



**Universidade de
Aveiro
2020**

Departamento de Comunicação e Arte

**Alice
Silva
Assunção
de Melo**

**Jornalismo pela Perspectiva dos Dados –
Reflexão Exploratória sobre a Literacia
Adequada**

**Journalism from a Data Perspective –
Exploratory Reasoning on Adequate Literacy**



Universidade de Aveiro Departamento de Comunicação e Arte
2020

**Alice
Silva
Assunção
de Melo**

**Jornalismo pela Perspectiva dos Dados – Reflexão
Exploratória sobre a Literacia Adequada**

**Journalism from a Data Perspective – Exploratory
Reasoning on Adequate Literacy**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Comunicação Multimédia, realizada sob a orientação científica do Doutor Vania Baldi, Professor Auxiliar do Departamento de Comunicação e Arte da Universidade de Aveiro, e coorientação do Doutor Adelino Gala, Pós-doutoramento no Departamento de Comunicação e Arte da Universidade de Aveiro, Novas Tecnologias da Comunicação.

To my incredible husband, family and friends.

Examination board

president

Doctor Luís Francisco Mendes Gabriel Pedro
Assistant professor at Universidade de Aveiro

Doctor Jorge Manuel Martins Rosa
Assistant Professor at Universidade Nova de Lisboa

Doctor Vania Baldi
Assistant Professor at Universidade de Aveiro

Acknowledgements

I would like to thank my husband Thiago for his inexhaustible patience and support, and my father Genivaldo and mother Francisca for their care and love even from far away. A special thanks to my family Israel, Rosa, my sister Aline, Moises, Jane, Anderson, and my nieces Sofia, Brigitte, and Laura, for the encouragement along the way.

To my brothers and sisters of faith who have been praying to God for me, my deepest appreciation for your love and friendship. To all my friends, from Brasil and Portugal, you have been a prompt help countless times.

At last, but not at least, I would like to thank my professor Dr. Vania Baldi who trusted in my work from the beginning and Dr. Adelino Gala who gave precious advice and made important contributions to this and future projects.

Palavras-chave

Jornalismo de dados, Big Data, epistemologia, jornalismo digital, cultura digital, literacia digital, Portugal, Brasil.

Resumo

A cultura da era digital articula-se em torno de um ilimitado número de dados gerados por parte de atores e objetos conectados em rede. O Big Data traz uma das revoluções mais emblemáticas à sociedade contemporânea, uma mudança de paradigma que envolve a esfera econômica, tecnológica e cognitiva, e proporciona desafios não só ao acesso à informação, mas à maneira como o conhecimento é obtido e retido. O âmbito jornalístico é um dos mais afetados pela consistência dos crescentes dados. Concomitantemente à digitalização, o jornalismo de dados configura-se como meio para estabelecer critérios de inteligibilidade e qualidade infocomunicacional em um exercício que integra princípios éticos do jornalismo e conhecimentos de estatística e de computação para recolher, tratar e interpretar dados e informar de maneira a enriquecer o debate público sobre fatos relevantes para a vivência coletiva. No entanto, permanecem questões sobre o quão longe o jornalista está apto a chegar no que se refere à análise de dados. Esta investigação sistematiza referências teóricas sobre a prática de jornalismo de dados com deduções baseadas em análise qualitativa de dados gerados através de entrevistas com especialistas na área. O objetivo é contribuir com reflexões para o projeto de ensino superior em jornalismo digital Erasmus+ Post-Crisis Journalism in Post-Crisis Libya: A Bottom-up Approach to the Development of a Cross-Media Journalism Master Program (<https://pages-project.ly/en/>).

Keywords

Data journalism, Big Data, epistemology, digital journalism, digital culture, media literacy, Portugal, Brazil.

Abstract

The digital culture is structured around an unlimited and increasing volume of data generated by as many actors as subjects connected through the network. So-called Big Data provides one of the most emblematic revolutions in contemporary society, a paradigm shift that involves the economic, technological and cognitive fields and challenges not only information access but the epistemic standards created until now. Historically attached to information and knowledge-acquisition, the journalistic field is one of the areas most affected by the incidence of increasing data. Simultaneously with digitalization, a practice called data journalism has been set up as a way to establish intelligibility and information quality criteria in a way that embodies the ethical fundamentals of journalism, statistics and computing skills to gather, treat and make sense of data to inform properly and therefore enhance the public debate about relevant facts. However, there remain questions about how far can a journalist go on data analysis. This research systematizes some of the main theoretical references on data journalism with deductions based on qualitative data from interviews with specialists in the field to support with insights the Erasmus+ Post-Crisis Journalism in Post-Crisis Libya: A Bottom-up Approach to the Development of a Cross-Media Journalism Master Program (<https://pages-project.ly/en/>).

Summary

1. Introduction	14
2. Literature review	18
2.1 Data as knowledge mediator.....	19
2.2 Data journalism.....	23
2.2.1 Cautious side.....	27
2.3. Data literacy – lessons to be learnt.....	29
2.4 Considerations.....	34
3. Data journalism examples	37
3.1 Data journalism practice in other regions, and circumstances.....	44
3.2 Transparency.....	49
3.3 Coronavirus in numbers.....	50
3.4 Considerations	58
4. Methodological trajectory	59
4.1 Interviews.....	65
4.2 Field observation: restrictions and future work.....	67
4.3 Considerations.....	68
5. Results	70
5.1 Confrontations between theoretical concepts and evidence.....	82
5.2 Analysis of the results.....	85
5.2.1 Perception of Big Data implications.....	88
5.2.2 Content characteristics.....	90
5.2.3 Educational provision insights.....	96
5.2.4 Perception of data literacy among professionals.....	100
5.2.5 Credibility associated to data journalism.....	105
5.3 Considerations.....	109
6. Conclusion	112
7. References	117

8. Appendices	123
Appendix 1 – Transcription of the interview with Dr. Alberto Marques.....	123
Appendix 2 – Transcription of the interview with Gabriel Zanlorenssi.....	133
Appendix 3 – Transcription of the interview with Rui Barros.....	138
Appendix 4 – Transcription of the interview with Fabio Takahashi.....	148
Appendix 5 – Transcription of the interview with Sergio Spagnuolo.....	154
Appendix 6 – Transcription of the interview with dr. Suzana Barbosa.....	160
Appendix 7 – Federal University of Bahia journalism degree reformulation.....	173
Appendix 8 –Catholic University of Brasilia data journalism course syllabus.....	184
Appendix 9 – Catholic University of Brasilia digital journalism course syllabus.....	185
Appendix 10 – University of Minho official note.....	187
Appendix 11 – Email exchange with Público newspaper.....	188

List of Tables

Table 1 - Analysis Model.....	62
Table 2 - Research consistency method.....	64
Table 3 - Analysis model/categories.....	72
Table 4 - WebQDA codes table.....	74
Table 5 - What is the relation between educational background and data literacy?.....	85
Table 6 - What is the relation between educational background and data literacy?.....	88
Table 7 - What is the journalist's perception of Big Data implications in their work?.....	89
Table 8 - What is the nature of data driven content produced?.....	91
Table 9 - What is the nature of data driven content produced?.....	95
Table 10 - What is being offered in Brazil in the academic context in terms of data literacy?.....	99
Table 11 - What is the perception of among professionals of journalist's data literacy?.....	100
Table 12 - What is the perception of among professionals of journalist's data literacy?.....	101
Table 13 - What is the perception of credibility associated with data journalism?.....	107

List of figures

Figure 1 - Screenshot of one of the stories resulting from The Guardian UK Riots 2011 project.....	40
Figure 2 - Screenshot of NYT's The Upshot front page.....	41
Figure 3 - Screenshot of the NYT's The Upshot subsequent highlights.....	42
Figure 4 - Screenshot of the data journalism webpage of Zeit Online.....	44
Figure 5 - Screenshot of the Folha de S. Paulo's Delta Folha webpage.....	45
Figure 6 - Screenshot of the Gráfico webpage hosted by Nexo newspaper.....	46
Figure 7 - Screenshot of the Núcleo highlights.....	47
Figure 8 - Screenshot of one the sections of interactive Broken Homes project.....	49
Figure 9 - Screenshot of analysis made by Tomas Pueyo in "Coronavirus: Why you should act now" for Medium.....	52
Figure 10 - Screenshot of an extract of the Núcleo's story showing part of the analysis and the links to each source used in the piece.....	53
Figure 11 - Screenshot of an extract of the Nexo's story about hospital bed provision in Brazil.....	55
Figure 12 - Screenshot of an extract of The Guardian's piece about hospital beds in the US.....	56
Figure 13 - Screenshot of an extract of Público's interactive about the spread of coronavirus	57

1. Introduction

In *Facts are Sacred*, Simon Rogers comments on the general feelings of distrust towards news sources, a condition, if not created, at least aggravated by the increasingly multichannel environment, in which countless sources capture the audience attention, but not always with reliable information or responsible interpretation of the facts (Gray et al., 2012). If on one hand, journalism faces threats to its status as the mediator of information and, why not, of knowledge, on the other one it might thrive under the circumstances as long as it consciously embraces the transformations related to the production and dissemination of information.

In other words, amidst this information deluge from multiple sources (Gray et al., 2012; Rogers, 2013), the journalistic practices driven by the visualization and analysis of data might be one possible resource to tackle the distrust that follows misleading information. Data journalism might help since it presents relevant stories supported by scientific methodology and, last but not least, the journalistic fundamentals of accuracy, scepticism, criticism, fairness, objectivity and accountability.

Naturally there is no clear definition of what precisely constitutes data journalism, as the term emerged in the late 2000s. Some theories describe it as a practice that results in pieces with visualizations (Knight, 2015), while others focus mostly on the process (Howard, 2014; Rogers, 2011; Fink & Anderson, 2015) as the creation of news is based on the analysis of a large amount of data. Data journalism advocates seem to agree that a piece data journalistic work is inevitably connected to the analysis of statistics and the creation of visualizations. Yet, the instinct for news continues to be a central element even when dealing with datasets (Gray et al., 2012)

Although the term reached the mainstream only really in the last decade, data journalism evidently is not new. In fact, it might be seen as an evolution from other practices in the field, such as Computer-Assisted Reporting (CAR), a pioneering form of journalism based on data, mainly used in the United States to predict election results (Howard, 2014). CAR, as well as data journalism, has been used mostly in investigative journalism (Felle, 2016). Data journalism still embodies some CAR features, but, in a digital culture mainly dominated by Big Data, it also offers the audience a range of visualizations, a new storytelling resource (Felle, 2016), which is the result of analyzing a great deal of data that requires computing support in order to first understand it.

A possible approach to the premise of data journalism is to consider the Big Data reality, the context in which data journalism work is placed, the proper literacy to practice it and its contribution as one of the resources of digital journalism to recover interest and credibility in press work.

To understand the implications of data journalism, a process still in progress, it is necessary to reflect on Big Data. The audience not only consumes information from multiple sources but also provides information to an increasing network of information / knowledge. Therefore, Big Data can be understood as a technological, scholarly and cultural phenomenon (Boyd & Crawford, 2012). It is technological because it engages algorithm and computation efficiency in order to manage data. The scholarly aspect applies to the use of data sets to make social and scientific claims, while it is a cultural phenomenon because of the understanding that large data sets provides a superior form of knowledge.

With reference to the level of data literacy required to be a data journalist, scholars agree that this is one of the main challenges. In a digital reality where information is continually produced and consumed, journalists face the task of improving their methods of verifying facts. Therefore, a data journalist able to collect, analyse and visualize large datasets through statistics, social science and computing skills

might fill the gap (Howard, 2014; Nguyen & Lugo-Ocando, 2016). At first glance, these features do not seem natural in journalism practice.

It is worth, then, emphasizing that data journalism work is not limited to social science and computing techniques, but covers different methodologies to address new source origins, enabled by the evolution of the relationship between humans and computers (Hammond, 2017), in order to still make sense of facts and present it to the audience. For this reason, new skills in themselves are not sufficient to do the work and it is crucial to maintain a critical view and consider the data, in several cases numbers, as part of a structure created under human consciousness for preconceived purposes (Gray et al., 2018; Godler et al., 2020).

Regarding the interest and credibility aspect, there is an opportunity to elevate the reputation of the press by using data journalism, as long as some features are kept in mind. The first one is that not all data has quality (Berti-Équille & Borge-Holthoefer, 2016) as it can be both created and analysed within misleading, incorrect, outdated and, therefore, unreliable contexts. Another important aspect is that transparency, which should be natural to any journalistic practice, cannot be preserved if the data journalist deprives the audience of access to the process that influenced the creation of a particular data-driven report (Appelgren, 2017; Allen, 2018). Moreover, data journalism loses its reinforcement credibility potential if the journalist has not a certain autonomy over the content and analysis particularly of the unusual or dramatic element which, in some cases, bring additional colour to the narrative (Cushion et al., 2017).

Therefore, this research seeks to present a considered discussion of data journalism characteristics rather than propose a definition that covers a practice still in development. The following chapters are organized into: the review of the literature; the main data journalism examples; details of the empirical research and the conclusion. Additionally, the purpose is to support the Erasmus+ Post-Crisis Journalism in Post-Crisis Libya: A Bottom-up Approach to the Development of a Cross-Media Journalism Master Program (<https://pages-project.ly/en/>), a higher

educational initiative co-funded by the EU Commission, with a theoretical discussion, considerations on the development of data journalism practice in different circumstances, and with insights into data literacy based on qualitative interviews with data journalists, data scientist professionals and professors in the digital journalism field.

The literature review section is supported by explorations on the concepts of Big Data, data journalism and data literacy. In the chapter on the main data journalism examples are considerations of significant data journalism projects and experiences around the world.

In the chapter on methodology, the methodological trajectory and procedures adopted in this qualitative deductive research used to verify the hypothesis are described, and a suggestion is made for a methodological approach in future work. In the chapter of results, the results are presented and analysed through the concepts explored in the literature review that supported the indicators used in both the collection and analysis of data.

2. Literature review

This literature review addresses how data journalism reset journalistic production by presenting a discussion between different theoretical references. The concepts explored are Big Data, considered in “Data as knowledge mediator”, data journalism, addressed in the second subtitled section, and data literacy, discussed in “Data literacy – lessons to learn”. Moreover, some considerations are made at the end of this chapter in order to organize the different perspectives presented in this discussion.

It is known that the rise of the digital age has led to a modification of epistemic standards in knowledge-acquisition. Traditionally knowledge was primarily acquired through observation and the verification of hypotheses, or even from the testimony of others, whereas nowadays data play a decisive role when obtaining knowledge since they provide the opportunity to correlate information from several different sources. Therefore, the processing of knowledge occurs not only through natural or physical, but also technological, agents (Godler, Reich, & Miller, 2020) and this assumption applies to different knowledge fields, especially journalism, which historically has been a means of creating knowledge through information. It is plausible, then, to affirm that the journalist’s understanding of source and evidence has been particularly altered by the establishment of technologies related to Big Data.

Thus journalism practice has been evolving while following new realities where massive datasets are placed as central figures in the production and acquisition of information. Therefore, journalism has now begun an attempt to master the extensive sources of information and even cross reference them with one another. In other words, data journalism has been consolidating a practice that combines scientific methodology, statistics and computing skills in order to make sense of large datasets.

Naturally, this form of news production has evolved from what has been done in the past to deal with a new big scale scenario, and, as with any reshaped format, it struggles to find its place among professionals. In his *The Curious Journalist's Guide to Data*, Jonathan Stray mentions the complex side of data journalism, a multiple skilled nature which is increasingly difficult to assemble (Stray, 2016). Journalists already face the challenge of dealing with subjects which, in the best of circumstances, are familiar to them, and then having to make the correct interpretation and correlation of the facts.

In the Big Data context, surrounded by coding, statistics and math elements, the professional has again to first become familiar with it and then tell the story. There is, however, a significant difference between news production from the past and the present: much more information generated by multiple channels in much less time, a condition that definitely creates opportunities, but also vulnerabilities.

2.1 Data as knowledge mediator

While knowledge has been primarily achieved through observation and the verification of hypotheses, or even from the testimony of others, data have been playing an increasingly decisive role in obtaining knowledge since it provides the opportunity to correlate information from several different sources. With the assumption that knowledge can also be a technological agent (Godler, Reich, & Miller, 2020), it is plausible, then, to suggest that the journalist's understanding of source and evidence has been particularly altered by the establishment of technology related to Big Data. Hence, it is necessary to assess aspects of Big Data in order to then observe the path that journalism is taking in order to address data as a major player in news production. Although a journalist cannot handle a massive and complex dataset like Big Data alone, as suggested by Splendore et al. (2016), it is necessary to understand at least the basis of Big Data to accomplish the mission of fully informing with data-driven content.

“The world has changed and it is data that has changed it” (Rogers, 2013, chapter 2, p.8), states Simon Rogers, award-winning data journalist, in *Facts are Sacred*, about the relevance of data journalism and the global context in which it is framed. This changed world is related to the digital footprints of several personal, social and business activities that occur on digital platforms (United Nations, 2019), which resulted in the Big Data age, a broader range of data increasingly available to individuals, organizations and corporations, and that are used for many different purposes. One theory basically describes data as unfiltered symbols resulting from several activities and inputs that can be processed through a series of transformations into information, which in turn can be used to support decisions made by individuals and organizations and contribute to the creation of thinking models and knowledge (United Nations, 2019). Still from a constructive perspective, Cukier & Mayer-Schönberger (2013) describe Big Data as an ability of groups and individuals to make use of information to produce insights and products that are useful and significant.

However, Big Data has also a controversial side. Clearly, not all data is used to generate value, meaning and knowledge as some of them are of low quality while failing to combine consistency, freshness, completeness and accuracy (Berti-Équille & Borge-Holthoefer, 2015). More realistic theoretical perspectives describe Big Data as a technological, scholarly and cultural phenomenon. It is technological because it engages algorithms and computational efficiency in order to manage data; scholarly because it applies to the use of datasets to make social and scientific claims; while it is a cultural phenomenon because there is an understanding that large datasets provides a superior form of knowledge (Boyd & Crawford, 2012).

An undeniable feature of Big Data that encompasses both the positive and realistic perspectives, relates to its substantial volume as “datasets whose size is beyond the ability of typical database software to capture, store, manage and analyze [...] as technology advances over time, the size of datasets qualified as Big Data will

also increase” (Manyika et al., 2011, p.1). Laney (2001) makes an interesting correlation to the word “big” present in the term, which can refer to the volume of data available, the wider variety which might require combinations and the velocity with which the data can be generated and interpreted.

The Big Data age added new and stronger shades to the journalistic field in the late 2000s as governments and companies begin to give the public access to their massive dataset. Rogers (2013) attributes the widespread data openness from governments around the globe to the Obama administration. One of the first legislative actions approved by the former US president was the creation, in 2009, of a website with thousands of datasets related to multiple subjects: crime rates, FBI data, population estimates and other topics.

If on the one hand governments appear to be on board for publishing data since this attitude makes them seem transparent, on the other companies embrace data usage because they seem to understand that a more transparent business profile provides larger revenues (Rogers, 2013). Therefore, it is reasonable to think that, according to the motivation, data openness can be wider or narrower:

While the US Library of Congress has an enormous amount of data on its hard drives – 235 terabytes up to April 2011 – that pales into insignificance besides the amounts stored by companies. According to the last figures, 15 out of 17 US business sectors have more data than this stored per company (Rogers, 2013, chapter 5, page 1).

Besides the understanding that being open and transparent might reinforce the public’s trust rather than increase competition, companies also count on another positive effect: business performance can be enhanced through the management and analysis of data stored (Rogers, 2013). Nevertheless, there are other implications to business as the data openness movement has led companies to take action in order to prevent leaks. Rogers (2013) mentions, for instance, the Bank of America which created a 20-person team with lawyers and specialists just in case WikiLeaks released information they claimed to hold on the bank.

Naturally, Big Data reality also raises concerns about privacy, the decrease of civil liberties and an increase in state and corporate control (Boyd & Crawford, 2012). A 2019 global survey on Internet security and trust conducted by the Centre for International Governance Innovation (CIGI) and Ipsos institute, in collaboration with Unctad and the Internet Society, found that 78% of Internet users in 25 economies were at least somewhat concerned about their privacy online (*CIGI-IPSONS Internet Security & Trust*, 2019). The many questions about how far someone can go with personal data led to the creation of the General Data Protection Regulation by the European Commission, which basically aims to protect the fundamental rights and freedoms of people in the European Union and their right to the protection of personal data (Union, Journal, & Union, 2018).

Alongside promising opportunities and reasonable fear, Big Data has undoubtedly changed the knowledge-acquisition process. Hammond (2017) emphasizes Big Data proponents claim that the 'datafication' phenomenon might improve the individual's understanding of the world and thereby enable people to make better decisions. Although with still uncertain implications, Big Data has been leading new discussions about the nature of knowledge, some of them yet to be better explored. Meanwhile, making sense of huge information sets is even more important than just the data themselves. Without context, data are meaningless (Boyd & Crawford, 2012).

The journalist and information architect Mirko Lorenz (2012) refers in his *Data Journalism Handbook* to a Big Data movement in which tech companies use software packages to dig through piles of data in order to extract insights. If the data age transforms knowledge-acquisition (Godler et al., 2020), it is key, then, to make sense of the increasingly available information, to make data visible and meaningful. Clearly, though, the "wealth of information creates poverty of attention and the need to allocate that attention efficiently among the overabundance of information sources" as Simon argues (as cited in Rogers, 2013). This means that although data are a rich resource, they have limitations and potential threats.

Rogers (2013) alerts us to the need of keeping expectations about data proportional to what they really represent as opportunity. In other words, it's fundamental to have a critical attitude towards data. Their value rest mostly on the structure in which they were created and the human capacity to give purpose to them. Data must also be faced with skepticism since "they are representations of the world gathered, generated, selected, arranged, filtered, collated, analyzed and corrected for particular purposes [...] they often serve as a starting point or a supplementary source for understanding a topic" (Rogers, 2013, chapter 9, p. 3).

Most of all, it is necessary to understand that having access to data is not the same as having the power to make the change, says Rogers (2013):

Democratization of information is different from the democratization of power. Knowing that something is wrong or there is a better way of doing things is not the same as being in a position of fixing things or to effect change. If you want to change laws, policies or practices it usually helps to have things like implacable advocacy, influential or high-profile supporters, positive press attention, hours of hard graft, bucketloads of cash and so on. Open information about government is not the same as open government, participatory government or good government. While it's easier than ever before to do things with data on computers and on the web, this does not necessarily entail that it's easier to know what a given dataset means. In this sense it isn't any easier to be a good data journalist than it's to be a good journalist, a good analyst, a good interpreter. Creating a good piece of data journalism is more an art than a science. I'm sure as time goes by, we'll have a more balanced and critical appreciation of the value of data and its role within our information environment (Rogers, 2013, chapter 9, p. 4-5).

2.2 Data journalism

In order to make sense from massive datasets and keep its mission to provide qualified information, journalism is evolving into a practice in which the main source is the data. It is, therefore, a practice that combines the abilities and tools to deal with data in order to first visualize it then explain the information through a story. Although some scholars define data journalism as pieces with visualizations

(Knight, 2015) or graphic information (Lowrey & Hou, 2018), the format of data-driven news can also vary from interactive charts to solely articles (Rogers, 2013). But the main feature of this kind of journalism is precisely related to its process, which encompasses different skills and roles from one reality to another (Fink & Anderson, 2015), as the creation of news is based on the analysis of a large amount of data. Data journalism advocates agree it is a practice that involves statistical analysis skills, basic programming knowledge and, most importantly, journalistic instincts (Gray, Chambers & Bounegru, 2012).

There remain, however, all sorts of aspects surrounding data journalism that need to be clarified as the practice is still evolving and its strengths and weaknesses are still being appraised. Although the widespread availability of information and the consequent growing interest in visualizing it promoted data journalism (Gray et al., 2012), the relationship between data and journalism dates back to the 19th century, as Howard (2014) reminds us, when Dow Jones & Company already published share prices to investors through newspapers.

Data-driven journalism can be seen as an evolution from other practices in the field, such as Computer-Assisted Reporting (CAR) - pioneering journalism based on data initially used in the United States to predict election results (Howard, 2014). Although data journalism still embodies some CAR features, as the method of investigation is still fundamentally the same (Felle, 2016), it differs from the latest journalism practices mostly because it offers the audience a piece of information which is the result of analysis of a great volume of data. This merging of investigative with new storytelling resources (Felle, 2016) is the most differentiated feature as the current digital culture is mainly dominated by Big Data.

