grain data was collected. Data shall not be read as an objective fact, but as the output of a social process.*

- **Data manipulation:** *The ability of technically manipulating data is related to the use of computational and statistical tools in order to transform data into information. Linking data with other sources is also and important skill. Data shall be manipulated based on explicit objectives.*

- **Data communication:** *The ability of communicating data comprehends finding the better matches between data types, such as distributions, temporal series, networks or comparisons, and communications tools, such as text, tables, several types of charts, maps or infographics combining these elements. Communicating data also comprehends a social evaluation of what message should be transmitted to which*

---

[6]http://index.okfn.org/place/

target audience. Data communication shall be done in a responsible and precise way, in order to avoid misunderstandings or invalid conclusions.

- ***Data production:*** *The ability of producing data includes deepening all elements listed at the* Data reading *ability. Additionally, knowledge about data formats and data publishing tools is required. Generally, data shall be published not only respecting the Open Definition[7], but also offering tools so that non-experts are able to use it.*

5. CONCLUSIONS

The fast spreading of ICTs in the society has, as one of its consequences, a recent publication of massive quantities of data over the Web. These can be either related to governments, through public transparency initiatives, or generated by companies or civil society organizations, or even originated from scientific research. This huge mass of new information brings with it a series of potential benefits, but also major challenges, which most of the time are hidden. There is an imminent risk of establishing an elite able to profit from these data, interpret it and act in the world through it, overlapping wide excluded masses. In this paper, we sought in the work of Paulo Freire inspirations for the construction of a critical data literacy, which is aware of this challenge.

Future works on this topic includes deriving more tangible examples of the application of this methodology in practice, followed by developing a strategy to assess and evaluate the outcomes. From the theoretical point of view, a deep analysis of the digital literacy literature could also bring more elements for data literacy.

It was not at random that Paulo Freire materialized his Popular Education pedagogy into a literacy method. For him, literacy is not useful only to read words, but to read the world. And imbued precisely of this spirit, we proposed an analysis of data literacy based on Freire's Literacy Method. By doing so, we hope to give a small contribution to the democratization of access to information. Data alone do not change the world, but we believe that people that critically understand the reality through data has better tools to transform it.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] J. N. Barato. *Codification/decodification: an experimental investigation on the adult education theory of Paulo Freire.* PhD thesis, San Diego State University, 1984.

[2] C. R. Brandão. *O que é o método Paulo Freire*. Brasiliense, São Paulo, 1985.

[3] J. Carlson, M. Fosmire, C. Miller, and M. S. Nelson. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Libraries Faculty and Staff Scholarship and Research*, 11(2):629–657, 2011.

[4] S. Chignard. A Brief History of Open Data, 2013.

[5] S. M. Corazza. *Tema Gerador: concepção e prática*. Editora Unijuí, Ijuí, 2003.

[6] A. R. Ferraro and D. Kreidlow. Analfabetismo no Brasil: configuração e gênese das desigualdades regionais. *Educação e Realidade*, 29(2):179–200, 2004.

[7] P. Freire. *Pedagogy of the Oppressed*. Continuum, New York, 30 edition, 2005.

[8] M. Gadotti. *Paulo Freire: Uma Biobibliografia*. Cortez Editora/Instituto Paulo Freire, São Paulo, 1996.

[9] A. M. Ghiso. Sistematización: un pensar el hacer que se resiste a perder su autonomia. *Decisio*, 1(28):3–8, 2011.

[10] J. Grey, L. Bounegru, and L. Chambers. *Data Journalism Handbook*. OKFN, 2012.

[11] M. B. Gurstein. Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2):1–7, 2011.

[12] M. B. Gurstein. Why I'm Giving Up on the Digital Divide. *Journal Of Community Informatics*, 11(1), 2015.

[13] N. Huijboom and T. V. D. Broek. Open data: an international comparison of strategies. *European Journal of ePractice*, 1(12):1–13, 2011.

[14] O. Jara. Los desafíos de la educación popular. In *Metodología de La Educación Popular*. Asociación de Pedagogos de Cuba, La Habana, 1998.

[15] C. Nuñez. Educar para transformar, transformar para educar. In *Metodología de La Educación Popular*. Asociación de Pedagogos de Cuba, La Habana, 1998.

[16] D. Schugurensky. *Paulo Freire*. Bloomsbury Publishing, London, 2014.

[17] A. F. Tygel, M. L. M. Campos, and C. A. S. de Alvear. Teaching Open Data for Social Movements - a Research Methodology. *Journal of Community Informatics*, 2015.

[18] A. F. Tygel and C. A. S. de Alvear. The Development of an Information System for the Solidarity Economy Movement. *Journal Of Community Informatics*, 7(1-2):1–12, 2011.

[19] Wikipedia. Data Literacy.

---

[7] http://www.opendefinition.org