This shift in journalism follows the remodelling of the roles that computers and data now play as an active party in every human relationship, as argued by Hammond (2017). "The more important change is in how we think about journalism and that change is not simply driven by new technologies but needs to

be understood in broader contextual terms” (Hammond, 2017, p 409). Additionally, if one reason that justifies the need for data journalism is the increasingly wide information environment that requires professional help to make sense of it, another one might be the digital culture which enables journalism to become transparent through giving the audience access to a portion of complex data. “In the first scenario, data is opaque and confusing, requiring special expertise to interpret it and turn it into meaningful knowledge; in the second, data is transparent and the only requirement is that we be given access to it” (Hammond, 2017, p 412).

There is, however, a delicate question towards data access given to the audience that refers to the fact that it has already been submitted to journalistic filtering during the news production cycle. In charge of the production and of what should be revealed and what should not, the data journalist might fall into the temptation of being what Appelgren (2017) classifies as ‘a paternalist’. This paternalistic side of data journalism suggests that the idea of a public access to the data itself might be misunderstood, since the journalist not only mediates but regulates what should disseminated.

Naturally, critical control over the information flow is inherent in journalism and this paternalistic behavior is adopted in journalism practice in general. Nevertheless, data-driven journalism can bring new aspects since journalism and technology professionals work together on a pre-defined angle and technological functionalities towards the content to be produced (Appelgren, 2017). Therefore, continues Appelgren (2017), the sense of control over raw data might be false since the audience are still being guided through a technological set previously designed to develop expected actions not fully consented to by the user.

Despite not being given full access to raw data itself, data journalism still plays an important role in a reality where the rich informational environment works for good purposes, but also for the bad ones. Not all the information available is reliable. In

fact, Rogers (2013) comments on the general feeling of distrust towards news sources at the current time when news organizations have to fight with other minor, but numerous, sources “for their very existence” (Rogers, 2013, chapter 2, p. 6). That scenario reflects the current multichannel environment where attention is captured through all kinds of information, but trust is an increasingly scarce resource (Gray, Chambers, & Bounegru, 2012).

The digital age also brought a news production cycle with a more irregular pattern due to an informational environment where there is always a novelty to report and consume (Bakker, 2012). Amidst this information deluge from multiple sources, data journalism might be a resource to tackle again the distrust that follows misleading information since it presents/ provides relevant stories supported by data and the journalistic fundamentals of accuracy, skepticism, criticism, fairness, objectivity and accountability:

Data journalists can help to collate, synthesize, and present diverse and often difficult sources of information in a way that gives their audience real insights into complex issues. Rather than just recycling press releases and retelling stories they've heard elsewhere, data journalists can give readers a clear, comprehensible, and preferably customizable perspective with interactive graphics and direct access to primary sources. Not trivial, but certainly valuable (Gray et al., 2012, p. 59)

The reconfiguration of purposes related to news coverage and the format of it is a continuing reality. For now, when it comes to hard news coverage, alongside the purpose of being the first one, another goal is settled: to be the one telling what any given event might actually imply while creating new data based storytelling products. It is about giving proper context and quality to the debate (Gray et al., 2012). Essentially, data journalism has the potential to reinforce credibility and interest in journalism in general because it offers a way to see world events that couldn't otherwise be observed:

Connections between powerful people or entities would go unrevealed, deaths caused by drug policies would remain hidden, environmental policies that hurt our landscape would continue unabated. But each of the above was changed because of data that journalists have obtained, analyzed, and provided to readers. The data can be as simple as a basic spreadsheet or a log of cell phone calls, or complex as school test scores or hospital infection data, but inside it all are stories worth telling (Gray et al., 2012, p. 10).

2.2.1 *Cautious side*

Despite their potential democratic force, Cushion, Lewis and Callaghan (2017) alerts us to the risk of data, as statistical inputs end up being used more to emphasize the unusual or dramatic than bring context or detail to news coverage, especially in the case of crime and terrorism.

Certainly, it is part of journalism to tell a good story, but the storytelling aspect of journalism must not surpass the fundamentals of objectivity and accuracy, otherwise it would lead to a misinformed view of the world, as warned by Cushion et al (2017).

They examined all statistical references used in reporting on several BBC media platforms and commercial television newscasts (ITV, Sky and Channel 4) from 12 October to 8 November 2015. The study revealed that situations where statistical information was deployed to bring context in order to understand the story were the “exception rather than the rule” (Cushion et. al, 2017, p. 1212). Moreover, if on the positive side there is indeed a more developed statistical expertise among the news organizations, on the negative one a less auspicious reality shows that this resource it is not well leveraged as a more traditional style / form of storytelling tends to prevail (Cushion et al., 2017).

It is undeniable that open data movement led by governments and companies paves the way to a kind of journalism in which the main sources are the data.

Nevertheless, this reality brings a collateral effect: journalists still rely on the same sources as before and, as argued by Cushion et al. (2017), these political, business and governmental sources also limit the range of the data in news coverage. According to the Cushion et al. (2017) study on BBC media platforms and commercial television newscasts, universities were the statistical sources in less than 1 in 10 data-driven stories. To reverse this reality, then, it would be necessary to set a broader understanding of what is news gathering while keeping some distance from classical patterns. “Three key areas— diversifying the range of everyday sourcing in news reporting, rethinking the practice of impartiality and questioning news values that stress drama over context—warrant further reflection”,(Cushion et al., 2017, p. 1212).

The practice of data journalism is an undeniable, but changeable, reality in the news coverage process. It has become a necessary tool in the contemporary digital culture where the informational environment provides every kind of data and sources, from reliable to misleading ones, to the audience. And mainly because of its transparent and high qualified information nature, data journalists must be assured of a certain editorial autonomy in order to not only make sense of data but also to question the data itself (Cushion et al., 2017).

Although numbers and statistics play an important role in data journalism, journalistic work still has to dealing with local culture, customs, laws and politics in order to find sense, as Stray (2016) argues. Yet, adding data skills and methodology to the news gathering process can help investigative journalists to find trends and explore hunches and hypotheses (Howard, 2014), related to this very same local culture, customs, laws, politics and world events.

On the bright side, media organizations could employ data journalism as an asset by selling their access to archives and APIs. This might be an alternative in a reality where few media organizations still rely on subscriptions as their main source of revenue while new models of low cost or even free content production

are increasingly dominating news production (Bakker, 2012). However, with the opportunity also comes challenges and the prime one is security of information:

In general, journalists must consider when it's appropriate to scrape data, access data, store it— or not. Does the story require storing personal information? If so, such sensitive data will need to be protected with the same vigor that journalists have protected confidential sources. Unfortunately, the information security practices of many media companies are not as robust as they will need to be to prevent determined intrusions by organized crime or nation states (Howard, 2014, p. 75).

As traditional journalism depends on the critical checking of facts and good quotes in order to communicate effectively, so data-driven journalistic content must result from good and clean, accurate and significant information (Gray et al., 2012). On this matter, Rogers (2013) affirms that data journalistic work as “a bridge between data and audience” is 80% perspiration, 10% inspiration and 10% output.

The critical perspective inherent to any journalistic work cannot be neglected. As happens in any case of good narrative, the journalist is again able to guide the audience through a complex story (Howard, 2014). Besides the technical skills, data journalists must reinforce ethics and security features on data-driven news production. The ability to publish data does not exclude the underlying responsibility with regards to the content, the sources involved and the audience. This means that not all kinds of data should be published, as Howard (2014) argues: “Does the story require storing personal information? If so, such sensitive data will need to be protected with the same vigor that journalists have protected confidential sources” (Howard, 2014, p. 75).

2.3 Data literacy - lessons to be learnt

If on the one hand a critical perspective is still not negotiable in the practice of data journalism, on the other there is an undeniable need for proper data literacy able to integrate basic computing and statistical skills. In fact, this data literacy should

be applied to other areas beyond the field of journalism as an answer to deal with the global process of “datafication” which encompasses all models of knowledge (Gray, Gerlitz, & Bounegru, 2018). However, those skills might prove a barrier to a journalist with no background in these fields. One primary difficulty would be to achieve a common understanding of what the appropriate level of data literacy for a data journalist should be. From a simpler and immediate perspective, an appropriate level of data literacy should be one that enables journalists to access, analyse, manipulate and deliberate on data.

As Nguyen and Lugo-Ocando (2016) emphasize, to master basic statistics competences has become “the journalist’s duty” since a Big Data society has been emerging and daily news work is increasingly supported by “number crushing”. Another reason for developing statistician skills and adding this to a critical perspective is the growing volume of inconsistent information, with fake news supported by numbers which again Nguyen and Lugo-Ocando (2016) called “statistics-based lies”, created from inappropriately produced data:

A basic competence in statistical reasoning would enable and empower journalists to harness the power of good data – namely their capacity to summarize patterns, to depict and predict trends and to identify possible causes and effects – for news purposes. Used wisely and properly, good data allow journalists, among other things, to overcome the sometimes tyrannical rule of intuition, to link anecdotal evidence to the big picture, to bust myths, to challenge prejudices, to connect seemingly distant incidents/events and to turn ostensibly boring, soulless numbers into vivid, meaningful representation of the world (Nguyen & Lugo-Ocando, 2016, p. 6).

Although still persists some accommodation with the lack of crucial skills to deal with data, as identified by Hewett (2016), this perspective tends to lose its strength while a more conscious attitude towards new methods of producing news is emerging. A study conducted by Heravi (2019) with 260 participants from 43 countries on who is teaching data journalism, who is learning and what is being taught and learned, shows that 98% of the journalists surveyed expressed interest

in acquiring skills to practice data journalism, while 81% reported being “very” interested. This perceived willingness might be the imitation effect considered by Splendore et al. (2016) in which an increasing number of journalists now aware of the data journalism reality start to adopt this form of newsgathering and, to do so, seek for adequate level of data literacy.

Still on the subject of literacy, there is a realistic approach which provides a broader relationship between data and its context. Data infrastructure literacy (Gray et al., 2018) is a critical understanding of the infrastructures through which data is created. This means that datasets are not totally free from bias and deliver a world in conformity with a given vision, culture and values. Even information created through technology embodies the human consciousness of its creator (Godler et al., 2020). As well as data creation, human consciousness is also present in data gathering, a process which Berti-Équille and Borge-Holthoefler (2016) reinforce can also be misleading, incorrect, outdated and, therefore, unreliable.

This understanding of data as a structure created by humans might help journalists to avoid giving in to the false authority of data, as argued by Appelgren (2017), in which the quantitative result tends to be considered as unquestionable. As a matter of fact, with regards to any journalistic production, it is central to first understand the data/information conception context and then make it clear to the audience. Once the process that leads to the story is not clear to the readers, then they are being deprived of a certain transparency to which they have the right (Appelgren, 2017).

As data journalism practice set its place, training initiatives start to form around the world as a response to the “datafication” described by Gray, Gerlitz and Bounegru (2018). In Europe, for instance, courses in United Kingdom, Italy, Germany, the Netherlands, Poland and Switzerland tend to focus on how to collect, analyse and present the data, while non-academic but well established institutions, such as the

European Journalism Center, have become references in the educational movement (Splendore et al., 2016). The European Journalism Center has offered data journalism training for more than 21,000 participants from 170 countries.

It is clear that technological systems, specially enhanced by the digital age, change the current epistemic standards (Godler et al., 2020). However, data journalism training still has a consolidation way ahead. Hewett (2016) alerts us to the fact that education has been responding slowly to the digital progress in news coverage. Heravi's (2019) study revealed that only 25 from 219 data journalism courses around the world provide complete programs since the remaining are basically modules on related topics. Yet, of these 25 complete data journalism courses, a smaller part are non-full degree, postgraduate certification programs and the remaining are mostly Master's degree offerings. Of the 219 courses analysed, 148 from the total are offered by institutions in the United States, 8 in Canada, and another 63 outside North America. The results exclude online courses. Apart from North America, the United Kingdom, the Netherlands, Ireland and Australia are together responsible for most of the offerings.

In terms of what is being taught, in European countries, the most used examples of good data journalism practice are The New York Times and The Guardian (Splendore et al., 2016). On the technical side, data journalism and Computer-Assisted Reporting (CAR) courses are currently the majority in training initiatives around the world, which means 65% of all offerings analysed by Heravi (2019). "Such courses have their focus on a complete workflow of data journalism, from finding, collecting and cleaning data, to analysis, visualization and communication, without diving deep into more complex topics or tools", (Heravi, 2019, p. 363).

Naturally, data journalism is still a discipline in progress as the technical term reached the mainstream mostly in the last decade. As a recently established practice, certainly it has space to grow, especially with regards to who is teaching

it. Heravi (2019) found that 40% of instructors have PhD level, another 34% have a Master's degree and 26% completed Bachelor's level:

This shows that while many academic/university courses are emerging in this area, not many instructors have the normal academic credential in this area, which, considering the age and (lack of) maturity of the field as an academic discipline, is not surprising [...] many of these courses are taught by professionals, or by academics with Master's level education, which is unlike many other academic disciplines such as social sciences or data and computer science, (Heravi, 2019, p. 364).

Besides different professional and educational backgrounds, data journalism professors also have different approaches. However, one of the main challenges in teaching data journalism might be to disprove any preconceptions that can exist about this discipline:

It is widely recognized that journalism students do not typically begin their education with an innate interest in numbers [...] Teachers of data journalism will be familiar with the complaint from students that 'I'm not good at math's' or, perhaps more worryingly, 'I'm not good with technology' (Bradshaw, 2018, p. 56).

2.4 Considerations

The beliefs about data journalism are still being formed as the digital age continues evolving. However, for now it is possible to define some aspects of this practice. Firstly, it is in course a shift in the understanding of what constitutes a primary source in news coverage since computers and data now play an active role in every human relationship (Hammond, 2017). Also, it is visibly a reconfiguration of purposes related to hard news coverage and format, since the purpose is not only to be the first, but to be the one explaining what a dataset might actually mean, while creating new storytelling products (Gray et al., 2012). On the other hand, although it is part of the journalistic job to tell good stories, Cushion et al. (2017) alerts us to the risk of misuse of data to emphasize the

dramatic instead of bringing context, which can happen when the principles of accuracy and objectivity are neglected to promote the unusual.

While auspicious, data journalism needs to be treated with caution. Appelgren (2017) alerts us to a side effect of data-driven stories which is the paternalistic face of the data journalist. As happens in most journalistic productions, the journalist is still in charge of making the decisions related to the data-driven content to be created. That is to say the journalist might overdo their control over the audience. Certainly, regulation is not a new facet of journalism and part of the job is indeed to act as a gatekeeper between the source and the readers, but the paternalistic side, according to Appelgren (2017), is exactly how much use the journalist makes of his authority to deprive the audience of getting in touch with the facts, in this case the data.

Secondly, data journalism entails developing statistical and computing skills, but critical observation remains as a strong characteristic of it, as in any journalistic practice. Essentially, specific skills allied to critical thinking and inquiring make data journalism a potential tool in reinforcing press credibility in a digital age where the news production cycle has a more erratic pattern due to an informational environment overwhelmed by the necessity to always report something and preferably something fresh (Bakker, 2012). Within this environment with extensive but unreliable information, data journalism could fill the gap, being the trusted source of information by collating, synthesizing and presenting diverse and difficult sources of information in a way that gives the audience real insights into complex issues. (Gray et al., 2012). However, data journalists must be assured of certain editorial autonomy in order to not only make sense of data but also question the data itself (Cushion et al., 2017).

Thirdly, with regards to levels of data literacy, it is not expected of journalists to master coding and statistics as well as an experienced professional in these fields would, but to develop new techniques to verify facts and turn them into information

able to mediate knowledge-acquisition. It is basic, then, for the journalist to understand the environment in which the data are placed, as made under human consciousness (Godler et al., 2020) with the purpose of revealing only a given side of the truth. Therefore, data infrastructure literacy (Gray et al., 2018) is essential to ponder the arrangement through which data is created. This means that datasets are not totally free from bias and deliver a world in conformity with a given vision, culture and value, while even extracting information from a dataset can be done in misleading, incorrect, outdated and, therefore, unreliable ways (Berti-Équille and Borge-Holthoefer, 2016).

With regards to the journalism training context, data infrastructure literacy should be taught in terms of the relationship between social and technical infrastructures which permeates data conception. Such understanding of the data as a structure created by humans might help journalists to avoid giving in to the false authority of data, as argued by Appelgren (2017), in which the quantitative results tend to be considered as unquestionable. As a matter of fact, as with any journalistic production, it is central to first understand the data conception context and then make it clear to the audience. Since the process that leads to the story is not clear to the readers, then they are being deprived of a certain transparency to which they have the right (Appelgren, 2017).

Finally, to understand part of this constantly changing reality it is necessary to comprehend the implications of Big Data, the context in which the data journalism work is placed. As Cukier & Mayer-Schönberger (2013) summarize, Big Data can be understood as an ability of both groups and individuals to make use of information to produce insights and products that are useful and meaningful. However, not every piece of data is used to generate value, meaning and knowledge as some of it is of low quality while failing to combine consistency, freshness, completeness and accuracy (Berti-Équille & Borge-Holthoefer, 2015).

More realistic theoretical perspectives describe Big Data as a technological, scholarly and cultural phenomenon. It is technological because it engages algorithm and computation efficiency in order to manage data, while the scholarly aspect applies to the use of datasets to make social and scientific claims, and it is a cultural phenomenon because there is an understanding that large datasets provide a superior form of knowledge (Boyd & Crawford, 2012). While questions and considerations on the nature of Big Data remain open, there is an undeniable feature that encompasses the different perspectives on it, and that is its substantial size - as “datasets whose size is beyond the ability of typical database software to capture, store, manage and analyze [...] as technology advances over time, the size of datasets that qualify as Big Data will also increase”(Manyika et al., 2011, p.1).

A journalist cannot handle the mass and complexity of Big Data alone, as argued by Splendore et al. (2016). Moreover, even journalists with developed statistics and computing skills cannot entirely master the data because no post-industrial system can be accessed completely by the human (Flusser, 1983), which means that access is only to a portion of what software programs and databases offer. However, it is crucial to understand at least the basis of Big Data and even more important to be aware of the shift in the pattern of news production due to the great volume of available information. In other words, alone or in collaboration with other professionals, the journalist will have to deal with non-100% human sources in order to make sense of information and then inform society.

3. Data journalism examples

In this chapter relevant examples of data journalism projects and experiences across the world will be presented. From classical references, such as the American New York Times, the British The Guardian and the German Zeit Online, to sources like ProPublica and newer, but so far, successful experiences in Brasil, Canada, Colombia, Kenya and Palestine. This chapter also proposes a discussion about the transparency level in data journalism projects by taking as reference works from the New York Times, the Washington Post and the Brazilian portal Nucleo.jor.br. At the end of this division, some considerations are also made about the data journalism work during the coronavirus pandemic.

It is not the job of a data journalist to become a professional coder or statistician, but to pursue knowledge for better dialogue with these fields, from which arises the journalistic sources of the digital era. It is clear, then, that data journalism practice joins critical perspective and team effort to provide to the reader context for daily news amidst the rich, and at the same time poor quality, informational environment. In an effort to understand the data journalist profile, it is constructive, therefore, to observe some of the best practices of it around the world. Data journalism reached the mainstream with a Wikileaks episode on Afghanistan war logs. Rogers (2013) considers the episode as a “game changer” for data journalism practice:

There are still reporters who don't want to know about data journalism. But there are other who see this new wave as a way to save journalism. It represents a new role for journalists as bridge and guide between those in power who have data, (Rogers, 2013, chapter 2, p.7).

At the time of Wikileaks, journalists from media organizations combined traditional journalism and technology in order to analyse spreadsheets with more than 92,000 rows of data about a detailed breakdown of a military event in Afghanistan. Simon Rogers was in charge of The Guardian's Datablog at the time and searched for

ways to tell one of the key stories of the war: the increase of Improvised Explosive Device (IED) usage. There were more than 7,500 IED explosions or ambushes between 2004 and 2009, and another 8,000 IED were found and cleared / deactivated. “This data allowed us to see that the south, where British and Canadian troops were based then, was the worst hit area – which backed up what our reporters who have covered the war knew”, (Rogers, 2013, Case Study Wikileaks, p. 3).

With data leaks and efforts from media organizations to analyse it, the Afghanistan conflict was the most documented war in history (Rogers, 2013). The data showed that most of the casualties were civilians. Wikileaks database recorded 109,032 deaths, of which 66,081 were civilians, 15,196 Iraqi security forces and 23,984 insurgents.

A second episode involving the WikiLeaks database on the Afghanistan war was the cables, communication made between 250 worldwide US embassies and consulates via the Secret Internet Protocol Router Network (SIPRNet). This global US military system keeps information separate from the civilian Internet. So, the WikiLeaks database recorded 50,000 SIPRNet documents during the Obama administration (Rogers, 2013).

In the data journalism spectrum there are some media companies which are already driven by data. This is the case with Bloomberg, Thomson Reuters and the Financial Times. It is in the nature of these companies to assess information based on financial data in order to deliver trends, support decisions and offer an economic overview. Although fully integrated data journalism staff in coordination with computing and design professionals, is not common yet in newsrooms (Splendore et al., 2016), some legacy media organizations are embracing data journalism practices for different purposes.

The Guardian, for example, holds the Datablog, a front-page website that provides tools for exploring public spending, trends and information on diverse subjects, such as economics, politics, and social behavior. Although it counts on development and design teams, most of the analysis work is based on spreadsheet matches. The data come from multiple sources and journalistic research. Hundreds of datasets are identified and cleaned, and the work covers various topics from carbon emissions or crime rates to entertainment. “The general approach is to look for the most authoritative place for your data”, (Rogers, 2013, Chapter 10, p.1)

The coverage of the UK riots in 2011 by The Guardian is a work of data journalism that stands out for its partnership with academia. Since UK authorities claimed the riots were mostly caused by incitement on social media, The Guardian, in collaboration with the London School of Economics, set up the *Reading the Riots* project to show how rumors were circulating on Twitter. The team analysed more than 2.6 million tweets provided by Twitter.

The process was divided into two phases. In the first one, to analyse the riots as they happened, the team used basic maps to check the location of confirmed riot spots and matched the information with where the riots actually took place. With this information it was possible to refute the main political excuse that the riots were not linked to poverty. Regarding social media, the group found, through the visualization of hashtag usage, that Twitter was mainly used to repost about the riot rather than to organize it. Through the campaign #riotcleanup, locals organized themselves to clean the places where rioting had occurred. The second phase was centered on working with transport mapping specialists to identify the most likely route traveled by the rioters as they made their way to different locations. This highlighted different patterns for different cities. At the end of the analysis process, Datablog journalists used Google Charts and Google Fusion (now extinct) in order to prepare graphics and other visuals /visualizations quickly (Gray, Chambers, & Bounegru, 2012).

datasets (Royal, 2010). The newspaper was indeed awarded in 2013 for the multimedia feature Snow Fall: The Avalanche at Tunnel Creek, as John Branch, the reporter who wrote it, was successful in integrating text, images, video and graphic animation (van Krieken, 2018).

Yet, the data-driven content at The New York Times appears to be commonly produced by small team focused on “hard news” and tends to present fairly complex visuals, with most of the data from institutional sources (Zamith, 2019). The referenced media has the statistical analysis section The Upshot, which, as described on The New York Times institutional webpage, “provides readers with digestible data, and interactive projects”.

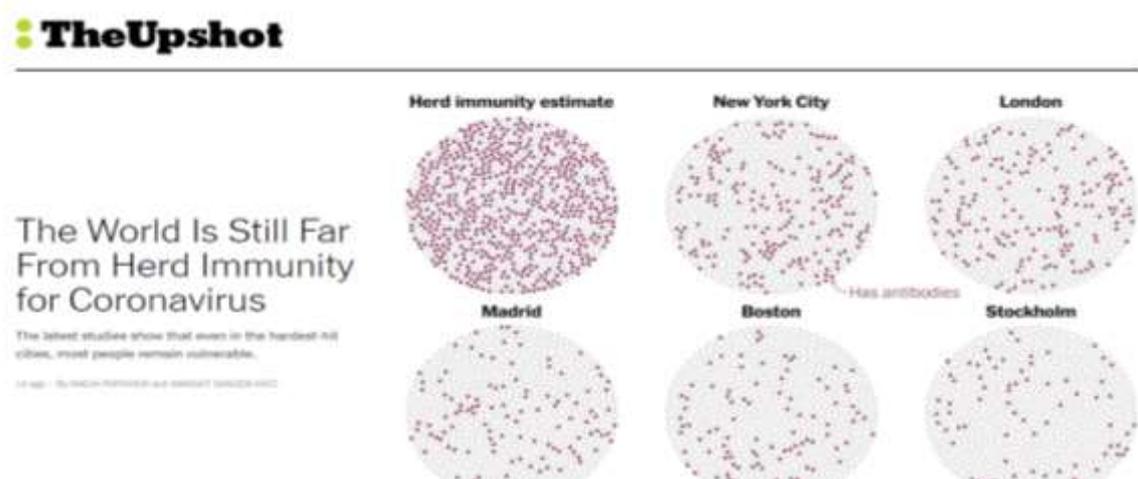


Figure 2 Screenshot of NYT's The Upshot front page. Retrieved from:

<https://www.nytimes.com/interactive/2020/05/28/upshot/coronavirus-herd-immunity.html>, on 05/30/2020 - 14:57.

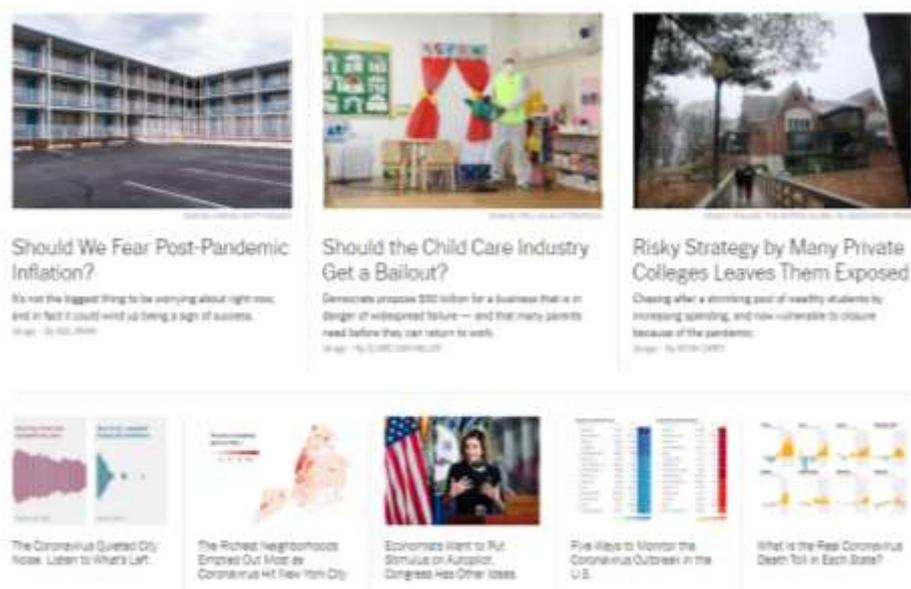


Figure 3 Screenshot of the NYT's The Upshot subsequent highlights. Retrieved from: <https://www.nytimes.com/section/upshot>, on 05/30/2020 – 14:57.

Other media companies, such as the newspapers the Chicago Tribune and the Financial Times, have also dedicated space to data driven stories. The American newspaper built a news app team where hackers work closely with editors and reporters to help on researching, reporting and illustrating stories online. Unlike many teams in this field, this one was founded by technologists for whom journalism was a career change. All app ideas come from the reporters and editors in the newsroom (Gray et al., 2012).

The international Financial Times created a web-page called Visual Journalism with visualizations of the most relevant stories of the year. In 2011 the newspaper started to investigate the private care home industry and exposed how some private equity investors turned elderly care into a profit machine through the elevated costs of a business model that favored investment returns over good care. The inquiry began with analyzing data obtained from the UK regulator in charge of inspecting care homes. The information was public, but it required extensive data cleaning because the reports provided by the Care Quality

Commission, for example, contained categorizations that were not uniform. The work was primarily done using Excel (Gray et al., 2012).

In Canada, specifically in Toronto, data journalist Patrick Cain works on creating maps of interesting and relevant information about the Canadian capital: historic records of everyone who died in the First War; which part of the city has an bedbug epidemic, and where the most guns are. The city's open data movement has been led by the Toronto's website department authority (City Tech). Canada's biggest open data sites (Toronto, Vancouver, Ottawa and Edmonton) have started getting together to set up a series of open data standards - one of them is the G4 Open Data Framework which links cities all over Canada (Rogers, 2013).

In Germany, Zeit Online, the digital publication of weekly newspaper Die Zeit, has been increasing the inclusion of data-driven stories in their news coverage. Since mid-2010, the German website has been developing data projects (Gray et al., 2012). In a comment to the Data Journalism Handbook (2012), Sasha Venohr, Head of Data Journalism and Editor Investigative Team at Die Zeit and Zeit Online, affirms that "at the Zeit Online, we've found that our data journalism projects have brought us a lot of traffic and have helped us to engage audiences in new ways", (Gray et al., 2012, p. 39). The data journalism team aims mostly to contextualize news stories through visual storytelling and with data support, according to Paul Blickle (2013), Data visualization and Infographics Editor at Zeit Online.

"During interaction with the graphic, readership is supported by short formulated pieces. A balance must be created between leading the readers and letting them explore the interactive contextualization—this binds them to the story", (Blickle, 2013, p.1). He adds that the suggestions for stories come from editors from all areas in Zeit Online and from the data journalism team. With regards to the editors in charge of data-driven contents, they must, "ideally" have knowledge of statistics. "Frequently, the ideas for visualizations originate from the editors. As they obtained

data initially, or are editing an issue that is full of statistical data, they are best able to explain the essential idea of their piece”, (Blickle, 2013, p.6).



Figure 4 Screenshot of the data journalism webpage of Zeit Online. Retrieved from <https://www.zeit.de/datenjournalismus>, on 05/30/2020 – 15:17.

3.1 Data journalism practice in other regions, and circumstances

In South America, data have been used mostly to deliver qualified information about government and company practices. In Brazil, the traditional Brazilian newspaper Folha de S. Paulo handles the Delta Folha, a webpage with articles supported by data analysis. Although the primary source is data, the format of stories is very similar to other Folha materials. The project uses as source data that disclosed by government bodies, universities and think tanks.

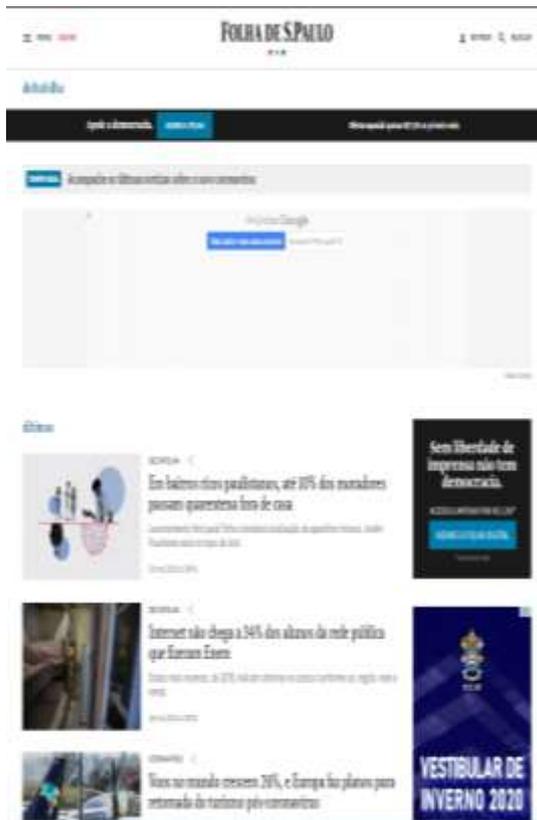


Figure 5 Screenshot of the Folha de S. Paulo's Delta Folha webpage. Retrieved from: <https://www1.folha.uol.com.br/deltafolha/#10>, on 05/30/2020 – 15:24.

Meanwhile, the Nexo Jornal, is a digital newspaper launched in 2015 with the main goal of bringing context to the news and expanding public access to data and statistics. The newspaper already has a team with more than 30 professionals and integrates the Brazilian unit of Trust Project, a global media organization where the goal is to apply transparency standards that help the audience to evaluate the quality and trustworthiness of journalism.



Figure 6 Screenshot of the Gráfico webpage hosted by Nexo newspaper. Retrieved from <https://www.nexojournal.com.br/grafico/>, on 05/30/2020 – 16:00.

Another project which has been bringing quality to data-driven journalism in Brazil is the Volt Data Lab, an agency focusing on journalistic content based on data. It has been serving newsrooms of other media organizations with on demand content related to public administration, the economy, politics and general topics. One of its standout works is the Atlas da Notícia project, a map which shows news production in Brazilian territory, specially from local vehicles. Recently, Volt Data Lab joined in the launch of Núcleo, a digital platform focusing on investigative journalism from public databases. The purpose is to bring more transparency to the public administration in Brazil.



Figure 7 Screenshot of the Núcleo highlights. Retrieved from: <https://nucleo.ior.br/>, on 05/30/2020 – 15:53.

In Colombia, a team of journalists, data scientists and social scientists have been developing stories, apps and activists' campaigns based on data-driven content. Datasketch is a digital platform which brings Colombians closer to a critical analysis of social realities through contextualized information. It offers free tools to develop simple charts and guides to practice data journalism. A recent work is the counting of social leaders murdered in Colombia since 2016. This used data from the human rights' digital platform Pacifista, the NGO protection program Somos Defensores, the international collaborative initiative Ojo a la Paz, the NGOs Indepaz and CINEP, the United Nations, and from the official body Defensoría del Pueblo y Unidad de Víctimas. The information was analysed using R programming

language and is updated as new data emerge. According to Datasketch, more than 500 social leaders have been murdered since January 2016.

A sort of data journalism practice can be also seen in some African countries. Kenya, for instance, is the first sub-Saharan African country to launch a national open data effort through KODI (Kenya Open Data Initiative), with more than 160 datasets organized under six subheadings: education, energy, health, population, poverty, water and sanitation. The data are sourced from national census, government ministries and the World Bank. However, the Kenyan digital reality faces at least two challenges: increasing the population's access to the Internet (only 25,9% of Kenyans have access to the Internet) and improving transparency (the right to information was granted only in 2010, which requires the government to publish and publicize any relevant information that affects the country). "An open data initiative, though, is a big step for any country – and how the newly accessible datasets will affect the relationship between Kenyans and their government is certainly something to watch", (Rogers, 2013, chapter 7, p.7).

Amidst the conflict between Israel and Palestine, in the Palestinian neighborhood of East Jerusalem, Al Jazeera tracked the houses and buildings destroyed by the Israel's policy of home demolitions in 2016. For the Broken Homes project, the broadcaster worked with the United Office for The Coordination of Humanitarian Affairs for a year monitoring every demolition that occurred in the occupied Palestinian neighborhood. It was destroyed 190 structures and more than 1,200 Palestinians were affected, a record year for demolitions since 2009, when the UNOCHA began to register them. Through an interactive platform it is possible to visualize charts with historical series from 2009 to 2016 and by month from January 2016 to December 2016, maps which highlight the affected areas, videos with testimonies of the families who were displaced and presentation texts with information of what structure was destroyed and who lived there. Al Jazeera reporters spoke with the affected families in the location of the demolitions.



Figure 8 Screenshot of one the sections of the interactive Broken Homes project. Retrieved from: <https://interactive.aljazeera.com/aje/2017/jerusalem-2016-home-demolitions/index.html>, on 05/30/2020 – 15:41.

Data journalism is also practiced by nonprofit initiatives. ProPublica, for instance, stands out by presenting itself as an independent newsroom that produces investigative journalism with moral force. Their mission is to expose abuses of power and betrayals of the public trust by government, business, and other institutions. The platform counts on more than 100 journalists to cover government, business, criminal justice, the environment, education, health care, immigration and technology issues. It was the first web portal to win a Pulitzer Prize, in 2010, for its investigative report *The Deadly Choices at the Memorial*, about the decision of a doctor and nurses to hasten the death of some patients by injecting them with lethal doses of drugs, during Hurricane Katrina at the Memorial Medical Center, in New Orleans. The ProPublica report revealed that mortuary workers carried 45 corpses from Memorial, more than from any similar sized hospital in the city affected by the hurricane.

3.2 Transparency

There are different theories on the limits of transparency. Yet, the one which unveils the factors that influence the creation of news (Allen, 2008) seems to be accurately done data journalism, as the practice is based on analysis of sources much more complex than was created through different methodological processes. In addition, journalists would not only be able to reveal the background, but also expose eventual fragilities related to the information (Allen, 2008).

Although data journalism reset, in given ways, the audience access to information based on more data, it still struggles to find a way to elevate its transparency levels by revealing its methodology. This is an important aspect for data journalism because if the readers are not informed about the process which led to the story, then their right for transparency is not maintained (Appelgren, 2017).

Of the examples mentioned in this chapter, all of them reveal the sources of the data, which is evidence of the understanding that by sharing data sources audience engagement and credibility tend to increase (Felle, 2016). Yet, Zamith (2019) found inconsistencies of transparency in data-driven content of elite newspapers such as The New York Times and The Washington Post after analyzing stories published on both websites for six months. According to the study, only 11.9 % of The New York Times' stories provided methodological information, while The Washington Post gave information of the methodology in 19.6% of the data-driven contents.

On the other hand, an example of transparency, in terms of revealing the methodology used to create the news, can be seen on Núcleo. The stories published in the recently launched website provide information about the methodology and links to the supporting data and other references used.

3.3 Coronavirus in numbers

Besides the concerns about the dissemination of the virus, a pandemic situation also raises another worry related to information about the disease. A side effect aggravated by the uncontrollable digital environment, the misinformation regarding Covid-19 has been felt almost as hard as the physical, and lethal, consequences of the disease. Among the poor quality, even fake, information involving the nature of the virus – from susceptible contamination situations to questionable procedures and drugs to prevent or treat the disease –, and related to the

participation of leaders and States on the dissemination of it, data journalism projects have been bringing light to what is known about the new coronavirus up till now.

It is important, though, to emphasize that while journalists are not in the position of health professionals to provide medical diagnoses, they can help by publishing reliable information, based on the responsible analysis of large databases, about the spread of the virus, the authorities' measures to tackle it and obviously the malpractice cases, both from public administration and civil organizations, that jeopardize the population. How the information is managed is a very important part of any pandemic fighting campaign and the data projects that have been published so far might help authorities and civilians to be prepared for the developments of the coronavirus pandemic and further epidemic situations.

As Bill Gates (2018) anticipated, the world is not prepared for an epidemic situation because it has been investing more in tackling armed threats. "In the case of biological threats, that sense of urgency is lacking. But the world needs to prepare for pandemics in the same serious way it prepares for war", (Gates, 2018, p.2059). A turning point would be, then, using good information to build efficient fighting models against epidemic outbreaks "So, we can have tools, but those tools need to be put into an overall global health system", said Gates at the time of a TED lecture in 2015.

Alongside classical news coverage - from press conferences with health authorities to personal stories about the families affected around the world - the journalistic content based on data analysis have been offering an overview of the spread of coronavirus with charts, part of them interactive, about the public health provision in different realities, the patterns found in the many cases of the disease, the amount of money invested to fight the virus.

This is the case in the article “Coronavirus: Why you should act now”, written by the engineer Tomas Pueyo for the digital platform Medium (Pueyo, 2020). In the piece, Pueyo shows several projections on the spread of coronavirus outside China. But the most important chart of the piece is the one with a timeline of the events in Hubei province. It shows the coronavirus cases with diagnoses versus the non-official cases, which were much more, and the decline of new cases after lockdown measures. For this chart specifically Pueyo used data from the Journal of the American Medical Association, based on raw case data from the Chinese Center for Disease Control and Prevention. To find out the cases without diagnoses, the Chinese CDC professionals asked patients during exams when the symptoms started. “Crucially, these true cases weren’t known at the time. We can only figure them out looking backwards”, said Pueyo in the article.

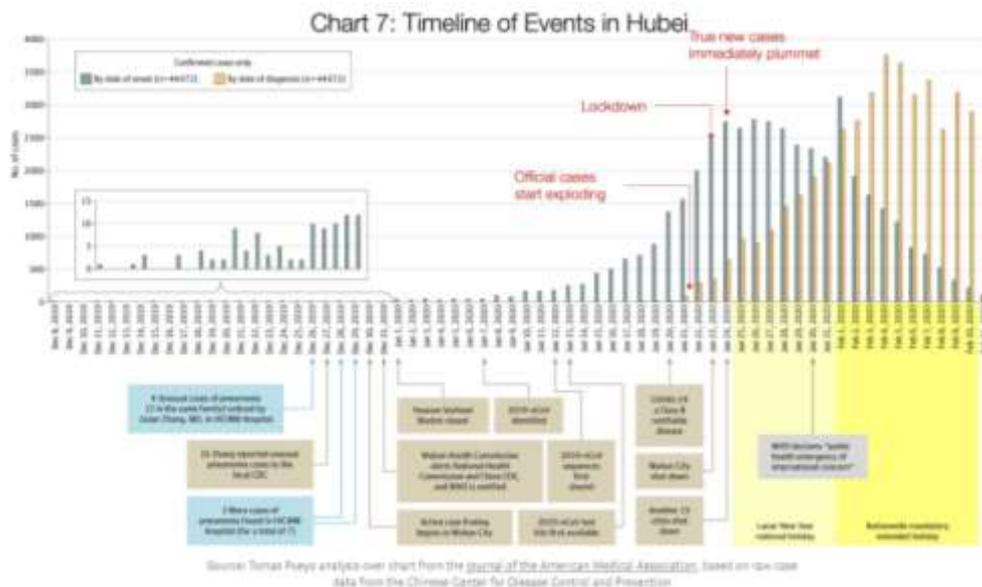


Figure 9 Screenshot of one the analyses made by Tomas Pueyo in “Coronavirus: Why you should act now” story for Medium. Retrieved from: <https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca>, on 05/30/2020 – 15:46.

The Brazilian Núcleo digital platform has launched several data-driven contents about the pandemic situation in Brazil. The pieces cover everything from the effect

of lockdown on the everyday life of Brazilians to the use of social media to monitor the virus spread. They published, for instance, a map showing the limitations of a specific lockdown endorsed by the Brazilian president Jair Messias Bolsonaro in the most needed regions of the country. This specific lockdown recommends the isolation of only certain groups of citizens considered vulnerable, because of their age or health situation. The maps show the residences in several Brazilian cities in which lives more than two residents per room. According to the Núcleo's data analysis, only 2% of the residences in Itaim Bibi, an upper-middle class neighborhood in São Paulo, have more than two residents per room, while in Paraisópolis, the largest Sao Paulo slum, the percentage jumps to 44%.



Figure 10 Screenshot of an extract of the Núcleo's story shows part of the analysis and the links to each source used in the piece. Retrieved from: <https://nucleo.ior.br/urbe/2020-04-02-mapa-densidade-saopaulo-coronavirus>, on 05/30/2020 – 15:53.

This means that in residences with too many residents per room it can be practically impossible to follow lockdown measures if a single resident exhibited symptoms of the coronavirus disease. The team used data from Censo 2010, a demographic research led by IBGE, the Brazilian government agency of geographic and statistical studies. To calculate the percentage of residents in the

regions with more than two residents per room, they multiplied the number of residents from each area by the percentage by individuals living in this condition. Other media organizations in Brazil have also published data-driven content related to the coronavirus pandemic. The Deltafolha, of Folha de S. Paulo newspaper, has been releasing, since mid-March, analysis of messages related to the president Jair Messias Bolsonaro exchanged on social media platforms, and data from official bodies, collaborative digital platforms and international organizations. Recently, Deltafolha published the information that 22 million Brazilians live in critical regions for the pandemic with poor infrastructure and a high level of chronic diseases. They analysed population data from IBGE, (the Brazilian Institute for Geographical and Statistical Studies), mortality and hospital structure data (SIM and CNES) from the Brazilian health department and Covid-19 information from Brasil.IO, a digital platform that facilitates access to Brazilian data of public interest.

The digital newspaper Nexo recently looked at the distribution of hospital beds in the Brazilian Sistema Único de Saúde (SUS) across the country. The team used data from Datasus, the department of SUS in charge of processing information about the public health situation. A set of charts showed the total number of hospital beds in the SUS from 2005 to 2020 compared with the number of non-SUS hospital beds; the number of hospital beds for every 10,000 people in February 2020 compared to October 2005; and the distribution of hospital beds per region and medical specialty. In February 2020, there were 14 SUS hospital beds for every 10,000 Brazilians, five fewer than in 2005. The SUS provision represented 69% of all 426,388 hospital beds distributed across the country.

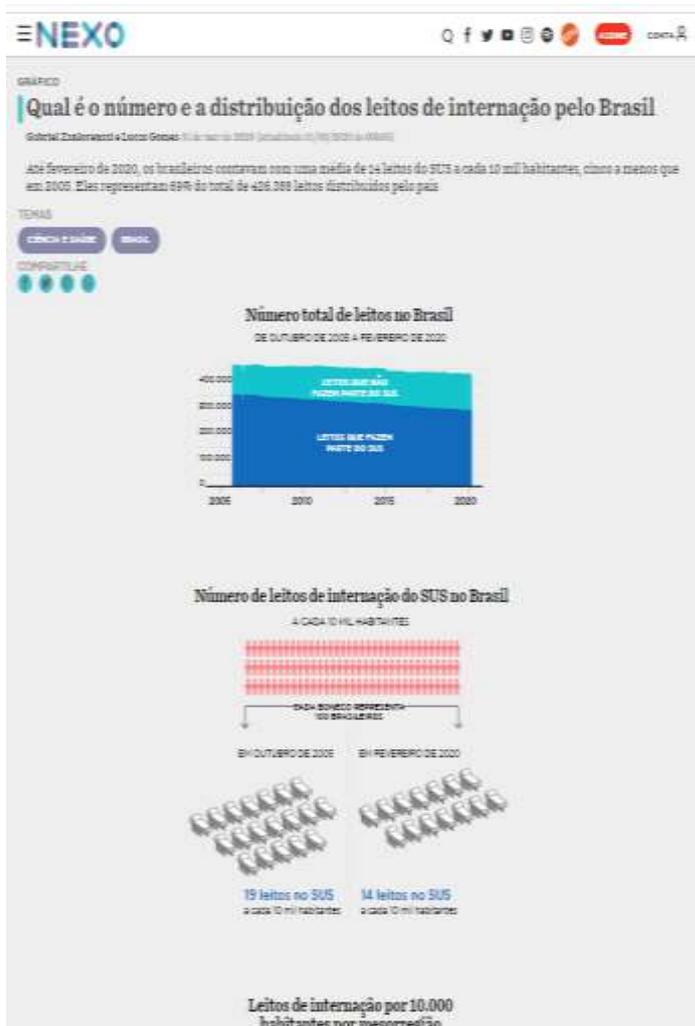


Figure 11 Screenshot of an extract of the Nexo' story about hospital bed provision offer at in Brazil. Retrieved from: <https://nucleo.jor.br/urbe/2020-04-02-mapa-densidade-saopaulo-coronavirus>, on 05/30/30/2020 – 16:09.

On the Infográfico webpage of the newspaper O Estado de S. Paulo, interactive graphics allow readers to see, by region, the spread of coronavirus in the smaller regions and cities of Brazil. The data used to construct the graphics were gathered by the digital platform Brasil.IO, and are based on reports released by local health authorities in the Brazilian provinces. The newspaper gathered the numbers according to the division of Brazilian territory made by the IBGE. In the piece, the reader can type a city or small region into the box of each animated graphic which then shows the increase in total deaths and new daily deaths. Another work shows through a model of the human body the trajectory of the virus responsible for the

Covid-19 disease. The story also contains explanations from a specialist from the Federal University of Rio de Janeiro’s microbiology department.

Naturally, the data journalism team from The Guardian have been covering the spread of the coronavirus and exposed some underlying fragilities. The piece “Coronavirus is revealing how broken America’s economy really is” shows how precarious the Health situation in the United States is, based on updated data from the World Health Organization, UNICEF, the World Bank and the Inter-Parliamentary Union.

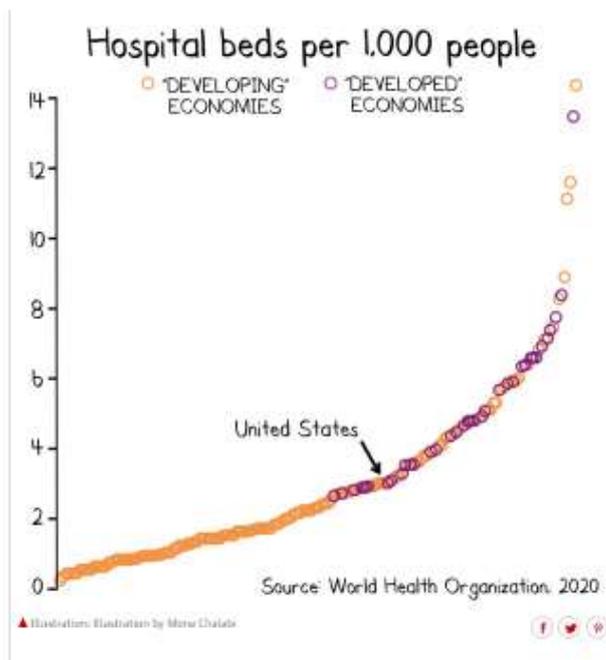


Figure 12 Screenshot of an extract of The Guardian’s piece about hospital beds at in the USA. Retrieved from: <https://www.theguardian.com/news/datablog/2020/apr/06/coronavirus-american-reaction-economy-covid-19>, on 05/30/2020 – 16:07.

The article then questions the “developed economy” status of the US while presenting simple charts with the position of United States in such world rankings as hospital beds per 1,000 people, medical doctors per 1000 people, maternal deaths per 1,000 live births and life expectancy in years. According to WHO, for instance, there are 2.9 beds for every 1,000 people in US, fewer than in

Turkmenistan (7.4 beds per 1,000), Mongolia (7.0), Argentina (5.0) and Libya (3.7). “In fact, the US ranks 69th out of 182 countries analysed by the World Health Organization. This lack of hospital beds is forcing doctors across the country to ration care under Covid-19, pushing up the number of preventable deaths”, according to the report. The United States has 2.6 doctors for every 1,000 people, fewer than Trinidad & Tobago (2.7) and Russia (4), while life expectancy at birth is lower in the US than in Chile or China.

In Portugal, the Público newspaper followed the spread of coronavirus in different countries from the pandemic situation in China and showed an overview of the dissemination of the virus in Portugal. Their sources were mostly the digital publication of data and research from Our World in Data and the non-profit organization Data Science for Social Good Portugal.



Figure 13 Screenshot of an extract of Público’s interactive chart about the spread of coronavirus spread. Retrieved from: <https://www.publico.pt/interactivo/coronavirus-como-achatar-curva-que-revelam-experiencias-paises>, on 30/05/2020 – 16:13.

3.4 Considerations

Data journalism can indeed be a useful tool for supporting decision making in different spheres, from political and economic decisions to those related to health.

As was observed in this chapter, data journalism can vary from interactive charts to articles with a similar structure to existing stories. References like The New York Times and Guardian have been contributing to the public debate with visualizations about the economy, politics, public health, international and local affairs, and human behavior. On the other hand, Folha de S. Paulo's Delta Folha, for instance, although the contents are based on data analysis, tends to look like the other articles published by the newspaper.

Yet, despite the classical references to data journalism, other practices have been flourishing across the world to reveal inequality and local issues. This is the case of the examples from Colombia and Palestine.

However, data journalism does not correct mistakes and solve imbalances by itself, but has to be applied in the exercise of deliberating over given facts, events or declarations. Therefore, data journalism, as in journalism in general, acts as an assistant to understand and measure complex world issues (Gray et. al, 2012).

Additionally, as the practice develops, a new transparency standard in data journalism seems to be needed in order to inform readers not only of the source, but also the process that produced the news story (Allen, 2008; Appelgren, 2017).

4. Methodological trajectory

In this chapter the methodology trajectory adopted to develop qualitative deductive research about how data journalism reconfigures journalism production in the Big Data context will be presented. So, this section explains the research objectives, data collection instruments, data analysis instruments and contributions to a methodology that could be adopted in future work. A brief considerations subsection is also part of this section and summarizes the strengths of the research methodology, as well as the weaknesses.

The purpose of this investigation is to analyse how data journalism reconfigures journalistic production in specific realities and to propose which concepts and journalistic fundamentals cannot be missed while reasoning about proper data literacy. The references used and relationships established, will also support the Erasmus+ Post-Crisis Journalism in Post-Crisis Libya: A Bottom-up Approach to the Development of a Cross-Media Journalism Master Program, a higher education proposal co-funded by the European Commission. Therefore, the main objectives would be:

- to propose a reflection on how the omnipresence of Big Data changes epistemic patterns of journalism as a knowledge mediator;
- to portray the main data journalism initiatives;
- to observe how data journalism might serve as tool to recover interest and credibility among readers;
- to propose a debate on appropriate level of data literacy required to practice data journalism, based on exploratory interviews with professionals and educators.

Initially, this study also had as an objective 'to analyse different training initiatives, master's degrees and courses in Europe, United States and Brazil'. However, it was considered that this analysis would be an appropriate objective for future

research exclusively dedicated to current educational practices in the data journalism field, given the range and diversified nature of each learning context.

In order to elucidate a real phenomenon, a researcher must pursue a methodology (Quivy & Campenhoudt, 2005). The research methodology is qualitative deductive, which will provide an integrated overview of the subjects through participants' perceptions and data that allows multiple interpretations (Gray, 2004).

Therefore, this investigation project proposes a multiple case study over the following objects:

Data journalism production routine – Semi-structured interviews with; Sergio Spagnuolo, founder of the Brazilian data journalism agency Volt Data Lab; Gabriel Zanlorenssi, data scientist at Brazilian newspaper Nexo; Fábio Takahashi, Data Journalism Editor at Folha de S. Paulo newspaper; and Rui Barros, a data journalist from Público newspaper.

Literacy required for data journalism practice – Semi-structured and structured interviews with professors from the Universidade Católica de Brasília, one of the few universities in Brazil which offers a post-graduate course exclusively in data journalism, and from the Federal University of Bahia, which is leading an effort to update the journalism course syllabus with disciplines related to data and digital platforms.

Attempts to contact the Columbia Journalism School and University of Missouri, which hold reference programs in data journalism, were made several times during the first semester of 2020. Unfortunately, by the conclusion of the first stage of this research, the subjects had not responded to the interview requests. Additionally, the University of Minho was also contacted for an interview about some classes dedicated to infographic design, but they also did not respond to the requests.

This research was divided into several different phases supported by Quivy & Campenhoudt's (2005) procedures and Gray's (2004) references.

During the exploratory phase (Quivy & Campenhoudt, 2005), a literature review provided three main concepts that would help to understand the object of this dissertation. They are Big Data, data journalism and data literacy.

With regards to the hypothesis, this investigation relies on the argument that, amidst the digital culture and information structured through Big Data, the journalist has to learn new skills which are not natural in the field of journalism to deal with non-human sources of information since the data structures are not yet as well-known as human sources are.

In order to verify the hypothesis, the data were collected through semi-structured interviews with data journalism professionals and professors related to the topic. This approach to data collection was applied since it provides opportunities for probing in case of more complex issues about the subject (Gray, 2004). Moreover, the interview is an effective instrument for different fields in social sciences, including journalism studies (Marconi & Lakatos, 2003).

In the case of semi-structured interviews, a list of topics to be covered allowed the interviewee to feel more comfortable while explaining about aspects of data journalism.

An analysis model (Quivy & Campenhoudt, 2005) was elaborated to guide the data collected through interviews. During the exploratory phase, some indicators were added and others deducted as new theoretical references were added.

Table 1 Analysis Model

Concept	Measure/Dimensions	Indicators/Categories
Data Journalist	Sociodemographic characteristics	Age Educational background Career trajectory Motivation to work in data journalism
Big Data	Volume	Correlate information from several different sources Data as central to knowledge acquisition (Godler, Reich, & Miller, 2020) Tools to analyse data Tools to clean data Big Data as technological, scholarly and cultural phenomena(Boyd & Crawford, 2012) Collaboration to handle Big Data (Splendore et al., 2016) Limitations to penetrate dispositive (Flusser, 1983)
Production	Content creation process	Data as primary source (Hammond, 2017) Reshaped format (Gray et al., 2012) Reconfiguration of purposes (Gray et al., 2012) Emphasis on dramatic (Cushion et al., 2017) Shift in the production awareness Collaboration with other professionals Tools/resources to create visualizations Complementary sources Standards and style to deal with non-human sources
Data journalism	Data literacy	Multiple skilled nature (Stray, 2016) Scientific methodology, statistic and computing skills Critical observation (Gray et al., 2018; Howard, 2014) Development of new techniques Data under human consciousness (Godler et al., 2020)

	<p>Data infrastructure literacy (Gray et al, 2018)</p> <p>Access to literacy</p> <p>Advice from other professionals</p> <p>Academically qualified in the data field</p> <p>Autonomy to deal with data</p> <p>Faculty qualification</p>
Credibility and interest	<p>Audience profile</p> <p>Repercussion among readers</p> <p>Repercussion among authorities and organizations</p> <p>Repercussion among other media organizations</p> <p>Authority of data</p> <p>False authority of data (Appelgren, 2017)</p> <p>Control over audience access</p> <p>Insights into complex issues (Gray et al.,2012)</p> <p>Editorial autonomy (Cushion et al., 2017)</p> <p>Shared methodology/process (Appelgren, 2017).</p> <p>Low quality of data (Berti-Équille & Borge-Holthoef, 2016)</p>

To obtain more detailed deductions from the content of interviews, the qualitative analysis software WebQDA provided by the University of Aveiro, was used. By using the software, the purpose was to assemble analysis categories and acquire the results through a systematic description of the interview content which allows inferences about the subject (Bardin, 1977).

An initial reading of the contents was useful to adjust the indicators/categories that would organize the sources (Bardin, 1977). A search for most frequently used words (Bardin,1977) was deployed during the stage of initial reading. This technique also helps to visualize eventual indicators that may/ might emerge from the statements that could be neglected.

After the characterization and description of the contents, the inquiring and analysis of the data was started. In this phase, it was possible to correlate indicators with the support of WebQDA software and create visualizations of the matrices. This procedure was effective to establish the connections between the characteristics and indicators and interpret them (Bardin, 1977).

To keep the internal coherence of the research the organization method presented by F.N Souza, D.N. Souza and Costa (2016) was adopted, which consists of assembling the research questions, the objectives and the analysis scheme followed. Therefore, it was possible to predict the subsequent questions that could emerge during the data analysis aligned to the research question and objectives.

Table 2 Research consistency method

Research consistency				
Research question	Research objectives	Data	Kind of analysis	Obs. Expect.
How has data journalism reset the production of news in a Big Data context?	1) Propose a reflection on the relation between data journalism and big data with data as a knowledge mediator; 2) Portray data journalism initiatives; 3) Observe how data journalism might serve to recovery credibility; 4) Propose reflection on appropriate levels of data literacy	Interviews with professionals	Content analysis through categories and subcategories (Bardin,1977). The analysis is based on the model presented by Quivy & Campenhoudt (2005)	Ask the data questions based on objectives and research questions

4.1 Interviews

To collect the data for this investigation, interviews were carried out with professionals whose activities are related to the data field: journalists, data scientists and professors. In order to prepare the interviews, the directives on readiness and awareness of the object field presented by Marconi and Lakatos (2003) were taken into account. With regards to the planning of the interviews, a script was produced based on the analysis model in order to cover the main concepts, dimensions and indicators.

Therefore, the scripts used the four concepts addressed in the analysis model: data journalist (data professional for the ones who are not journalists) for the sociodemographic characteristics dimension, Big Data for the volume dimension, production for the content production process dimension and data journalism for the dimensions of data literacy and credibility/interest.

The subjects were selected according to the relevance of their work with data. Gabriel Zanlorenssi, for instance, is data scientist at Nexo newspaper, a relevant, though recently founded, media in the field of visual journalism which provides journalistic content that explains the context of recent events.

With regards to the journalists and professors selected for the interviews, they play relevant roles in their professional and educational environments. Sergio Spagnuolo is an experienced journalist in the fields of economics and politics, as well as in investigative journalism. He is a member of the Brazilian Association of Investigative Journalism and the founder of the Brazilian data journalism agency Volt Data Lab. Fábio Takahashi is also an experienced journalist from Folha de S. Paulo, one of the most traditional Brazilian newspapers. Takahashi is the leader of the Delta Folha, a division of the newspaper in charge of producing data-driven contents.

In the educational area, Dr. Alberto Marques is relevant to this investigation since he is one of the founders of the Data Journalism post-graduate and Digital Journalism course at the Catholic University of Brasilia. Marques is also in charge of the reconfiguration of the Journalism bachelor's degree at the Catholic University of Brasilia- the course started with its first class in the first semester of 2020.

From the Federal University of Bahia, Dr. Suzana Barbosa is the new director of the journalism course at the university and is leading a reconfiguration in the course syllabus to aggregate data and digital skills. Barbosa is a member, and one of the leaders, of the Grupo de Pesquisa em Jornalismo Online (GJOL) at the university. In 2008, Barbosa was awarded the Prêmio Adelmo Genro Filho, from the Associação Brasileira dos Pesquisadores em Jornalismo (SBPJor), for her thesis "Jornalismo Digital em Base de Dados (JDBD) - Um paradigma para produtos jornalísticos digitais dinâmicos".

Although the data collection instruments are based on the analysis model above, it was necessary to adapt the subjects for some of the interviews and field observations. Such was the case in the interview with Sergio Spagnuolo, Volt Data Lab founder. Since Volt Data Lab delivers data-driven content to its clients, such as media companies, some indicators in the credibility measure defined in the analysis model were modified.

Indicators such as "usage of content by official bodies, institutions" and "usage of content by other media organizations" lost relevance. Therefore, other indicators verified in the credibility measure of data journalism were used: "at which level is the news credibility associated with statistical data" and "how the agency assesses the public's understanding of data".

Other modifications were made as the data scientist Gabriel Zanlorenssi is not a journalist – his background is in social and political science –, but works as data

scientist at a Newspaper. The indicators related to journalistic practice were most altered. It was the case with the concept of Data Journalist, which had to be adapted to Data Professional, and the indicator on “motivation to work with data journalism” had to be altered to “motivation to work with data journalists”.

Similarly this occurred with the indicator “which subject the journalist covers”, which was modified to “the journalistic areas which require more data science effort”. In the production section/concept, the indicator “the need for PR between the journalist and the information” was removed.

Additionally, the indicators of the Data Journalism concept were adapted to consider Zanlorenssi’s perception on the autonomy of his journalist colleagues. Therefore, the indicators on data literacy dimension were “the perception of the autonomy of the journalist to search, clean and interpret the data”, “the perception of the autonomy of the journalist in creating the visualizations of the data”. Also, with respect to credibility and interest dimension, only one indicator was adapted in the case of this interview, which was “how the professional assesses the relation between data and news credibility”.

4.2 Field observation: restrictions and future work

A field observation in the areas of education and the production routine had been planned for this research, to get beyond the individual’s opinion and self-interpretations (Gray, 2004). However, with coronavirus spreading to a pandemic situation across the world, this procedure had to be canceled. Nevertheless the observation can still be deployed in an advanced stage of this research in the future.

The field observation would have occurred in two different contexts: a newsroom and a class at university. The purpose would be to gather information about the

production process of a data-driven story and the dynamics of a class. The observation of the mentioned realities would be guided by a systematic procedure involving recording and then analyzing and interpreting the subject's actions.

The observation procedure chosen would be the systematic one (Marconi & Lakatos, 2003), in which the observer follows already determined conditions, although the parameters would not be too strict in order to seize all the input. Additionally, it could be deployed in a non-participant manner (Marconi & Lakatos, 2003) since the researcher does not necessarily need to interact with the observed reality in order to extract the data.

4.3 Considerations

Naturally, a limitation to this research was the impossibility of moving forward with the field observation at the Portuguese newspaper Público's newsroom and the University of Minho's Communication Science master's degree. The University did not respond to requests, however it published on its website an official notice announcing the suspension of academic activity across campus (appendix 10). On the other hand, David Pontes, one of Público directors, had agreed to the visit (appendix 11), but with the coronavirus pandemic the professionals adopted working from home as a protection measure. The visit was, then, cancelled by the data journalist Rui Barros during his interview.

The semi-structured interviews, however, earned results able to support the research with qualified information from professionals who are highly regarded in their fields. During the data collection process, the methodological approach of semi-structured interviews (Gray, 2004; Marconi & Lakatos, 2003) proved to be an effective instrument as the subjects are experienced in their fields. The ability to probe into more complex issues (Gray, 2004) was possible because the interviews were not tied to a strict sequence, but to a script with topics which gave the interviewee the possibility of expanding on their impressions.

Naturally, this kind of interview resulted in a large volume of data to be analysed. There were six interviews of approximately 50 minutes each, except in two situations where the interview lasted for more than an hour.

To analyse this volume of data properly, the software WebQDA was essential to organize data into categories and sub-categories (Bardin, 1977). It is worth stressing, though, that the categorization of the data was made based on the concepts, dimensions and indicators referred to in the analysis model suggested by Quivy and Campenhoudt, (2005).

Additionally, the research coherence scheme suggested by N Souza, D.N. Souza and Costa (2016), which consists of assembling the research question, the objectives and the analysis scheme followed, was effective in keeping the data analysis in line with the research objectives.

5. Results

This chapter presents the results of the research organized by the questions asked using the data gathered with the support of the WebQDA software. Therefore, this division starts with a superficial analysis of the general interpretations of the interviewees and the most frequently cited words and terms, through tag cloud visualization. Then, the questions asked about the data are organized through different subdivisions for discussion of the results.

This research proposes an analysis of specific aspects of data journalism. They are the Big Data context in which the practice is placed, an outline of the data journalist profile limited to the scope of the professionals interviewed for this study, and the literacy obtained in different circumstances to deal with data.

Naturally, these are references to a deductive approach towards data journalism practice and literacy retrieved from contexts related to practices in Brazil and Portugal. As more professionals are consulted, more educational and professional characteristics might be added to the general view.

Dr. Alberto Marques (appendix 1) led the organization of the Data Journalism *latu sensu* postgraduate programme at the Catholic University of Brasília, one of the few postgraduate degree courses offered in Brazil exclusively about data journalism. Marques is also in the charge of the reconfiguration of the journalism bachelor's degree at the Catholic University of Brasília. The course, now reformulated to Digital Journalism, started its first class in the first semester of 2020.

Fábio Takahashi (appendix 4) is an experienced journalist from Folha de S. Paulo, one of the most traditional Brazilian newspapers, and Lead Editor at Delta Folha, a department of the newspaper in charge of producing data-driven content.

Gabriel Zanlorenssi (appendix 2) is the data scientist of Nexo newspaper, a relevant digital media in the field of visual journalism which provides journalistic content that explains the context of recent events.

Rui Barros (appendix 3) is data journalist at Público newspaper. His experience includes around three years working as data journalist for Renascença Broadcast Media, one of the oldest radio stations in Portugal.

Sergio Spagnuolo (appendix 5) is an experienced journalist in the fields of economics and politics, as well as in investigative journalism. Spagnuolo is a member of the Brazilian Association of Investigative Journalism and founder of the Brazilian data journalism agency Volt Data Lab. Spagnuolo has also been a leader of the Núcleo website, a portal with content exclusively driven by data launched in January this year.

Dr. Suzana Barbosa (appendix 6) is the director of the journalism course at the Federal University of Bahia and one of the leaders of the online journalism research group known as GJOL, also at the University. Like Marques, Barbosa has been leading a reconfiguration in the bachelor's degree in Journalism in order to aggregate digital skills to the course syllabus. In 2008, Barbosa was presented with an award by the Brazilian Journalism Research association for her thesis Digital Journalism in Databases (DJDB).

The categories and subcategories (Bardin, 1977) used in the qualitative analysis software WebQDA were based on the concepts, dimensions and indicators of the Quivy and Campenhoudt (2005) method.

Table 3 Analysis model/categories

Concept	Measure/Dimensions	Indicators/Categories
Data Journalist	Sociodemographic characteristics	Age Educational background Career trajectory Motivation to work in data journalism
Big Data	Volume	Correlate information from several different sources Data as central to knowledge acquisition (Godler, Reich, & Miller, 2020) Tools to analyse data Tools to clean data Big Data as a technological, scholarly and cultural phenomenon (Boyd & Crawford, 2012) Collaboration in handling Big Data (Splendore et al., 2016) Limitations to dominating the machine (Flusser, 1983) Limitations to accessing data
Production	Content creation process	Data as primary source (Hammond, 2017) Reshaped format (Gray et al., 2012) Reconfiguration of purposes (Gray et al., 2012) Emphasis on the dramatic (Cushion et al., 2017) Shift in production awareness Collaboration with other professionals Tools/resources to create visualizations Complementary sources Standards and style to deal with non-human sources

Data journalism	Data literacy	<p>Multiple skilled nature (Stray, 2016)</p> <p>Scientific methodology, statistics and computing skills</p> <p>Critical observation (Gray et al., 2018; Howard, 2014)</p> <p>Development of new techniques</p> <p>Data under human consciousness (Godler et al., 2020)</p> <p>Data infrastructure literacy (Gray et al., 2018)</p> <p>Access to literacy</p> <p>Advice from other professionals</p> <p>Academically qualified in the data field</p> <p>Autonomy to deal with data</p> <p>Faculty qualification</p>
	Credibility and interest	<p>Audience profile</p> <p>Repercussions among readers</p> <p>Repercussions among authorities and organizations</p> <p>Repercussions among other media organizations</p> <p>Authority of data</p> <p>False authority of data (Appelgren, 2017)</p> <p>Control over audience access</p> <p>Insights into complex issues (Gray et al., 2012)</p> <p>Editorial autonomy (Cushion et al., 2017)</p> <p>Shared methodology/process (Appelgren, 2017).</p> <p>Low quality of data (Berti-Équille & Borge-Holthoefler, 2016)</p>

The table below lists all the indicators/categories, named in the WebQDA software as codes, with the number of references each source made to each one of the topics referred to.

Table 4 WebQDA codes

Nome	Refs	Fontes
Big Data - Volume	6	2
Correlate information from several different sources	2	1
Data as central to knowledge acquisition	9	4
Tools to analyse data	12	6
Tools to clean data	0	0
Technological, scholarly and cultural phenomenon	1	1
Collaboration to handle data	4	2
Limitations to dominate technology	4	2
Limitations to access data	2	1
Production	11	3
Data as primary source	15	6
Reshaped format	17	5
Reconfiguration of purposes	28	6
Emphasis on dramatic	5	3
Shift in the production awareness	27	5
Collaboration with other professionals	25	6
Tools/resources to create visualizations	9	4
Complementary sources	13	6
Standards and style to deal with non-human sources	28	6
Data literacy	21	4
Multiple skilled nature	28	9
Scientific methodology, statistic and computing skills	29	9
Critical observation	27	6
Development of new technics	24	7
Data under human consciousness	2	1
Data infrastructure literacy	3	1
Access to literacy	29	6
Advise from other professionals	5	1
Academically qualified in the data field	20	5
Autonomy to deal with data	5	2
Faculty qualification	13	3
Credibility and interest	7	3

Audience profile	8	4
Repercussion among readers	11	4
Repercussion among authorities and organizations	6	3
Repercussion among other media organizations	1	1
Authority of data	10	5
Low quality of data	3	1
False authority of data	3	3
Control over audience access	4	4
Insights into complex issues	8	3
Editorial autonomy	4	4
Shared methodology/process	11	3

The content analysis covered nine sources, from the professionals' interviews to syllabus courses, and resulted in 500 references categorized through 43 indicators/categories. Within this chapter some excerpts from the interviews will reinforce the statements made and will also be organized in tables according to some of the indicators/categories established to analyse the data. Since all the interviews were conducted in Portuguese, each excerpt was translated into English, but the original files are included in the appendix.

The figures below present the 50 most frequently cited words by each interviewee regarding the practice of data journalism. In this case it is possible to start with an analysis, although still superficial, about the first impressions of the interviewee on data journalism and the elements that, in their view, most influence the practice.

Alberto Marques



Portuguese words for journalism, data, information, course, professor, think, research, qualification, also, different, agenda, difficulty, times, always, something, do, several, professional, know, newsroom, market, catholic, post-graduate, university, taking, production, people, format, bachelor, trajectory, Brasilia, digital, now, process.

Possibly influenced by the nature of his work as a researcher and educator in the field of digital journalism, Dr. Marques focused his speech mostly on data journalism skills and the reconfiguration of production realities due to global digitalization, and how journalists are reconsidering ways to deliver news to their audience. These ideas are mostly represented by the terms with equivalence to English words such as “data journalism”, “people”, “format”, “course”, “thinking”, “training”, “production”, “digital”, “doing”, “information”, “professor”, and “professional”.

Fábio Takahashi



Portuguese words for data, group, Folha, journalism, do, education, little, something, politics, example, people, professionals, general, question, public, reporter, months, times, other, coding, team, editor, different, journalistic, great, project, music, sense, reportage, newspaper, analysis.

Fábio Takahashi is the lead editor of Delta Folha, the data journalism division of the traditional Folha de S. Paulo newspaper. His partnership of years with one of the main Brazilian newspapers is clear in his speech. The name “Folha”, as the newspaper is commonly called, was one of the most cited words during the interview, alongside data, group, journalism, doing, reporter, question, professionals and coding. The frequent use of terms like “professionals” and “group” seems to reflect his work with a multidisciplinary team in charge of the data-driven content at Folha de S. Paulo. Meanwhile, the word “question” appears many times as the professional insists on the need for questioning the data thoroughly, as any journalist usually questions authorities, specialists, organizations and individuals.

Rui Barros



Data, do, journalist, journalism, work, tell, people, statistic, also, public, coding, notion, learn, question, science, university, degree, great, information, formats, master's degree, interest, jobs, reader, process, important, mathematics, course, analysis, accomplish, understand.

Rui Barros made several references to “data”, “work”, “example”, “journalist”, “journalism”, “people”, “statistics” and “coding” in his interview. There is a bigger emphasis on “data” and the verb “do”. One of the explanations might be the fact that Barros has been developing the literacy to deal with datasets in journalism mostly through non-academic ways. As will be further detailed, a great part of the knowledge he attained came from reading, experimentation and advice from other professionals.

Sergio Spagnuolo



Portuguese words for data, journalism, do, example, story, analysis, journalist, difficult, speak, source, Brasil, world, conclusion, people, Núcleo, transparency, government, work, consulting, content, technique, also, someone, need, come, start, win, accomplish.

Founder of the data journalism agency Volt Data Lab, in charge of the data journalism portal Núcleo and member of the of Brazilian Association of Investigative Journalism, Sergio Spagnuolo demonstrated in the interview a closer relationship to government issues, reportage patterns and, evidently, data journalism. His most frequently cited words were Portuguese equivalents for “Brasil”, “Núcleo”, “story”, “transparency”, “doing”, “journalist”, “world”, “conclusion”.

Suzana Barbosa, FACOM/UFBA director



Portuguese words for journalism, data, qualification, professor, digital, reality, do, need, true, course, techniques, communication, way, technology, know, work, university, formats, research, journalist, Brasil, contents, degree, important, sources, information, project, moment, aspect, professional, strong, team.

As an educator and researcher in the digital journalism field, like Dr. Marques, Dr. Barbosa’s speech was mostly permeated by terms such as “journalism”, “data”, “training”, “communication”, “courses”, “techniques”, “to do”, and “digital”. As will be detailed further in this chapter, the frequency of these words shows the consciousness of the interviewee about the deployment of specific skills to properly develop data journalism characteristics.

5.1 Confrontations between theoretical concepts and evidence

The intention of this study is to provide deductions enlightened by three concepts that relate to data journalism practice. These concepts were also used to organize the literature review and, therefore, the data collection and analysis. The concepts are Big Data, for the volume dimension; journalistic production, from a content creation perspective; and data journalism, through the lens of data literacy and credibility.

With respect to the Big Data concept, the interview subjects seem to agree that it is increasingly difficult for a journalist to dominate Big Data, yet this is a reality which the journalist cannot avoid in the production routine, as argued by Splendore et. al (2016).

Regarding the production concept, the data indicate an awareness among those interviewed with respect to the shift in information now produced and distributed, as described by Hamond (2017) and Howard (2014). Yet, in the context of this research, there seems to be an understanding that journalism still moves with some uncertainty to a data-driven reality, despite the acknowledged advances in the field over the last decade. Sergio Spagnuolo states the majority of news coverage is still based on statements:

Journalists still have their minds mostly focused on a declaratory form of journalism, that journalism based on non-quantitative sources, right? Now it is starting to acknowledge the value that data journalism can bring to news coverage in general, right? Yes, there is a lot to improve yet (Spagnuolo,2020).

Meanwhile Rui Barros, affirms there was not a data journalism culture within journalism in Portugal until recently.

So, I mean, there was no data journalism culture in Portugal in 2016. There was no, I mean, there was this notion of what the New York Times, The Guardian, El País, I mean the great newspapers, were doing at the time, but there was no culture and, therefore, it was necessary to develop here a line of work as well (Barros, 2020).

Dr. Barbosa also makes reference to an erratic development of the practice within newsrooms, a situation that seems to be enhanced by several world crises that also affected journalism practice. In Dr. Barbosa's view, difficult times led to the decommissioning of data groups inside the newsroom and might have made the data-driven coverage more dispersed.

We had a moment where many newsrooms had their data teams, specific groups working content driven by data. Nowadays, in many cases there is not a desk shaped this way, but the journalists remain in the newsroom, some of them learning by themselves. So, we are in a very hard period where the newsrooms around the world are smaller and these data teams, in most of cases, already passed this phase because journalism was confronted by a new cycle of crisis now aggravated by the pandemic situation. What I am saying is that many teams do not exist anymore as specific desks. But whoever has learned it disseminates it through the newsroom (Barbosa, 2020).

Naturally, the journalistic work of verifying sources will not be solely a data-driven procedure. Dr. Barbosa acknowledges that in the future not every journalist would be able to work guided by datasets. "Also, I do not believe that this is something that everybody in given newsroom will master. It won't be like that", Barbosa declared / said / stated.

Another important aspect identified in the literature review is the possible emphasis on the dramatic over the information element to support a worthy

narrative (Cushion et al., 2017). With regards to this side, Spagnuolo, from Data Volt Lab and Núcleo, Zanlorenssi, from Nexo newspaper, and Barros, from Público newspaper, seem to agree that some sort of balance is needed to tell relevant stories. Spagnuolo states that in certain circumstances, a journalistic content that purely informs without necessarily reaching any conclusion might also serve the informative mission of journalism:

“I think it is very interesting when journalism make conclusions about something and I think the main purpose has to be to reach some conclusion and have something to say. But you can also present information in such a way that not everyone who can [...] So, we go there and do this for others and people will get informed as well from that” (Spagnuolo,2020).

Zanlorenssi defends the need to weight on journalistic relevance from the beginning. “Another aspect, which I believe also concerns the journalistic agenda, is to understand the journalistic relevance of something. In some situations, some things require research, so we cannot be negligent, right?”, (Zanlorenssi, 2020).

Rui Barros, from Público newspaper, adds that balance is indispensable in producing data journalism content. In his opinion, the data journalist must consider the appeal to the general public as well as the numbers. “There is sometimes this situation in which the data journalist occasionally focuses too much on numbers and might forget to elaborate the human side of the work”, (Barros, 2020).

From Takahashi’s perspective, however, data journalism has this inseparable technical feature that interferes in the tone of the content. “In our case, things are less, let us say, human in the form of telling the story, a bit more technical”, (Takahashi, 2020).

With regards to data literacy, perhaps, academics seem to be a better choice to collaborate with on data journalism projects. Clearly, they have already mastered

the methodology and statistics skills (Howard, 2014), but they are also used to elaborate research questions that might be useful while making sense from datasets to deliver news.

They are professionals, use coding languages and they are able to visualize the problem, see? They are able to answer a question. In general, this kind of professional is in the academic sphere and I think that is a good convergence, (Takahashi, 2020).

5.2 Analysis of the results

Fifty-six pages of data were analysed, the interviews and documentation related to the syllabus of the Universities was consulted for this stage of the study. To obtain more accurate results, a range of questions were made to the dataset through the creation of grids with the assistance of WebQDA software. The first question is related to the influence of educational background on the interviewees and how they pursue for specific skills to deal with data.

In the table below, some excerpts of the interviews show the different paths each interviewee pursued in their respective work in the data journalism field:

Table 5 What is the relationship between educational background and data literacy?

Alberto Marques	My career is one foot in the academic sphere and the other one in the market. I have a master’s degree from the Federal University of Bahia with [Marcos] Palacios. It was the way I dreamed, but there is an important detail: I did not research journalism, but cyberculture. Now in the PhD program, which was about convergent journalism, I have proposed a study about tablets.
Fabio Takahashi	I have a major in journalism from the Universidade Metodista de São Paulo. In 2016 I went to Columbia University to develop research through a fellowship program in the field of journalism with emphasis on

	education. I stayed there from 2016 to 2017 [...] there I had just one class about data analysis that helped a bit, mostly with theory.
Gabriel Zanlorenssi	I have a major in Social Science and a master's in Political Science from the University of São Paulo. My experience is mostly in the area of social sciences with application to quantitative aspects because of Political Science. Even coming from human sciences, turn out to have a pretty quantitative nature.
Rui Barros	I majored in Communication and Media Studies. In my first year I had a course on quantitative methods, pure statistics [...] indeed I appreciate statistics because it finally makes sense for me. Back at the time I was not even familiar with the data journalism concept, I wanted to be a journalist.
Sergio Spagnuolo	I have a major in Journalism from Mackenzie and a master's degree in Social Science from the Pontifical Catholic University of São Paulo. I like the profession; I am a journalism enthusiast.
Suzana Barbosa	I have a master's degree and a PhD from the Federal University of Bahia's program in communication and contemporary cultures, held by the communication school. I did my internship for my PhD at University Beira Interior, in Portugal, and a post-doctorate at the University of Santiago de Compostela. The study in journalism driven by data has come since the PhD.

Although the interviewees have an educational background in social science areas, it was sometimes the case that part of their data literacy was achieved through experimentation in a non-academic context. It is the case for Barros, from Público newspaper, who has a major in journalism from the University of Minho. He studies/ studied in the master's degree program in the same University, but developed most of his data skills reading references in the field and, most of all, by trying and failing.

Honestly, most the skills I have been developing are not through formal ways, it is more a kind of informal education, I have been reading a lot

about statistics and doing something that is not probably the best learning process, but it has been working well, which is to learn something as I need to develop a new skill (Barros, 2020).

Although both Sergio Spagnuolo and Fabio Takahashi have educational backgrounds in journalism, they searched for subsidiary knowledge outside journalism to relate to their profession.

Gabriel Zanlorenssi is the only interviewee that has no major in journalism. His educational trajectory includes a major in social sciences and a master's degree in political sciences from the University of Sao Paulo (USP). Possibly because of the nature of his educational background, most of his skills related to the analysis of data were initiated, and still are occurring, in the academic context:

I think we have several courses in bachelor's degree and postgraduate which are specifically on learning how to deal with data, references, statistics and coding. For example, I have learned R language during the postgraduate course because I have a course exclusively about it. Now I will start a new class, though not as a regular student, about Bayesian statistics (Zanlorenssi, 2020).

These distinct educational realities reveal that there are educational areas that better covers the new needs especially with regards to data-driven journalism. So, journalists have been pursuing a specific literacy outside the journalistic environment to better perform in the multiple skilled data journalism (Stray, 2016).

Despite this, Barbosa and Marques' perspectives on access to literacy seems to corroborate the accommodation with the lack of data skills perceived by Hewett (2016).

Table 6 What is the relation between educational background and data literacy?

Suzana Barbosa	See, from what we can check on reports published in several countries, obviously from distinct realities, it still fails in the sense that we still are failing to master the techniques
Alberto Marques	Journalism in general has a lot to gain if professionals qualify with specific skills because this is an existing gap

5.2.1 Perception of Big Data implications

The consciousness that a journalist cannot handle the mass and complexity of Big Data alone, as argued by Splendore et al. (2016), seems to be present in the work of the professionals interviewed. It can be seen that there is a search for the proper tools for data journalism work rather than a pursuit of an understanding of the implications of Big Data in the journalistic routine.

As Cukier & Mayer-Schönberger (2013) summarize, Big Data can be understood as an ability of both groups and individuals to make use of information to produce insights and products that are useful and meaningful. Although the professionals interviewed, especially those in the production fronts, express a close relationship with the forms of producing in data context, they do not demonstrate a similar relationship with this idea that Big Data is created from chunks of information.

The assumption of Rui Barros, from Público, that little data journalism work is produced from Big Data might serve as an indicator that Big Data might be seen more from the perspective of the person accessing it.

Right away there is here something which is: little data journalism is still done with Big Data. Why is that? Because when we talk about Big Data it means that it really is a lot of data. It is not even the complexity factor, but the fact that is necessarily a structure, it is necessarily a basic thing that is the raw material which are the data. Often there is none, or every time there is none (Barros, 2020).

Similarly, Dr. Marques seems to associate Big Data with the access the third parties usually offer, or not, to their information sets.

What is new in all this set of technologies is the availability of information at an unprecedented volume. In some cases, the access is not even given, they block all ways. There is a kind of confusion in terms of legislation towards data transparency and those in charge have no interest in this [access] happening (Marques, 2020).

As an attempt to exhaust the datasets through devices, a movement similar to the black box dynamics of Flusser (1983) in which the human try to penetrate the post-industrial devices, the subjects seem to resort to the same coding and statistics programs in their work routines.

Table 7 What is the journalist's perception of Big Data implications in their work?

Tools to analyse data	
Alberto Marques	The idea is that the student goes deep inside Python coding language, or APIs, structures regarding data frames and so on. So, this is about basic visualization skills, public data mapping, filtering algorithms, the basics.
Fabio Takahashi	Everyone has their own methods, tools. Some professionals use more R language, others use Python. Actually, they are pretty much autonomous in this case.

Gabriel Zanlorenssi	We usually work basically with R language. I would say that more than 90% of the work is made on R. Obviously, once in while we use spreadsheets to collect something, but we do not generate graphics from them.
Rui Barros	I mainly use R coding language to analyse the data. Initially I used Excel to make those pretty basic analyses and quickly I noticed that there were too many things that could not be done in Excel and Excel was too constrained and, for that reason, writing the script is the best thing in the world because it allows procedures to be automated [...] If some update in the data happens, in Excel I have to do all over again, but with R language I already have things programmed and it is just going back and clicking and everything runs again
Suzana Barbosa	The use of the Internet is for sure more expanded. People usually know basic searches, but many times they do not master the advanced techniques.

5.2.2 Content characteristics

A question about the nature of the data-driven content produced in the context of those interviewed was also made to the data collected. To obtain more accurate answers, the professionals from an educational context were excluded from this question deliberately. Therefore, this sample is supported by the answers of Fabio Takahashi, data journalism editor at Folha de S. Paulo, Gabriel Zanlorenssi, data scientist at Nexo newspaper, Rui Barros, data journalist at Público newspaper, and Sergio Spagnuolo, founder of data journalism agency Volt Data Lab and leader of Núcleo data journalism website.

While there has been over time a shift in the understanding of what constitutes a primary source for news coverage given the central role of computing in every human relationship (Hammond, 2017), journalists are also adapting journalistic fundamentals and patterns to dialogue with these sources. Takahashi explains that the questions made to a dataset must have the same elements of a question

made to the sources used previously (Cushion et.al, 2017), which could be a spokesperson, specialist, politician, an authority on the subject or a member of the general public.

I think we put on the journalistic prop by choosing wisely the question, which is a journalistic question, a question that could be made by a reporter to an official source, for example (Takahashi, 2020).

All the interviewees seem to conform to a similar pattern where the discussion led to the production of content. This logic slightly differs from the path that news coverage usually takes, which is to follow the sources statements.

Table 8 What is the nature of data driven content produced?

Standards and style to deal with non-human sources	
Fabio Takahashi	So, this is one possible way, it comes from us. Another way is when it comes from other departments or from the newspaper board, they believe we can join then we talk and see if it makes sense to join indeed.
Gabriel Zanlorenssi	So, we write down all the ideas, even those that might not be plausible, or practical, or interesting. We write down everything. Then we start by studying the feasibility of that idea. After that, we usually take these ideas to the meeting with the team and from that they turn into a journalistic agenda. I even insist a lot with my team that we cannot put on the agenda anything that we don't have a 100% feasibility certainty.
Rui Barros	If there is an idea, or a fact worth being investigated. I prepare myself to investigate it, read about it, research what kind of data might exist about it, what information exists, and how I have to collect this.

Sergio Spagnuolo	We usually look to what is happening in Brazil now and then look at certain subjects. We do not do too much brainstorming. Basically, the other editor, Alexandre Eurico, and I discuss: ‘Well, this subject is worth doing so let’s follow it. [...] I presume that we will have a more organized process to define the subjects and agendas. For now, it is pretty much experimental.
-------------------------	---

On the Núcleo website, the most covered subjects are transparency of public resources, transparency of public politics, media, technology and information security, the environment, smart cities and human rights politics (policies?). As the initiative is in its first semester of existence, some of the procedures now adopted will mature with time.

Still on the perspective of new sources in journalism, Spagnuolo makes an interesting point about a practice that might converge with the reconfiguration of purpose demonstrated by Gray et al. (2012) in which data journalism helps us to understand complex issues. In Spagnuolo’s statement it is possible to see that this supporting nature also applies to journalists themselves and not only to the final readers.

For now, data journalism is there to help traditional journalism to deal with these data [...] What is journalism? What is not journalism? Nowadays it is possible to develop an application and call it journalism, see? Journalism does not need to have that rigid formula with lead, graphic, video, declaration, instead you can have several formats of journalism (Spagnuolo, 2020).

Barros emphasizes, though, that in this reconfiguration of purpose, statistical knowledge is crucial for dialogue with these sources. As Nguyen and Lugo-Ocando (2016) argue, to master basic statistics competences has become inevitable for the journalist in a Big Data context where news coverage is

increasingly supported by “number crunching” (Nguyen & Lugo-Ocando, 2016, p. 6).

The questions I ask of the datasets depend on my knowledge in statistics, for example, and because of that I insist on constantly reading about statistics, about new elements, new examples (Barros, 2020).

It is worth noting that the challenges involving data analysis require a diversified team from the perspective of those interviewed. Takahashi, for instance, says that a multidisciplinary team helps to bring quality to data-driven projects. However, the absence of journalistic instincts among data-related professionals might set some limitations.

There are a lot of professionals with no educational background in the field of journalism actually. Here we have two from Biology, they are biologists, so they have experience of working with genetics and this sort of thing. So, they know and are pretty well able to code, do good statistical analysis as well, but they do not deliver the final product, do not write the text. During their time here, they begin to develop their journalistic instinct, they know when there is a story, which could result or not in an article. Yet, there is a restriction, they will not delivery everything (Takahashi, 2020).

The multidisciplinary element seems to permeate the journalistic work at Nexo newspaper as well. Besides the journalists in charge of news coverage, the newspaper also has a four-person team dedicated to more complex data-driven stories and visualizations.

So, the team is an *infographer* named Lucas, and Lucas has an intern named Carol, my intern Gabriel who was just promoted, and I. The work is somewhat divided like this: data and art, precisely data and infographics. We are a team inside the newspaper, but we have a certain autonomy,

see? We work for an editorial division of the newspaper, but we are not in the editorial hierarchy. Usually, when we work with journalists it is for some special content. Once in a while, we create some surveys the writers ask us for, but that is not our job description (Zanlorenssi, 2020).

Barros, from Público, explains that collaboration through a multidisciplinary team is welcomed in data journalism practice because it brings quality to the interpretation much as a qualified data professional brings caliber to the analysis of the datasets.

The data journalist is also thankful for that [collaboration] for two reasons. The first one is that there are things which I have not mastered and if you are specialized in Health, for example, I prefer that you work with me because you will be able to give the proper angle. To have two professionals working together is important because you help one another. Besides, when you have finished a data analysis you cannot think straight anymore and to have someone to be there available to assist you and write is very important. Therefore, I end up writing a lot of projects in partnership indeed (Barros, 2020).

The accountability element seems to be an important side of data work in the realities studied as all those interviewed share similarities in their perspectives regarding complementary sources. The statements reveal again that most data journalism work is indeed supported by the journalistic fundamentals rather than the technique itself (Cushion et al., 2017; Howard, 2014; Gray et al., 2012). Yet, the rigor adopted in the data analysis seems to surpass the treatment of classical sources. It was noticed, especially in Spagnuolo and Barros' statements, signs of a slight increase in the patterns of classifying a source as adequate in a data-driven piece.

Table 9 What is the nature of data driven content produced?

Complementary sources	
Fabio Takahashi	Our stories are like any other Folha de S. Paulo reports. If there is an accusation against the Ministry of Health, for example, probably we will also check it with Health specialists, and find witnesses for the story.
Gabriel Zanolenssi	If we decide to do, for example, a graphic showing how to prevent coronavirus, anything that suggests a person taking an attitude based on what you recommended. I really do not like this very much because we are not health specialists and cannot say: 'do this or that'. So, it is complicated for a journalist to act like that. It is a case of check with a specialist, always seek for confirmation.
Rui Barros	The work with data does not exclude traditional journalism, on the contrary. When I have a dataset I still need to question people, reach them. The advantage is this: in one scenario I call PR and say 'my data are telling me that the government has increased by three the costs for doctors', in another one I asked the PR 'how much has the government been spending?'. Of course, the PR work is legitimate, but I am depending on someone who regulates the information.
Sergio Spagnuolo	If it is a description, we will only describe the data available. So, it depends on the format we are adopting at the time. What I do not want is something like a war of versions. Also, I do not want someone to corroborate my data, I want a person somehow unbiased. If we have to consult a specialist, it must be someone unbiased. I will not talk to someone, for example, that I know for sure that will defend a side, no matter what. So, it has to be someone more inclined to science than to politics.

5.2.3 Educational provision insights

Still on the topic of literacy, the data was assessed to see what kind of educational provision is offered in Brazil in the data journalism field, especially from universities. The data collected are from two educational institutions from two distinct Brazilian regions and with different organizational models.

At the Catholic University of Brasilia, a private university, there is a post graduate program with an emphasis on data journalism. On the course, according to its syllabus, the student would be able to develop the multiple skills needed (Stray, 2016) through classes that would be teaching how to code, CAR elements (Howard, 2014; Gray et al., 2012; Felle, 2016) and statistical analysis.

We have identified an existing demand from the market. I have conducted research in all the universities in the country and could not find any specific postgraduate offered in Brazil (Marques, 2020).

According to Dr. Marques, the one-year program starts debating by the role of data in journalism, more specifically towards the local freedom of information laws, and strategies to verify and refine data sets. Yet, in the first module the students are already learning statistical analysis. In the second and third modules the students begin to learn how to code and create visualizations. Marques highlights, though, that content taught in the program provides a more basic approach as he acknowledges that not every interested student has previous experience with data journalism.

Actually, here we start with basic knowledge because many professionals do not have any educational background in coding. We know, they do not have the understanding. But, well, the idea is that the professional becomes able to develop, to handle instruments to create maps, graphics, statistical analysis, that they become able to differentiate, to use

automated resources and so on. The idea is that the professional develops a good understanding over the course (Marques, 2020).

Although Marques affirms that they have found demand for data journalism practice, he believes that the high cost of the course, created in 2019, has been a barrier to many journalists. The tuition for the one-year course could reach R\$ 12,510 (approximately € 2,120). For comparison, the minimum wage in Brazil is R\$ 1,045 per month.

I have six entrants, but I can only start a class with at least eight. I have to wait. [...] The cost is high. Who has better conditions to do it? Those ones in corporate communications. We have been organizing lectures about data journalism precisely to engage more people (Marques, 2020).

Besides the postgraduate course, the Catholic University of Brasilia also reformulated the undergraduate journalism program according to the new demands from Brazilian Ministry of Education established in 2013. The University's updated bachelor's degree program now offers three module courses related to data journalism, from concept to coding and visualizations skills. The class Professional Practice – Data journalism (English for Prática profissional – Jornalismo de Dados) has a total of 160 hours, while the classes Data Visualization and Coding in Journalism (English for Visualização de Dados e Programação em Jornalismo) have a total of 80 hours respectively.

We do not verify facts the same way we did in the past. So, we need to master several tools. All I am saying is that the professional who will work with this kind of agenda, he will really need to develop this kind of profile (Marques, 2020).

Dr. Marques' concerns about the ways of dealing with sources in news coverage reflects the shift in journalism that follows the remodelling of the roles that

computers and data now play as an active party in every human relationship as argued by Hammond (2017).

The coordinator at the Catholic University of Brasilia's also seems concerned as to who will teach the new skills on both courses. The demand is not only for a professor who masters the techniques, but also for those who have the journalistic knowhow.

Because even the professional that will teach statistics and mathematics, or something related to it, must know what journalism is indeed and through which ways this very journalism operates, develop agendas (Marques, 2020).

Additionally, the geographical arrangement of the professors able to do the job is an issue with a specific impact on the cost of the course.

Another great difficulty is to have professors to teach here. There are professionals in São Paulo. In Brasília, you can count them on one hand. So, there are costs to bringing professionals here and accommodating them. We must bring someone, it won't be possible to bring everyone because the costs will multiply (Marques, 2020).

The problem of finding Professionals capable of transferring the new skills into training contexts is a remark already made by Heravi (2019). She found that many instructors do not have the expected academic credentials in the field, which might be a side effect of data journalism being a practice still in consolidation. In Marques' view, the challenge is centered on how updated the professionals in general are, considering the current digital reality. Barbosa, on the other hand, highlights that this should be an inherent feature of any academic professor.

Table 10 What is being offered in Brazil in the academy context in terms of data literacy?

Faculty qualification	
Alberto Marques	The design professor, for example, who is familiar with teaching graphic design but did not stay updated won't get there. These transformations end up stressing out the updating process for the professors.
Suzana Barbosa	This is part of the job description of the professor in any field. He needs to keep updated. [...] On the other hand, researchers have this feature, which is reality data, that obviously means to say that as a researcher you need to keep updating yourself

The Federal University of Bahia also started in 2018 a reformulation of its journalism program, based on the same Brazilian Ministry of Education demands established in 2013, and concluded the phase with a journalism bachelor's degree structured in six integrated modules. According to an article written by Dr. Barbosa for the publication *Práticas e Tensionamentos Contemporâneos no Ensino do Jornalismo*, edited by the University, the classes Integrated Journalism V and VI (English for *Jornalismo Integrado V e VI*) have a greater level of complexity in terms of using software to deal with datasets, coding skills and visualization techniques. Yet, Barbosa remarks that the main change is that the entire syllabus was designed to be applied as an integrated process.

Specifically in the case of our pedagogical project, one of the changes, in fact, I think that is the main pillar, what really changes, is the fact of it being integrated [...] Nowadays, the student needs to understand the multimodal character, which is specifically work through distinct arrangements, by being aware that it is probable to have to work with not

only one kind of language, format, platform. We know that journalism basically works that way (Barbosa, 2020).

To respond to the new educational needs, the University hired a new professor with a digital journalism background and with a multimodal profile.

We organized a public tender last year precisely to hire a professor from the journalism field, but keeping in mind this format of the needs identified in the pedagogical project elaboration. So, the professor must have the profile of a professional that works in those realities. We have already accomplished the public tender and have an approved professional with this profile (Barbosa, 2020).

5.2.4 Perception of data literacy among professionals

Those interviewed also express their interpretations of the level of data literacy that their peers have been demonstrating. It is worth highlighting that these comments concern the professionals that have been in touch with the interviewees in work or educational contexts.

From an academic point of view, the multidisciplinary element of generating news should be developed yet in an educational context, alongside the required techniques. This might suggest a more comprehensive approach while supporting the teaching of new digital elements based, however, on the convictions of traditional journalism:

Table 11 What is the participants perception of the level of journalists' data literacy?

Educators propositions	
Alberto Marques	To teach coding, [...] it is ideal that the professor comprehends it, but I believe that this is a field in which we have to work with multidisciplinary groups, as we see happening in journalistic

	organizations in general. [...] I believe this interdisciplinarity approach must also occur in the educational context
Suzana Barbosa	Like it or not, however small, I have seen transformations in the production process. So, journalism will still be journalism, the journalism canons do not disappear, likewise the fundamentals, the ethical issues, but new elements are added

From the multiple skilled nature (Stray, 2016) perspective, Sergio Spagnuolo states that acquiring the coding skills might be laborious rather than complex.

In general, the technical part is easy, once you have mastered / grasped it. I am not the greatest coder ever, the greatest specialist etc. There will always be someone able to extract data, formalize it, it will be laborious, but the technical part is workable most of the time (Spagnuolo, 2020).

On the other hand, as was verified in the production routines studied, journalists usually have a certain autonomy to deal with data from the start until the final product when the data skills required are basic rather than complex.

Table 12 What is the participants perception of the level of journalists' data literacy?

Autonomy to deal with data	
Fabio Takahashi	Here we have indeed journalists, with a major in journalism, that do a little coding. So, they can deliver a final product, but these are simple things because they are not that advanced in coding
Gabriel Zanlorenssi	Everybody here, the writers, they basically handle data. Normally, they do what that with Excel and with the internal visualization tool that we adapted for the web site, right? So, they just insert the data and it generates a graphic". Yet,

	Zanlorenssi highlights that data literacy among the journalists is even more important to detect the inconsistencies that might appear in data sources. “You can only notice errors, inconsistencies, when you have a certain level of familiarity with these things
--	--

Yet, the understanding of the data as a result of an infrastructure with purpose and preconceived motivation (Gray et al., 2018) seems to be more effective than solely using technique to detect the inconsistencies that might appear in data sources. “You can only notice errors, inconsistencies, when you have a certain level of familiarity with these things” (Zanlorenssi, 2020).

Still from a production perspective, both Takahashi and Zanlorenssi present similar views on a journalist’s autonomy in dealing with data. As the complexity of the sources and the project itself increases the autonomy level decreases.

There is great coding capacity, fine statistics knowledge, but sometimes the project is too big and the professional cannot even process information properly anymore after dealing with so much data. So, here we have these cross-checks, let us say, from outside [...] the professional will never have total autonomy, I will check what exactly is happening and usually another person from the team will also check (Takahashi, 2020).

Additionally, Zanlorenssi stresses that the time element also dictates the autonomy of the journalist since a more complex and dense analysis could take hours in the hands of a journalist with data familiarity, but minutes in the hands of an experienced professional in the data field.

I believe the hardest part for them is when some research is a little more demanding than usual, more complex. For us it will be a trivial thing, something you handle in five minutes, but they will spend the entire morning doing it. [...] So, this kind of work that sometimes involves a larger amount of data, right? Something bulky. Normally, well, a lot of things that they do are studies and reports with tables, graphics and summaries already made (Zanlorenssi, 2020).

Possibly, this critical observation remains as an unavoidable characteristic of the realities studied. Sergio Spagnuolo, for instance, reinforces that the most difficult aspect of data journalism work is to make sense of the datasets, as argued by Boyd and Crawford (2012). Reasonable results might come, however, from repetitive work.

The hardest part of all is to make sense of data, see? Because, well, sometimes you did not quite understand that data, even published data is sometimes wrong, it happens with everybody, at least I believe so. So, to do the right analysis and focus on the data, understand the data, it is the hardest part of all parts. It is tricky, it takes time, it is even frustrating, repetitive (Spagnuolo, 2020).

Spagnuolo compares the density of the data to noise from which critics are required to retrieve the meaning. "Most difficult is to make sense from all the noise existing in the data" (Spagnuolo, 2020).

Fabio Takahashi, for example, counts on the critical element even when choosing his data journalism team. And for those who are not journalists, he was conscious that it would take time to develop 'a nose for the news' (Gray et al., 2012).

I believe that without the journalistic instinct being a super coding professional will not serve much. Even when I was assembling the team

here [...] so, if they are not journalists, it will take time to acquire the instinct for news (Takahashi, 2020).

Yet, Dr. Marques, raises the concern that the journalistic instinct might be missing even amongst journalism students. “I am talking about looking at the information and not knowing if it is new or old” (Marques, 2020).

Dr. Barbosa seems to share the same concerns as Dr. Marques. She is emphatic about how crucial it is that the student knows how to use digital tools to mediate knowledge to the audience rather than solely developing digital skills. From her perspective, this critical observation development is based on the journalistic fundamentals.

First of all, and as always, the critical view aspect is essential above all, and much more. Because, sometimes, people may think that it is easier now [...] to find these sources and deal with them. But it does not mean that I do not have to follow journalistic fundamentals and criteria, those concerning the validation of information and which are the basic principles of journalism (Barbosa, 2020).

The understanding of data as a product from a given infrastructure, with predefined purposes, is at some point associated to critical observation reasoning. Godler et al. (2020) argue that even information created through technology embodies the human consciousness of its creator. As well as in the data creation, the human consciousness is also present in data gathering, a process which Berti-Équille and Borge-Holthoefer (2016) reinforce can also be misleading, incorrect, outdated and, therefore, unreliable. And to understand these features is to understand that datasets are not totally free from bias and deliver a world in conformity with a given vision, culture and value (Gray et al., 2018). A critical view of the data infrastructure element might help the journalist to avoid giving in to the

false authority of the data (Appelgren, 2017). This debate on the nature of data seems to coincide with Zanlorenssi's perspective on biased data.

Usually, data from consultancy firms are a bit odd, especially when the firm has some interest. Normally, what is blocked by the ethical barrier of the newspaper are any kind of graphic, stories that, for example, are endorsed by companies with specific interest in that data, right? So, I also believe that this is something evil that several companies paid for stories in newspapers to show some data, some positive information for them. So, this is something that we never do, this is one of our ethical limits (Zanlorenssi, 2020).

5.2.5 Credibility associated with data journalism

The news production cycle has acquired a more irregular pattern due to an information environment overwhelmed by the need to always report something, and preferably something fresh (Bakker, 2012). With this aspect of the information reality in mind, the last question made to the data is about the perception of those interviewed of the credibility associated with data and the interest in data-driven stories.

Sergio Spagnuolo considers that data journalism has a great potential for credibility since the sources, at least in his experience, usually can be revealed, a feature that helps bring transparency to journalism.

A positive side to working in data journalism is that there does not exist, or at least in most of it, 95% of the time, off the record sources, the sources are usually open. So, I think that data journalism has a huge potential for credibility, gigantic. Everything is documented. So, well, the return that this bring to us in terms of transparency is great, see? (Spagnuolo, 2020).

Naturally, the data are not all used every time to bring transparency and credibility to a piece of information. As Nguyen and Lugo-Ocando (2016) stress, even fake news can be supported by numbers, created from the misuse of data. Although numbers are at the center of data journalism work, in both a positive and negative sense, Dr. Barbosa reinforces another side of data journalism concerning a broader context where the numbers are not necessarily the final news product.

Data journalism is not solely the numbers. The numbers are indeed very important, but there is a lot of information that is not numerical. Data journalism might not just be necessarily to observe how many stories are using statistics, if there are spreadsheets, graphics, etc. This is so much more complex (Barbosa, 2020).

This broader context seems to be part of the work developed by the data team at Folha de S. Paulo, where the focus is to tell worthwhile stories based on datasets and not show precise graphics with numbers.

I see a slight difference from other data journalism groups to ours, which is the fact that other projects are pretty much based on visualizations, great interactive projects and this sort of thing. Here we do great Folha de S. Paulo reports, but we are specifically doing it from great datasets (Takahashi, 2020).

The audience profile was also verified among the professionals who work in media organizations. Some patterns are visible as most of them classified the regular readers as professors, data scientists, and university students. Even at Folha de S. Paulo, where the final results are data-driven stories usually with the same appearance as the other stories in the newspaper, there is a portion of readers that differs from the general readership.

There is a segment of our public that is indeed a bit more specific than the general readership of Folha de S. Paulo. We cannot help it, although I have said during this conversation that we all seek to match our materials as much as we can to the other Folha materials. I think there is a public that search a bit more for evidence of the news, see? More concrete facts, less emotion (Takahashi, 2020).

Table 13 What is the perception of credibility associated with data journalism?

Audience Profile	
Fabio Takahashi	Although I believe we can appeal to all kinds of public, those who comment most on Twitter are, for instance, university professors, from institutions like USP. It seems to be something more related to the academic field.
Gabriel Zanlorenssi	There are many postgraduate students, university professors. I would say that it is an audience with a high level of education. So, if we make a mistake it never passes without being noticed.
Rui Barros	We have a loyal public who are people with an interest in statistics, data science, coding. Therefore, this kind of audience read your stories not only because of the subject. They read it because they know it is data journalism.
Sergio Spagnuolo	Lawyers, managers from the third sector, communication coordinators. We sort of write to this kind of readership [...] There are people from the government texting me, telling that they will use it in inquiry commissions, or telling me that the work we are doing is pretty interesting.

It is true that data driven journalism might guide the audience on more complex issues (Gray et. al, 2012). However, that feature of data journalism can be enhanced or diminished by the level of editorial autonomy (Cushion et al., 2017) and transparency (Allen, 2008) in the journalistic practice. As argued by Appelgren (2017) if the process that leads to the story is not clear to the readers, then they

are being deprived of a certain transparency. Yet, this transparency towards the process shared, seem to have room for improvement

All the professionals from media companies interviewed affirmed that they share the source of their information. Spagnuolo, however, argued that to have clarity it is necessary to share, besides the sources, the methodologies used to create the content.

I believe that credibility is directly associated to transparency, right? When you disclose your data, you inform from where did you take it, you cite the sources, when you use an off the record source you explain why it is (off the record), and give as much information as you can about this source. [...] If you access nucleo.jor.br, you can see that all the stories have a link to the data, and at the bottom of the article there is the methodology used, original links to the source, everything in terms of transparency (Spagnuolo, 2020).

Rui Barros, from Público, also expressed clearly his efforts to share the process that led to the production of specific content (Allen, 2008).

Make sure of explaining the methodology very well : this and that was made, that was excluded. I really insist on that [...] because it is very important. So, that kind of reader who is more interested, that goes beyond only reading the methodology, can himself reproduce the analysis I have made and eventually see where I might have committed a mistake (Barros, 2020).

In Barros' reasoning regarding reader participation it is possible to note the trend to increasingly involve the reader in the process of journalistic verification. In this scenario, however, audience participation would not be through commentaries and shares only, but in the checking of the facts.

5.3 Considerations

The following considerations are organized by the theme of each question made to the data sets. Therefore, with regards to the influence of educational background on the search for proper literacy, the professionals interviewed search for proper literacy outside the journalistic scope, mostly through related areas from social science and by themselves. This behavior reinforces the multiple skilled nature of data journalism (Stray, 2016).

Although there seems to be some consciousness among those interviewed that a journalist cannot handle the mass and complexity of Big Data alone (Splendore et al., 2016), there is an attempt to exhaust the datasets through devices, a movement similar to the black box dynamics of Flusser (1983), in which humans try to penetrate the post-industrial devices by playing with them.

In relation to the characteristics of the content, the stories seem to present similarity in format, with graphics and interactive pieces. The exception seems to be Delta Folha's stories which resemble other Folha de S. Paulo materials, as stated by Takahashi.

In relation to the way the professionals carry out data analysis, there seems to be a pattern regarding the questions that are made to datasets - they are based on journalistic fundamentals. However, statistical knowledge is an essential tool for asking effective questions as advocated by Barros: "The questions I make to the datasets depend on my knowledge in statistics, for example, and because of that I insist on constantly reading about statistics, about new elements, new examples" (Barros, 2020). Additionally, Zanlorenssi, from Nexo, reinforces the value of having a "certain familiarity level" with statistics to "notice errors, inconsistencies" (Zanlorenssi, 2020).

An important topic identified in the data is related to the educational provision for data journalism in Brazil, verified by Dr. Marques and Dr. Barbosa's statements. In this specific context data skills are still being integrated into the course syllabus. Although there is an acknowledgement of the value and necessity of developing this new literacy associated to statistics and coding, structural issues remain as the main challenges.

The main concern of both educators seems to be how the professors will adapt their experiences and knowledge to a developing teaching form that includes the digital element in the critical observation of facts. The need for professionals capable of transferring the new skills into training contexts is a remark already made by Heravi (2019).

Still on the topic of literacy, the perception of a journalist's data literacy among the interviewees coalesces into a belief that there still needs to be better developed skills to handle more complex sources. Dr. Marques and Dr. Barbosa agreed on the point of view that the multidisciplinary element of generating news based on data should be developed in the educational context, alongside the required techniques.

However, a journalism instinct supported by the understanding of the data as a result of an infrastructure with purpose and motivation (Gray et al., 2018) tends to be seen by those interviewed as more effective than solely using the technique to detect the inconsistencies that might appear in data sources.

Finally, the credibility element of data journalism was also explored in the interviews and identified in the data. Most of the participants recognized that the regular readers of their stories are specialists, members and workers from the field of education and authorities.

Spagnuolo, for instance, affirms the materials of the Núcleo portal are sometimes used by official bodies in enquiries commissions, as well as lawyers, communication leaders and professionals from the third sector. Takahashi acknowledges, despite the effort to keep the stories in the Folha de S. Paulo's format, most of the regular readers of Delta Folha materials are university professors. A similar audience is identified by Zanlorenssi, from Nexo, while Barros, from Público, considers the specific audience for his stories as professionals interested in statistics, coding and data science.

One issue, however, remains uncertain specifically how much of the process through which the news was created should be revealed to the audience (Allen, 2008; Appelgren, 2017). While Spagnuolo and Barros clearly advocated for revealing details of the methodology used in the data analysis, the others interviewed seemed to focus only on the responsibility to reveal the source of the data, with no emphasis on the process as a whole.

6. Conclusion

Clearly journalism now faces a reconfiguration of production routines and especially of what constitutes a source. Datasets, especially enhanced by an open data movement led by governments and companies (Rogers, 2011), have been giving journalists new possibilities for checking facts and creating new narratives.

Indeed, specific skills allied to critical thinking and questioning make data journalism a potential tool in reinforcing press credibility in a digital age where the news production cycle has a more erratic pattern due to an informational environment overwhelmed with the necessity of always reporting something and preferably something fresh (Bakker, 2012). Now data journalism practice is starting to become better known, a literacy demand for a specific academic education starts to become visible.

Thus data journalism can be seen both as a development of past techniques such as CAR (Howard, 2014; Felle, 2016) and as the acquisition of the new skills required to deal with much bigger datasets in the Big Data context (Nguyen and Lugo-Ocando, 2016). While the practice is much better established in the United States and some European countries, as is data literacy – as Heravi (2019) highlights most data journalism related courses are in the United States – data-driven journalism shows signs of flourishing in countries like Brazil and Portugal.

This assumption suggests it is time to develop further studies about the data journalism experiences in these countries, with special attention on how the educational establishment are preparing to offer a proper grounding with a balance between techniques and journalistic fundamentals. It is a case of considering a complete literacy rather than emphasizing one, or another, aspect (Heravi, 2019).

Naturally, not every journalist will become a data journalist and not every data scientist will eventually work in journalism. But educational provision balanced with

elements from both realities might accelerate knowledge-acquisition for those interested in this field. Moreover, every journalist will have to deal with data, to a greater or lesser extent, since computers, and consequently data, have begun occupying a central position in human relationships (Hammond, 2017).

The data collected for this research confirms the hypothesis that journalists have to learn a new and specific literacy to question datasets, which are increasingly occupying the place of primary source for news coverage. Indeed, a journalist has to develop new skills to deal with non-100% human sources. As a matter of fact, the skills to be developed are not natural in journalism, but pretty much based on social science methodology, with the emphasis on statistical analysis. In relation to that, coding skills should be learned in order to obtain a faster, more accurate analysis.

Part of the data collected for this study shows that journalists and professionals related to data journalism have been pursuing data skills outside the epistemic journalism environment and even developing the skills by themselves. It is the case of Rui Barros, from Público newspaper. “I have been reading a lot about statistics and doing something that is probably not the best learning process, but it has been working well, which is to learn something as I need to develop a new skill” (Barros, 2020). Other professionals seem to pursue additional qualifications to address data properly. Sergio Spagnuolo, for instance, has a master’s degree in Social Science from the Pontifical Catholic University of São Paulo, while Fabio Takahashi was a fellowship researcher in the field of journalism with emphasis on education at Columbia University. In this period he studied data journalism concepts.

Moreover, for more complex work with data it was noted there is a trend to search for professionals from other fields in order to question this new type of journalistic source. As a matter of fact, regarding the production routine, a plausible

consideration could be that data journalism can better serve if merged into a concept of work in which content creation is similar to an industrial production chain. In this case, the content, as a product, is created by a multidisciplinary chain with different phases, professionals and tools devoted to each part of the production. This chain production was verified in the routines studied.

At Folha de S. Paulo, for instance, a data journalism team is formed by journalists and professionals from different fields, such as Biology, Political Science and Social Science, with experience in dealing with data. At Nexo, a four-person team of data scientist, designer and coder supply the newspaper with data-driven content and support the journalists on major data projects. At Público, the data journalist Rui Barros works in collaboration with an experienced journalist in charge of covering different subjects. The Data Vol Lab is an agency that provides content and consultancy to other newsrooms for data journalism projects.

It is important to note, though, that this chain production concept is not new in journalism, as highlighted by Dr. Barbosa. Reporters from newspapers and magazines are used to working in collaboration with photographers and designers to add images and other visual elements to their pieces. However, more complex work related to how to handle sources that are not as well-known as before, has been added to the notion of chain production. While data can be considered as a higher form of knowledge (Boyd and Crawford, 2012), a new professional profile emerges, then, to help to dig through datasets and ask the right questions.

Meanwhile, some academic courses in Brazil are starting to be reformulated while the few options dedicated to data journalism studies are being organized. It seems to be a time of thinking about a higher educational provision that embraces both journalism fundamentals and specific coding and statistics skills.

As data journalism practice is still evolving, and initiatives outside mainstream reality are starting to consolidate, an interesting idea to develop would be how to

properly share the methodology, not just the source of information, as an element of transparency in journalism. From the understanding of data as the main feature of digital culture and knowledge mediator (Godler et al., 2020), it is plausible to reconsider the levels of transparency in journalistic production since solely sharing the sources might not be sufficient anymore. Meanwhile, sharing the process by which the news was generated (Allen, 2008) can be an alternative for the audience who can persevere in the acquisition of knowledge-acquisition by replicating the models and procedures used by the journalist.

Future research on data journalism should continue to study how data journalism has reconfigured journalistic production specially from the perspective of new initiatives regarding the production cycle and learning context.

With regards to the production cycle, it would be valuable to carry out more extensive research into the development of the website Núcleo, not only for the nature its data journalism, but also for the transparency standards adopted on the portal, an important element in a digital reality permeated by multiple sources. All the details about the methodology used in the analysis process are shared at bottom of Núcleo's articles.

With respect to the learning context, further research about the advances and challenges faced in the field of higher education in Brazil could be interesting. The Federal University of Bahia recently reformulated the Journalism bachelor degree with new digital competences related to data as source in an integrated learning environment. Besides the reformulation of their journalism bachelor's degree, the Catholic University of Brasilia went further and organized a postgraduate degree exclusively in data journalism. Therefore, to monitor this initiative that is in its initial stages could bring interesting insights about data literacy in regions outside the United States and Europe.

Finally, data journalism seems to have reconfigured journalistic production specially related to data literacy and transparency. As with any transformation that has ever occurred in the profession, it is time again to incorporate new and more complex skills to deal with technology, whether working alone or in collaboration with others. However, the level of transparency seems to gain new layers with endless possibilities to address and share new sources of information, and consequently give the audience the option to continue the knowledge-acquisition by themselves.

7. References

- Allen, D. S. (2008). Journalism Studies THE TROUBLE WITH TRANSPARENCY
The challenge of doing ethics in journalism ethics in a surveillance society
<https://doi.org/10.1080/14616700801997224>
- Al Jazeera (2017). Broken homes – A record year of demolitions in occupied East
Jerusalem. Retrieved from:
<https://interactive.aljazeera.com/aje/2017/jerusalem-2016-home-demolitions/index.html>.
- Appelgren, E. (2017). An Illusion of Interactivity: The paternalistic side of data
journalism. Journalism Practice.
<http://dx.doi.org/10.1080/17512786.2017.1299032>
- Bakker, P. (2012). Aggregation, content farms and huffinization: The rise of low-
pay and no-pay journalism. Journalism Practice, 6(5–6), 627–637.
<https://doi.org/10.1080/17512786.2012.667266>
- Bardin, L. (1977). Análise de Conteúdo. In Revista Educação (Vol. 22).
<https://doi.org/10.1017/CBO9781107415324.004>
- Barros, R. (2020) Coronavirus: Como achar a curva? O que revelam as
experiências dos países. Retrieved from:
<https://www.publico.pt/interactivo/coronavirus-como-achatar-curva-que-revelam-experiencias-paises>.
- Berti-Équille, L., & Borge-Holthoefer, J. (2015). Veracity of Data: From Truth
Discovery Computation Algorithms to Models of Misinformation Dynamics.
Synthesis Lectures on Data Management, 7(3), 1–155.
<https://doi.org/10.2200/s00676ed1v01y201509dtm042>

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, 15(5), 662–679.

<https://doi.org/10.1080/1369118X.2012.678878>

Bradshaw, P. (2018). Data Journalism Teaching, Fast and Slow.

<https://doi.org/10.1177/1326365X18769395>

Cushion, S., Lewis, J., & Callaghan, R. (2017). Data Journalism, Impartiality and Statistical Claims Towards more independent scrutiny in news reporting. *Journalism Practice*, 11(10), 1198–1215.

<https://doi.org/10.1080/17512786.2016.1256789>

Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional "fourth estate" function. *Journalism*, 17(1), 85–96.

<https://doi.org/10.1177/1464884915593246>

Fink, K., & Anderson, C. (2015). Data Journalism in the United States. *Journalism Studies*, (4), 467–481. <https://doi.org/10.1080/1461670X.2014.939852>

Flusser, V. (1983). Vilém Flusser. *Film, Theory and Philosophy: The Key Thinkers*.

https://doi.org/10.1007/978-3-531-91158-8_36

Folha de S. Paulo (2020). Delta Folha. Retrieved from:

<https://www1.folha.uol.com.br/deltafolha/#10>.

Gates, B. (2018, May 31). Innovation for pandemics. *New England Journal of Medicine*, Vol. 378, pp. 2057–2060. <https://doi.org/10.1056/NEJMp1806283>

- Godler, Y., Reich, Z., & Miller, B. (2020). Social epistemology as a new paradigm for journalism and media studies. *New Media and Society*, 22(2), 213–229. <https://doi.org/10.1177/1461444819856922>
- Gray, D. (2004) *Doing Research in the real world*. Sage Publications. London: Sage Publications
- Gray, J., Chambers, L., & Bounegru, L. (2012). *The Data Journalism Handbook*. Retrieved from <https://datajournalism.com/read/handbook/one>
- Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. (December), 1–13. <https://doi.org/10.1177/2053951718786316>
- Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism*, 18(4), 408–424. <https://doi.org/10.1177/1464884915620205>
- Hewett, J. (2016). Learning to teach data journalism: Innovation, influence and constraints. *Journalism*, 17(1), 119–137. <https://doi.org/10.1177/1464884915612681>
- Heravi, B. R. (2019). 3Ws of Data Journalism Education: What, where and who? *Journalism Practice*, 13(3), 349–366. <https://doi.org/10.1080/17512786.2018.1463167>
- Howard, A. (2014). *The Art and Science of Data-Driven Journalism*. Tow Centre for Digital Journalism - a Tow/Knight Report. <https://doi.org/10.1017/CBO9781107415324.004>
- Laney D (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6(70): 1

Loureiro, B. (2020). Mapa mostra limitações do isolamento vertical nas periferias.

Retrieved from:

<https://nucleo.jor.br/urbe/2020-04-02-mapa-densidade-saopaulo-coronavirus>.

Lowrey, Wilson, and Jue Hou. 2018. "All Forest, No Trees? Data Journalism and the Construction of Abstract Categories." Journalism. Advance online publication. doi:10.1177/1464884918767577.

Manyika, J., Michael, C., Brad, B., Jacques, B., Richard, D., Charles, R., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, (June), 156. <https://doi.org/10.1080/01443610903114527>

Marconi, M., & Lakatos, E. (2003). Fundamentos de metodologia científica. In Editora Atlas S. A. <https://doi.org/10.1590/S1517-97022003000100005>

Megan Knight (2015) Data journalism in the UK: a preliminary analysis of form and content, Journal of Media Practice, 16:1, 55-72, DOI: 10.1080/14682753.2015.1015801

Mona, C. (2020) Coronavirus is revealing how broken America's economy really is. Retrieved from:

<https://www.theguardian.com/news/datablog/2020/apr/06/coronavirus-american-reaction-economy-covid-19>.

Nexo Jornal (2020). Gráfico. Retrieved from: www.nexojornal.com.br/grafico/

Nguyen, A., & Lugo-Ocando, J. (2016). The state of data and statistics in journalism and journalism education: Issues and debates. Journalism, Vol. 17, pp. 3–17. <https://doi.org/10.1177/1464884915593234>

NYTimes (2020). The Upshot. Retrieved from:

<https://www.nytimes.com/section/upshot>

Núcleo (2020) Jornalismo, dados, transparência. Retrieved from:

<https://nucleo.jor.br/>.

Popovich, S. (2020) The World is still far from Herd Immunity for Coronavirus. Retrieved from:

<https://www.nytimes.com/interactive/2020/05/28/upshot/coronavirus-herd-immunity.html>.

Pueyo, T. (2020). Coronavirus: Why You Must Act Now. Retrieved from:

<https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca>.

Rogers, S. (2011) Data journalism reading the riots: what we know. And what we don't Retrieved from:

<https://www.theguardian.com/news/datablog/2011/dec/09/data-journalism-reading-riots>.

Rogers, S. (2013). Facts are Sacred. London, England: Guardian Faber Publishing.

Splendore, S., Di Salvo, P., Eberwein, T., Groenhart, H., Kus, M., & Porlezza, C. (2016). Educational strategies in data journalism: A comparative study of six European countries. *Journalism*, 17(1), 138–152.

<https://doi.org/10.1177/1464884915612683>

Shoemaker, J.P., Vos, P.T. (2009) Gatekeeping theory (pp. 9-22). New York, USA: Routledge.

Stray, J. (n.d.). The Curious Journalist's Guide to Data - Columbia Journalism Review. Retrieved March 17, 2020, from Tow Center for Digital Journalism website:

https://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php

Union, T., Journal, O., & Union, E. (2018). General Data Protection Regulation. 2, 1–78.

United Nations. (2019). VALUE CREATION AND CAPTURE: IMPLICATIONS FOR DEVELOPING COUNTRIES. Retrieved from <https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2466>

V., Mayer-Schönberger & K., Cukier. (2013). Big Data: A Revolution that Will Transform how We Live, Work, and Think. New York, USA: Houghton Mifflin Harcourt Publishing Company.

Zamith, R. (2019). Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. *Digital Journalism*, 7(4), 470–489. <https://doi.org/10.1080/21670811.2018.1554409>

Zeit (2020). Datenjournalismus. Retrieved from: <https://www.zeit.de/datenjournalismus>.

8. Appendices

Appendix 1 – Transcription of the interview with Dr. Alberto Marques

Alberto Marques – 02/24/2020

Professor and course coordinator at Catholic University of Brasilia

[Sociodemographic characteristics]

Minha formação é em jornalismo, me graduei em jornalismo. Eu tenho uma trajetória que é um pé na academia, um pé no mercado. Eu me formo na universidade particular com uma vocação muito de mercado em Sergipe, Universidade Tiradentes. No sétimo período eu descubro uma coisa chamada web jornalismo, inclusive minha professora não era uma pessoa que estava muito ligada à pesquisa, mas que já falava de tecnologias, mudanças, e isso fez meus olhos brilharem. Foi o primeiro momento que eu tive contato com a UFBA, com a produção da Federal da Bahia e aí escutei falar do Marcos Palácios, André Lemos, que não era nem tanto jornalismo, mas que já ouvia falar de cibercultura. E todo mundo falava: 'Ai! Lá na UFBA tem rede WiFi, tudo é muita tecnologia'. Então os meus olhos brilhavam, né? É incrível porque eu acho que foi a primeira que ouvi falar e tinha certeza, digo: 'É isso. Agora eu descobri o que eu quero pra minha vida'. Na época você não tinha os gerenciadores de conteúdo, tudo era feito na mão. Primeiro manual de jornalismo do Elias Machado e o Marcos Palácios dizia: 'Se for publicar uma foto, cuidado com o tamanho da imagem porque demora a carregar'. Então tinha uma série de limitações técnicas, mas eu sabia que era aquilo que eu queria. Então comecei a fazer todos os cursos possíveis, web design...o que tinha da parte técnica e o que tinha de leitura, de possibilidades de leitura eu saía abraçando, correndo atrás freneticamente. E antes de terminar minha graduação eu me inscrevi numa pós-graduação Lato Sensu que se chamava Mídias Digitais. Era uma pós-graduação nova em Sergipe, também numa universidade particular e que não tinha nem definido ainda seus professores. E o que eu faço? Eu vou conversar com a coordenadora e digo: 'Você tem que contratar esses caras'. Então, aí veio o Palácios, vem a Carla Schwingel que na época já dava aula, era doutoranda, mas já dava aula em vários lugares na Faculdade de Salvador. Aí veio a Simone Bortolheiro que era uma moça que já falava de TV digital e outras possibilidades e isso foi praticamente uma abertura pra eu me aproximar dos grupos de pesquisa em jornalismo online, né, que era coordenado pelo Marcos Palácios na época. O Elias também ainda estava lá na UFBA. Então a minha trajetória foi calcada dentro daquele grupo, assim, quando consigo me aproximar e começar a participar das reuniões foi o momento que o blog bombava (gjol.net), escrevi igual louco, se não tivesse o que postar eu sofria, é sério. Hoje eu reconheço que na época era como uma doença, algum distúrbio, porque eu ficava tremendo quando não tinha o que postar. Isso abriu diversas portas porque passei a conhecer todo mundo. Até hoje tem pessoas que falam: 'Ah! Você é o Alberto, que publicava'. O blog era muito conhecido. A gente tinha disciplinas com gente de fora, com o

Scolari, com o Javier Diaz Noci, com uma galera do mundo, né? E eu devo toda a minha trajetória a esse grupo. Ele realmente foi que me preparou, forneceu o arcabouço teórico para que a gente de alguma forma conseguisse produzir. Diferente da maioria dos membros do grupo, todo mundo foi logo engatando direto, fazendo mestrado, doutorado, e eu [fui] pro mercado. Na verdade, eu resolvi fazer esse mestrado em 2010, foi quando tive uma experiência no Congresso trabalhando no portal vermelho, portal de esquerda, e eu percebi 'olha! agora já deu minha vida profissional e eu quero agora rumar para a vida da pesquisa'. Porque no fundo, assim, eu sempre achei que o jornalismo fosse um instrumento de transformação social, uma ferramenta. Com essa ferramenta eu consigo de alguma forma transformar realidades e contribuir socialmente de algum jeito para coisas que eu acreditava que seriam boas. Eu acho que com a maturidade, porque tem isso, né, na vida acadêmica ela nem sempre você tem maturidade suficiente para perceber, ou para avançar ou sei lá, não sei. Bom, de alguma forma, a trajetória seja no mercado, ou seja, na pesquisa, ela é muito difícil, né? Aí depois que eu entrei no mestrado eu dei continuidade. Fiz mestrado na UFBA com o Palácios, do jeito que eu sonhava. Aí tem um detalhe: eu não pesquisei jornalismo, fui para cibercultura. Como a gente escrevia muito no blog, eu fui pensar o que que levava alguns blogs ter mais visibilidade do que outros, reputação. Eu era tão apaixonado por blogar, e toda aquela dinâmica, e tava tão impregnado, assim, todo mundo que tava ali dentro respirava isso. Tanto que diversos trabalhos lá dentro no âmbito do grupo foram sobre blog: Gonzalo, a Raquel. Algumas pessoas trabalharam blog ali dentro em diferentes perspectivas. Bom, só que eu fui criado dentro, no centro de um grupo de pesquisas de jornalismo e eu tinha um débito assim...tinha a sensação que não tinha feito nada. E aí no doutorado, como eu tava dentro da pesquisa da Suzana que era aquela coisa do jornalismo convergente eu fui e fiz uma proposta de pesquisa sobre tablets. Então eu fui pesquisa rotina de produção dentro das redações de dispositivos móveis (tablets) que eram uma produção exclusiva ali pra dentro. Só que como eu bebi muito dentro da católica, eu digo: 'Pronto! agora eu preciso de uma universidade nova e preciso ir pra mais para os clássicos. Eu preciso ler aquilo que vai me dar base para seguir, eu tenho que discutir os textos seminais, né? Eu queria ler. E a UNB sempre ficou no meu imaginário, eu digo: 'Ah! Esses caras! Ah!'. É engraçado que eu chego na UNB e eu não tenho a mesma produção que eu teria na UFBA no sentido dos grupos de pesquisa, daquela produção coletiva. Aquelas discussões entre as pessoas. Isso me fez muita falta. A produção coletiva. As pessoas pensando junto com você no seu projeto. Isso faz muita falta. Por mais que tenham as disciplinas, mas é diferente você conviver com as pessoas de 15 em 15 dias, se encontrar pra discutir textos, sabe? É uma descoberta coletiva. Você descobre texto juntos, debate junto. Por que por mais que sejam temas diferentes, é jornalismo digital, ou ciberjornalismo, ou é jornalismo online, isso de alguma forma faz toda diferença. Bom, no meio da minha trajetória o meu objeto de pesquisa morre, né? Os jornais começam a fechar os dispositivos [9"30']. Foi me dando um desespero. Eu volto do meu sanduíche (Portugal) querendo mudar de projeto, chegando com projeto de apuração, fake news, nada a ver, e aí minha orientadora diz: 'Tá é louco! Vai

mudar nem a pau, você já qualificou, agora volta. E aí eu resolvi trabalhar com pauta, pensar pauta, produção de pauta para o jornalismo móvel, ou para o tablet, pior ainda. Bom, é isso. É o que dá. No fundo, você descobre com o tempo que tese e dissertação o que vale é a trajetória, não é exatamente o seu produto final. É fato. Quem é doutor não é doutor que fez uma tese, mas porque viveu o processo. Viveu isso. Eu trabalhei em televisão, trabalhei em rádio. Trabalhei cinco anos com artístico, trabalhei no portal vermelho, fiz muita assessoria de imprensa que é diferente de Portugal. Aqui a gente mistura. Inclusive tem gente que defende que a assessoria de comunicação pode ser jornalismo.

Data Literacy

Tem dois pontos aí centrais: tem a pós-graduação que é de jornalismo de dados e tem o curso de jornalismo digital. A graduação não é mais jornalismo, é jornalismo digital agora. A gente mudou tudo. A gente já tinha feito uma reformulação a partir das novas diretrizes, né? Porque todos os cursos no Brasil mudaram com as novas diretrizes do jornalismo, que eu acho que é 2013, não tenho certeza. Mas a gente identificou, primeiro, eu acho que é importante, tem que pontuar que é uma universidade particular. Isso é um ponto central porque a gente depende de pessoas ingressando, pagantes, para que o curso exista. Então por muito tempo a gente tem tido dificuldade em ensalamento na graduação. Essa é uma dificuldade geral, não é da Católica de Brasília, inclusive os cursos das federais, em algumas também o ensalamento reduziu. E aí não sei dizer exatamente porque, mas a gente suspeita que um dos reflexos seja justamente inicialmente a queda do diploma e automaticamente, mais recentemente essa campanha toda contra o jornalismo. Na verdade, as pessoas que possuem mais visibilidade têm colocado em cheque a profissão e tem uma série de elementos aí. Eu realmente não tenho muita clareza para afirmar o motivo do que está acontecendo. Esse semestre a gente instalou e conseguiu fechar uma turma. É a nossa primeira turma do curso de jornalismo digital. E por que virar jornalismo digital? Há mais ou menos dois ou três anos a gente vem discutindo no NDE (Núcleo Docente Estruturante) que existe uma série de transformações, é um conjunto de transformações muito complexo que eu acho que o termo convergência ele dá conta, não exatamente na perspectiva do Jenkins, mas a convergência jornalística mesmo. Porque o cenário de produção de informação ele está tão complexo e envolve tantas questões que isso sem dúvida sem dúvida afeta também a profissão [jornalismo]. Aí, a partir de uma série de discussões a gente começa a pensar, não sei se você conhece, em 2019 a Digital Journalism publica uma edição, o volume 7, o número é o 3, é uma edição inteira pra definir o que é o jornalismo digital. E é engraçado porque a gente tinha fechado a nossa grade e quando sai aquela edição eu vejo e digo: 'Meu Deus! Olha pra isso!'. Aí mando lá no grupo. Por que? Porque o termo jornalismo digital, ele não está mais circunscrito apenas no suporte a um ambiente. A gente tem uma digitalização generalizada, e essa digitalização possibilita um fluxo dessas informações por diferentes estratégias, seja crossmedia, transmedia, então a produção está extremamente complexa. É óbvio que isso vai mudar dependendo de estado pra estado, nas organizações elas vão ser diferentes, isso natural que aconteça. Mas, por mais simples que seja essa

organização, você chega a uma rádio all news (CBN ou Band News) o que você tem é rádio pura e simplesmente, mas você tem o repórter ali se preocupando em fazer uma passagem pro Instagram, preocupado em como empacotar aquele conteúdo para colocar para aparecer não sei aonde. Queira ou não queira, por menor que seja transformação, eu tenho práticas, eu tenho transformações no processo de produção, isso vai demandar tempo de alguma forma. É mais uma preocupação. Então, o que a gente foi pensar é que precisa criar um curso, isso na graduação, que de conta desse processo de circulação com ambiência digital. Porque quando eu tô falando de jornalismo digital, eu tô pensando numa coisa muito mais complexa que é a ambiência que é de que forma essas informações vão circular e de que forma esse profissional vai se preparar para pensar nisso. E o que a gente faz? Coloca uma disciplina chamada jornalismo de dados, coloca uma disciplina de programação, coloca uma disciplina para visualização, coloca disciplina de redes e mídias sociais, coloca disciplina para pensar em mobilidade. Então, você tem as disciplinas clássicas, porque jornalismo continua sendo jornalismo, porque os cânones não somem, os valores, as questões éticas, só que acrescenta outros elementos. Porque não adianta colocar redes e mídias sociais e não fazer discussões que são básicas, sabe? A disciplina precisa dar conta dessa complexidade. Aí a questão é que a gente precisa também de um nome que seja mais vendável porque eu estou numa universidade particular, ou seja, além de ter toda essa transformação você precisa também mostrar às pessoas que tem diferencial, eu tô propondo diferente, ter uma redação integrada para produzir e distribuir nos diferentes suportes [17"4"], eu tenho estágio supervisionado que tem que dá conta de produção de podcasts para circular nos suportes que são do curso. Isso a gente repensou. Está da forma que a gente queria, não está. Mas Brasília tem particularidades. Então jornalismo político é obrigatório, a gente precisa ter. A gente teve ele [graduação] aprovado em 2019. O MEC aprovou e a gente tem a primeira turma agora. E no primeiro semestre, a gente já entra pra discutir, é uma disciplina... a Católica cada semestre tem uma disciplina chamada práticas profissionais, então são disciplinas com carga horária alta de 160 horas e que tento já trazer uma aproximação deles para produzirem conteúdos digitais, pensar em plataformas, eles precisam criar narrativas, pensar no conteúdo nessa perspectiva. Aí no segundo semestre eles já têm jornalismo de dados e eles começam a ver.

A pós-graduação em jornalismo de dados. A gente identificou que existia uma demanda mercadológica. Eu fiz uma pesquisa em todas as universidades do país e não encontrei nenhuma pós-graduação Lato Sensu no Brasil sendo oferecida. O que eu fiz? Identificado que existia essa demanda, que a universidade tinha uma demanda por novos cursos, então a universidade queria criar novas pós-graduações e tal. Aí eu comecei a me debruçar sobre todos os cursos que existia de oferta no mundo, Columbia, aquela universidade de Miami, de Londres, aí vi as ementas e material e tal e fiz uma proposta de curso. É um curso que tem duas disciplinas generalistas, que é de formação geral, né? Criatividade, empreendedorismo, assim, que trabalha bem numa perspectiva de uma formação para todos. E uma ética. Por que? Porque pelas regras da Católica, é preciso de

duas disciplinas que ensinam geral. Então eu não tenho como fugir. Diferente das outras pós-graduações do País, na católica [20"49] sistema de 60 horas por módulo, ou seja, mesmo que você tenha módulos curtos que ensinem pinceladas 15, 20 horas, a gente concentra no assunto e bota o cara para pensar ali, lendo, estudando aqueles módulos, em vez de fazer aqueles módulos rápidos e curtos. Geralmente a gente bota um ou dois professores por módulo pra também não passar 60 horas praticamente com um. Praticamente são oito ou 10 encontros só sobre uma temática. Depois dessa, eu tenho uma disciplina que se chama numeralização do jornalismo, do computador aos dados. Essa disciplina, vamos dizer assim, tem outro detalhe que acho importante: esses módulos são pensados para serem independentes. É. Isso é um desafio. Por exemplo, nesse módulo, o cara vai pensar o papel dos dados no jornalismo, é o comecinho. É essa discussão que a gente estava tendo da digitalização da informação e a disponibilização de dados, é o uso de lei de acesso à informação, apuração e estratégias de verificação a partir de dados. É um módulo bem basilar, informação e transparência, legislação e estratégia jornalística para conseguir dados. E aí tem o detalhe: nesse módulo entra também matemática e análise estatística. Por que? Porque...eu tô falando isso porque tem o seguinte: o cara que procura esse curso, que jornalista geralmente é um bicho curioso, ele quer saber de tudo. Então muitos deles chegam e não têm a dimensão de que eles vão ter que ter noções de matemática, de programação. Ele precisa ter. Então alguns se assustam, eu já ouvir gente dizer assim: 'Não, então não. Isso não dá pra mim não. Não tem condições, não'. E eu culpo isso talvez seja pela própria baixa formação que a gente tem, né? Então as pessoas dizem muito assim: 'Ah! Eu sou de humanas. Não vou disso'. Geralmente é alguma experiência traumática que o cara teve e ele quer correr. Na minha graduação eu tive estatística e eu lembro que penei pra passar, não foi simples. Eu tenho uma outra que é programando dados, eu tô falando numa ordem mais ou menos solta, aí aqui a gente pra Python. É Python puro. Python, os pandas, aí vai tudo: variáveis, entradas, saídas, operadores lógicos, estruturas de decisões. A ideia é que o cara mergulhe pura e simplesmente em Python aqui, nas APIs, data frame, a estrutura do Panda, é isso. O cara tem que entrar aqui e não precisa saber de lei de acesso à informação, ele vai aprender a programar. Aí eu tenho uma outra que é jornalismo computacional. Observe que aí praticamente é uma continuidade daquele primeiro que eu estava falando. Aqui eu tenho base de dados estruturados, formato mais comum, raspagem, limpeza, análise. Eu já estou aqui, vou, vamos dizer assim, se eu dissesse que há um encadeamento, este seria o segundo módulo, depois daquela primeira que tem que pensar pauta, receber e pedir dados, lei de acesso à informação, aí eu venho para cá. Eu vou ter aqui visualização básica, mapeamento de dados públicos, agrupamento, modelo de espaço vetorial do documento, algoritmo de filtragem, é o básico. E depois eu tenho visualização, mas é bem na perspectiva de D3, é o básico. Na verdade, aqui vai do básico, porque muita gente não tem formação nenhuma de programação. É porque a gente sabe...ninguém tem noção. Mas, assim, a ideia é que o cara consiga já construir, que já saia manuseando, criando mapas, gráficos estatísticos, tentando diferenciar, tentar usar mapas, criar mapas, criar coisas

automatizadas, e por aí vai. A ideia é que o cara saia daqui com uma noção boa. [O curso foi criado] o ano passado também. Eu tenho seis escritos, mas eu só consigo abrir turma com oito. Tem que esperar. Não tenho opção. Outra dificuldade absurda, ter professor para ensinar isso aqui. Tipo, São Paulo, você tem. Brasília, você conta nos dedos. Porque no fundo é assim: você tem gente com experiência no mercado, mas que nunca lecionou nada. Então, não adianta você pegar um cara de redação e colocar dentro de uma pós-graduação que você vai ter um problema. Porque esse cara não tem o timing e o feeling natural. Então, assim, tipo, não vou falar nomes, não é legal. Muita gente que está no mercado está indo pra lá. Tem um cara pronto que dá aula de jornalismo de dados, que é o Marcelo, acho que é Marcelo, ele foi da folha meu Deus, Marcelo Alves? Ele é muito bom, tem experiência, deu praticamente aula em todos os cursos, mas é de São Paulo. Então tem custos de trazer, de hospedar, tem uma série de coisas que a gente não tem como fugir. Vai ter que trazer alguém, mas não vai dar pra trazer em todos, senão o custo do curso dobra. Quem procura muito é gente de redação ou gente que está se formando, acabou de se formar na universidade e enxerga como possibilidade de atuar. Em Brasília, só o Metrôpoles, que é um jornal online, tem um setor de dados. Você tem também o J, mas o Jota faz um meio termo entre o jornalismo e ciências de dados para parlamentares, tem consultoria. Você fica no meio termo, o Jota tem uma redação. Você tinha aquele Vortex, que acabou fechando e ... tinha uma sucursal lá. As sucursais dos grandes jornais não têm equipe de dados em Brasília, geralmente. Equipes de dados geralmente estão dentro dos jornais online, é curioso isso porque, inclusive eu tô trabalhando nisso num artigo como Rodrigo Cunha a gente tá tentando entender. Se você está dentro de uma redação convergente, que empacota conteúdos para vários lugares, você tem um setor de dados, esses dados são trabalhados para quais suportes? De que forma eles circulam? Tipo: 'Ah! Eu trabalho na Globo. Se eu trabalho na redação da Globo que é integrada e eu produzo para dados, os dados vão pra quem, pro Jornal Nacional? Ou vai pro G1, pro Globo? A gente não sabe como eles circulam e como é que se pensam pautas para que suporte. No Estadão geralmente é o próprio setor, eles são demandados, isso eu ouvi do Rodrigo Menegat, às vezes eles são demandados, raramente por política ou economia, mas muitas as vezes as pautas surgem deles que são independentes. Eu tenho dificuldade de dizer qual é o perfil exatamente. Tinha uma galera do Correio Brasiliense que queria fazer. Mas qual é o impeditivo? Grana. O custo é alto. Em Brasília, as pessoas ganham bem. Meu primeiro salário em Brasília eu me assustei, eu não sabia que era possível. E mesmo assim pessoas têm dificuldade. Assim, você não tem universidade pública oferecendo Lato Sensu, né? Especialização. E os que querem fazer as vezes têm dificuldade, ou seja, quem é que tem mais condições. O pessoal que tá na assessoria de comunicação. A gente fez uma palestra sobre jornalismo de dados, justamente para atrair mais pessoas. Então, eu tinha praticamente metade do público era de assessoria de comunicação, sabendo que os jornais demandavam também dados estruturados, fontes de visualização. E os caras preocupados em produzir também conteúdos nessa perspectiva. Sabe que a maior reclamação é a estruturação desses dados. Cada lugar disponibiliza dados em um formato

diferente. A Anac tem uma API, um dado estruturado, e os caras estavam lá reclamando, discutindo na assessoria de imprensa, que mesmo assim os jornalistas ainda continuavam errando dados, passando informações que não eram corretas, justamente pela dificuldade de ler os dados.

A pós-graduação e a graduação são dois momentos distintos, sabe? O aluno de graduação ele tem dificuldade básica mesmo, dificuldade em criar pauta, em olhar para uma informação e perceber que aquela informação é mais importante que a outra. Para eles... eles têm uma série de dificuldades mesmo. Eu não estou nem falando de extrair uma pauta de uma segunda camada, olhar para os dados e ligar eles a uma outra coisa e pensar que assim: 'Ah! Eu já tenho dos dados das eleições. Então agora eu vou no IBGE e vou pegar os dados para cruzar'. Eu não tô falando disso. Eu estou falando de olhar para informação e saber que aquela informação é nova ou velha. Eu estou falando do básico. Eles não querem ler. Não estão informados. Não percebem a importância de estar atualizado. Olha que coisa mais boba. Por que é importante saber o nome do Presidente da República? Por que é importante saber de onde ele vem e tal? Então eles não percebem. É uma geração que vem desconectada desse universo. Em compensação, se você vai conversar de Anitta, Lady Gaga, qualquer coisa, eles vão matar a pauta. Eles vão saber tudo. Qualquer MCs desses eles sabem tudo. Mundo pop é encantador. Isso só chega com a maturidade, só chega com alguns anos depois. Chega e chega com força. Você encontra um aluno do primeiro período com os olhos brilhando, encantado e tal. E vai encontrar com ele no quinto, ele está de saco cheio, já sabe tudo. Às vezes você encontra com ele no final do curso e ele já está com outra... continua de saco da graduação, já sabe tudo, mas agora ele está preocupado com aquilo que ele abraçou em algum momento. Eu tenho um caso de um aluno que hoje é repórter do Globo, cobre energia e tal e que no começo ele ... então assim, você tem diferentes experiências. Cabe ao papel do professor de tentar a todo instante despertar, mostrar...é uma coisa que eu costumo fazer na sala de aula. Eu começo todas as aulas das disciplinas práticas discutindo acontecimentos, locais, nacionais e internacionais. O cara precisa me dar conta. E eu não quero saber porque o Kim Jong[-un] explodiu a bomba do não sei o quê. Eu quero mais, eu quero entender questões geopolíticas, quero entender o que têm por trás. Se eles não me disserem, eu vou abrir o mapa, vou olhar onde é que fica, vou atrás de uma informação, da economia, como é que gira naquele lugar e tal. Mando entrar no celular. Isso vai despertando de alguma forma o interesse. Agora, esses exercícinhos básicos de fazer um pedido da LAI, isso funciona. Eles não sabem o que é LAI. Então se você começa um semestre com um pedido da LAI, com a formação de um pedido, pra lá no final fazer uma pauta isso super dá certo. Então, eu acho que uma questão de, assim, é igual a questão da leitura de feed, ter um feed alimentado por notícias, por segmentação e tal. Então isso funciona. Eles de alguma...é óbvio que tem algumas ovelhas que desgarram, que não estão fazendo. Mas você realmente transformações. No fundo, assim, minha experiência como professor me mostra que muita gente está descolada do processo pedagógico, assim, você não lê o PPC do seu curso, o plano

pedagógico, e você não entende qual é a função da sua disciplina. Porque as vezes para sua disciplina, você pega ela constrói qualquer coisa, isso acontece muito, você dá do seu jeito, porque você tem interesse nisso ou aquilo, e você acaba não percebendo o conjunto da obra que você está formando, sabe? Que profissional é aquele que você está formando. E isso vai fazer toda a diferença depois. Então esses pequenos grãos, é a função de cada disciplina, esse esforço, no final você tem um profissional que não é completo, porque nenhum profissional dá conta de tudo, mas que sai profissional habilitado, capacitado a atuar aí em diferentes lugares.

[Professores da pós] Então em especialização você precisa ter um professor que tenha experiência docente e que atue profissionalmente. Eu não posso pegar um cara acadêmico e empurrar lá dentro. Porque até um cara que vai dar aula de estatística e vai dar aula de matemática ou alguma coisa dessa ele precisa saber o que o jornalismo e de que forma esse jornalismo operacionaliza, constrói pauta. [O quadro de professores], não tenho professor na graduação para dar conta da maioria das disciplinas. Não tenho. O coordenador está por dentro, está todo mundo por dentro disso. Daqui a pouco, já que tem uma turma rodando, eu vou ter de ter professores que deem conta, de Python, é o processo natural. Tipo, o professor de design, que trabalha o básico de design, que esteve acostumado a vida toda em ensinar diagramação, isso e aquilo, não se atualizou em termos de visualização e isso aquilo não vai conseguir. Eu venho já avisando os professores: 'Gente, eu preciso de um professor com esse perfil. Alguém?'. Aí ninguém. Aí eu falo: 'É, daqui a pouco a gente vai precisar na graduação, gente'. A católica tem um curso de design rodando, que é novo, está no terceiro período. A gente tem nesse semestre, por enquanto, quatro professores de design, novos. Daqui a pouco eu tenho um colegiado inteiro, muitos professores que vão precisar dar conta de visualização. É isso. Essas transformações, de alguma forma, tensionam a atualização dos professores. Se eles não se atualizam, são engolidos por esses novos profissionais. E quando você nas redações, quem tá ocupando as redações de jornalismo de dados é uma galera nova, você não tem jornalistas superexperientes que se atualizou e foi programar. O meu primeiro emprego eu consegui porque eu sabia programar, sabia HTML, sabia vetorizar, e eu sabia escrever. Eu suspeito que os profissionais que vão ter domínio de um maior número de ferramentas, que vão ter maior facilidade para se adaptar a essas coisas, vão ter mais espaço.

Big Data – Volume

Primeiro que essas terminologias são super polifônicas. Jornalismo digital, jornalismo computacional, jornalismo de dados, Big Data, e todas essas tecnologias são utilizadas de diferentes formas, acho que cada um vai dando conta de um jeito. O que a gente tem novidade de todo o conjunto dessas tecnologias é a disponibilização de informação em volumes nunca vistos. Porque não é só a disponibilização, mas também uma memória que não se esgota, né? A grande questão que está por trás é como utilizar esses dados que estão disponíveis para produção e criação de narrativas, esse é ponto central. Porque

às vezes as pessoas dizem assim: 'Ah! Eu tenho Instagram. E aí? Instagram faz parte desse Big Data? Quantas fotos as pessoas sobem por dia? Quantos vídeos no Tik Tok? Esses dados que foram disponibilizados, eles podem produzir alguma coisa? Eu sempre dou como exemplo em sala de aula, sempre digo assim: 'Ah, gente! Acabou de explodir uma bomba aqui no metrô de Taguatinga, que lá próximo da universidade, e todas as comunicações foram cortadas, menos a Internet. O que a gente faz?'. Aí todo mundo fica assim: 'Liga pro bombeiro' - 'Os telefones dos bombeiros estão todos ocupados. E agora o que vocês fazem?'. Então, assim, os alunos têm muita dificuldade de por exemplo: 'Você não publica no Instagram toda hora? Por que você não vai olhar para aquela localização se tem alguma publicação? Vocês não conseguem imaginar uma hashtag? Tentem identificar aí na rede, entrem rapidinho para ver se tem uma hashtag e tal'. Existe uma dificuldade muito grande de entender que está todo mundo munido de dispositivos, independente quais sejam, são moveis e que estão hiperconectados. Essa hiperconexão e esses dispositivos multimidiáticos produzem uma infinidade de arquivos e formatos que circulam na rede. Esse é um dos fatores que geram todo esse boom de produção e informação disponíveis. Eu tô falando de produção social. É óbvio que eu tenho uma série de legislações e leis que vão demandar das instituições e dos órgãos, sejam públicos ou privados, a disponibilização de seus dados. Isso também é motivo de tensionamento de crescimento de todos esses dados que são públicos. A questão central é que, e olha que engraçado, você tem primeiro a cultura do proprietário, a cultura de que eu sou dono dos dados e eu disponibilizo de qualquer forma, e eu disponibilizo meus se eu quero ou não porque eu sou o dono. E tem umas outras pessoas que pensam: 'Não. Esses dados precisam circular, eles precisam ser lidos.' Porque não adianta eu colocar todas as minhas informações na rede se eu não consigo ler. Eu vi um Twitter agora do editor do jornalismo de dados do Estadão dizendo que encontrou uma cartilha do Ministério da Saúde ensinando como responder às perguntas da LAI. E uma das dicas de como responder era identificar se o cara era da imprensa ou não, porque poderia render pautas negativas. Olha que absurdo. Em vez da lógica ser transparência é evitar. Então você tem um desafio muito grande de como acessar essas informações e como essas bases se comunicarem entre elas. Talvez um dos lugares do Brasil que tem uma melhor API, que fornece mais dados, é a Camara dos Deputados, porque eles conseguiram estruturar um canal ali e fornecer informações rápidas. Não é à toa que são produzidas pautas e pautas sobre eles porque eles disponibilizam dados estruturados. Porque a questão é lê. Porque tem gente que às vezes dá acesso, mas dificulta na linguagem. Tem gente que nem acesso dá, fecham todas as suas faixas. Existe uma confusão em termos de legislação dessa coisa de transparência de dados e tal que os caras não têm interesse que isso aconteça. Ninguém tem interesse em disponibilizar porque se enxergar como um inimigo, como o cara que vai queimar, que vai prejudicar. Porque a coisa pública ainda é tratado como algo que...que eu sou detentor, eu sou governante, eu não sou o Estado, eu sou Jair Messias Bolsonaro. Então os caras pensam que são proprietários, e não é só ele, são todos os poderes. Eu acho que a grande dificuldade disso tudo é ter uma sistematização, uma organização desses dados para que as pessoas tenham

acesso. Porque você tem muita gente com literacia, não pode dizer que dentro da profissão não há, é um grupo reduzido. Mas é um processo de transformação. Porque há pouco tempo as pessoas não sabiam o que era sistema de publicação. As pessoas não sabiam o que era nada. Hoje as pessoas estão...gente...você entra no Tik Tok, aquelas narrativas que os caras estão construindo são complexas, não é uma coisa boba. Os meninos estão chegando na adolescência produzindo coisas. Então, assim, isso vai acontecer. Hoje, é engraçado porque as pessoas não enxergam os processos. Você tem software que não lhe demanda. Tipo, os Pandas, os sistemas, as coisas já prontas, se você é iniciado e tal, você isso prontinho e você roda muito rápido. Acho que tem muito mais relação com o interesse, sabe? Porque cada vez mais você tem ferramentas que possibilitam que o cara ...

Data journalism – Credibility

Você vai ter diferentes profissionais. Tem caras que correram atrás, que se especializaram e fizeram acontecer. Eu acho que vai continuar existindo perfis diferentes nas redações, eu não estou fazendo exercício de futurologia, então vai ter os caras que vão realmente trabalhar com o processo. Porque, no fundo, o trabalho em redação acontece muito em parceria, né? Tudo bem, tem o repórter setorista, o cara que tem uma pauta menor, que está trabalhando mais rápido, a pauta do dia a dia. E tem aquelas pautas mais encorpadas, geralmente essas pautas mais importantes, que tem maior densidade, elas são trabalhadas em algumas mãos. E essas mãos, geralmente, o cara que é...nas redações integradas tem uma figura que se chama o produtor, o produtor é o cara que é o antigo pauteiro. Só que é esse cara que enxerga o perfil que o repórter vai ter ...essa pauta é pra fulano, ciclano. Então, se eu tenho uma pauta que vai trabalhar com muitos dados e isso e aquilo, certamente fulaninho vai estar ali dentro, ou seja, pautas que se aproximam da ideia, é que no fundo a própria ideia do jornalismo de dados,

Quando você me perguntou desse conceito eu dei uma tangenciada para lhe dizer assim: 'Olhe! Jornalismo de dados ou Big Data, não tô falando de jornalismo de dados, o Big Data é amplo está lá atingir todo mundo. E ele envolve etapas diferentes do processo. Apurar, não se apura mais da mesma forma. Precisa ter o domínio de uma série de ferramentas. Tudo isso pra dizer que o cara que vai atuar nesse tipo de pauta, ele realmente vai precisar desse perfil. Só que preciso ter o apertado de tecla. Porque você tem aquele repórter mediano. Eu vivo dizendo aos meninos: 'Ah! Você quer o cara que pega o release, esquentar os releases e publica?'. Porque sempre vai ter nos jornais. Esse é repórter mediano. Que não vai receber uma pauta complexa porque ele não tem capacidade de ir além daquilo ali. Então a tendência é que isso cresça, que as redações busquem mais esse tipo de profissional que tem essa competência. Mesmo que elas não busquem, esse profissional é mais completo, no sentido de que ele tem uma capacidade além, ele corre o risco de errar menos porque ele vai ter uma noção de construção de amostragem, ele não vai me chegar e fazer um povo fala e dizer que isso representa a voz da sociedade. Isso é muito a cara desses jornais sensacionalistas. O jornalismo só tem a ganhar se as pessoas se especializarem

mais porque é uma lacuna existente. Os dados sempre existiram, talvez em menor oferta. Os valores estatísticos e os dados as pessoas só não manipulam bem, alguns profissionais, outros são muito bons e estão fazendo um bom trabalho. Essa é uma realidade mundial.

Appendix 2 – Transcription of the interview with Gabriel Zanlorenssi

Gabriel Zanlorenssi - 02/21/2020

Data Scientist at Brazilian newspaper Nexo

Sociodemographic characteristics

Eu fiz ciências sociais na USP e depois ingressei no mestrado em ciências políticas, terminei faz nem um mês que foi minha defesa. Minha experiência vem mais dessa área de humanas com uma certa aplicação quantitativa porque ciência política mesmo sendo na ciência humanas ela é bem quantitativa. Por exemplo, acho que a gente tem várias matérias na graduação e pós-graduação que são especificamente pra você lidar com dados, referências, estatísticas, programação. Então por exemplo eu aprendi o R durante a pós-graduação porque tinha uma matéria que era exclusivamente disso. Agora vou fazer alguma como aluno especial que é estatística bayesiana. É uma outra abordagem de estatística que parece mais complexo porque eu justamente vou fazer aula pra entender melhor.

Sempre gostei de mapas, gráficos e tudo mais. Mas foi durante a iniciação científica. Minha orientadora de iniciação científica era professora de métodos quantitativos. Então quando fiz iniciação com ela foi quando tive oportunidade de ir mais a fundo nisso, aprendi o Stata e minha iniciação era um grupo de pesquisas que é o Cepesp da FGV, por mais que estive na USP, grupo era na FGV. Então lá fiz essas pesquisas assim de setor público, governo e tudo mais que fui me desenvolvendo pro Stata. Mas tive de me converter pro R, né? E me converti pro R justo nessa disciplina que tive no mestrado.

[Primeiro trabalho com dados no Nexo] Sim! Foi uma coisa até meio...assim...inesperado na minha vida porque nessa época que estava na iniciação no mestrado eu fazia várias coisas e postava no meu próprio Facebook...uns gráficos, uns mapas, umas coisas e postava no Facebook. Aí em um deles um amigo meu marcou minha atual chefe no post e ela viu minhas coisas e me chamou pra fazer a entrevista porque no mesmo momento o responsável por isso, o cientista do Nexo tinha saído de lá e ido pra Folha. E coincidiu de abrir essa vaga, ela viu meu perfil lá, gostou e chamou pra entrevista. E aí foi meio assim. E também não fui chamado de primeira assim porque acho que elas ficaram um pouco receosas porque eu estava fazendo mestrado, né? E de como eu iria equilibrar essas duas semanas. Mas as pessoas que elas chamaram não ficou nem uma semana lá e foi pra outro emprego.

Production

[No Nexo] desde julho de 2017, vai fazer ... dois anos e sete meses. Então, a

gente é uma equipe de quatro pessoas. Sou eu, um infografista que é o Lucas, o Lucas tem uma estagiária que é a Carol e o meu estagiário Gabriel acabou de ser efetivado, vai ser meu assistente. Nós somos uma equipe de quatro e meio que dividido assim: dados e arte, dados e infografia pra ser mais preciso. A gente é uma equipe dentro do jornal que a gente tem uma autonomia assim, sabe? A gente não está na estrutura editorial. A gente trabalha na parte editorial do jornal, mas a gente não tá na hierarquia do editorial. A gente não responde diretamente pro editor chefe, a gente responde pra outras pessoas. Então, porque lá não tem uma pessoa exclusiva de dados, um editor de dados, não tem. E quem acaba fazendo a revisão das nossas publicações é a Marina Menezes, ela tem uma função mais administrativa lá no jornal, né? Diria que ela tem o cargo mais alto, depois das fundadoras. O diretor de redação é o Conrado Corsalette. Mas a Marina tem uma função mais administrativa, mais bastidores do jornal. Então ela é quem edita principalmente o nosso texto, né? Mas a parte de dados quem realmente olha é a Renata Rizzi, ela é uma das sócias e fundadoras. A Renata é doutora em economia e ela gosta bastante disso de dados então normalmente os feedbacks mais voltados a dado e ela quem faz.

Raramente a gente trabalha com outros jornalistas para fazer alguma coisa. Normalmente quando a gente trabalha com jornalista é algum conteúdo mais especial. Vez ou outra a gente faz alguns levantamentos que os redatores pedem pra gente, mas não é a nossa função ali. Então a gente publica três conteúdos por semana, normalmente a gente mesmo pensa quais serão as pautas e qual será a programação. Normalmente a gente mesmo pensa quais serão as pautas e qual será a programação e a gente tem uma reunião diária, né? Que são nós quatro, nossa equipe, a Marina e a Renata e a gente ... a gente... fecha qual que é a programação. A gente fecha três semanas, um mês de programação, três gráficos por semana. Mas eu diria que a maioria das pautas é a gente mesmo, a equipe mesmo, que pensa. A gente anota as nossas ideias pra não esquecer aí pena qual é a viabilidade delas pra poder transformar isso numa pauta, né?

Data literacy

Todo mundo ali, os redatores eles mexem basicamente com dados. Normalmente eles mesmos se viram ali no Excel, na ferramenta interna de visualização que tem assim pro site, né? Só insere os dados e ele gera um gráfico. Chama Chart Builder (Quartz), mas ele está bem defasado porque não foi feito mais atualizações nele. Porque a Quartz desenvolveu o Chart Builder e o nexo fez uma adaptação do Chart Builder só que essa adaptação preciso reescrever quase tudo dos códigos e tudo mais para ficar com o estilo do Nexo. Só que aí o Chart Builder avançou e a nossa versão ficou atrasada então é uma coisa que a gente discute. Não é muito pra gente, a gente não usa muito essa ferramenta, mais os redatores.

Production

Eu e o Gabriel que trabalha comigo a gente trabalha basicamente no R, eu diria que mais de 90% é no R, mas obviamente a gente usa de vez em quando spreadsheets pra fazer uma coleta, alguma coisa, mas não para gerar o gráfico.

Pra gerar o gráfico é só no R. Às vezes a gente usa Qgis, mas também é bem raro. E o pessoal da arte do design, eles usam Illustrator. Eu não consigo dizer se eles usam outras ferramentas para além disso, mas eu diria que 90% do trabalho deles é Illustrator. O Illustrator é vetorial, né? Então, a gente exporta em formato vetorial os nossos gráficos do R. Então é por isso que é o Illustrator que a gente usa. Mas eu diria que basicamente maior parte do trabalho deles é no Illustrator, talvez eles usem ou outra Photoshop pra fazer um GIF uma coisa assim, mas também não é tão frequente.

Então, primeira coisa é ter a ideia. Então a gente anota todas as nossas ideias, mesmo que elas não sejam plausíveis, ou viáveis, ou interessantes. A gente anota tudo. Depois começa a estudar a viabilidade de sair aquela pauta, aquela ideia. A gente leva isso pra reunião de equipe e da reunião isso vira uma pauta. Até insisto muito com a equipe assim de que a gente não pode colocar na pauta nada que a gente não tenha certeza da viabilidade. Se a gente acha que não vai dar, melhor não pautar porque depois tem que acabar tirando se não dar certo. Melhor estudar antes pra ver se tá ok do que ter a ideia. Acho que esse estudo de viabilidade vem antes da pauta. E depois da pauta a gente começa a produção. Começa primeiro em dados. A gente coleta dos dados, faz os gráficos, faz as análises que precisa fazer. E aí exporta pra infografia. Na infografia, eles vão fazer todo o layout final da página, né? E...toda a aparência, o conceito da visualização. As visualizações, a gente tenta pensar meio em conjunto, sabe? A parte de dados e tratamento e tudo mais é mais gente e eles fazem mais a parte final do acabamento e da montagem. Da ideia a pauta pode demorar de uma semana a um ano. Como a gente tem um ritmo de três por semana, eu diria que tem mais coisa normalmente que a gente precisa fazer é... eu diria que provavelmente a gente tem uma média de um dia pra cada segmento da nossa equipe. Se demora mais de um dia, a gente começa ficar correndo contra o tempo. Se a gente atrasa em dados, vai atrasar infografia. Mas acontece as vezes de a gente não ter nada pra fazer e as vezes o pessoal de infografia não tem nada pra fazer porque a gente não conseguiu acompanhar a produção. Eu acho que tem um desafio de ordem prática que é fazer aquele tratamento, análise que as vezes envolvem desafios de programação porque nem sempre os dados vêm tao limpos, bonitinhos. Tem vez que precisa fazer muito trabalho em cima [dos dados] antes de fazer qualquer coisa. É...eu diria que esse é de ordem prática. Mas isso você consegue contornar porque você pesquisa na Internet tem ...normalmente, assim, problemas de programação alguém já se deparou com o mesmo problema antes então já existe uma resposta. Isso eu acho mais tranquilo de lidar por mais que seja as vezes cansativo. Mas o que eu acho mais difícil mesmo é entender o dado, né? Não só a interpretação, mas a qualidade daquele dado. Se ele tem um nível de qualidade suficiente pra gente ...e até um bom parâmetro de qualidade que a gente usa as vezes. Vamos dizer assim: tem um estudo de emissões de carbono que a gente não sabe se é tão confiável assim. A gente pesquisa se outros jornais lá fora já fizeram, por exemplo NY Times, Guardian, se eles já trabalharam com essas fontes, né? Também tem a interpretação que é entender o dado e tudo mais. E tem também uma outra parte que eu acho que é também de pauta de entender qual a relevância jornalística pra

aquilo. As vezes algumas coisas precisam de pesquisa então não dá pra fazer qualquer coisa, né? Por exemplo, a gente publicou um agora essa semana, um de orçamento dos estados, qual que é os impostos que compõem o orçamento dos estados, de onde vem a arrecadação dos estados. Então esse foi um tema que mexi muito no meu mestrado, então eu sabia bastante coisa. Mas mesmo assim eu fiquei uma manha pesquisando pra não dizer nada errado. São coisas bem complexas e assim, a gente não é especialista em todas as áreas que a gente trabalha com os dados. E muitas vezes algum especialista dessa área pode acabar lendo o nosso gráfico, né? E isso acontece porque o Nexo, não sei exatamente qual é o perfil do leitor do Nexo, mas é um leitor que tem muito estudante de pós-graduação, muito professor universitário. Eu diria que é um público com alta escolaridade. Então se a gente comete qualquer deslize, nunca passa em branco. Então dificuldade é um pouco essa também de falar de áreas que a gente não está confortável, né?

Normalmente esses dados de consultoria são meio esquisitos, ainda mais quando a consultoria tem algum interesse... pode ter algum eventual conflito de interesse, né? Dados às vezes que tem algum tipo de inconsistência metodológica, mesmo que seja de uma fonte confiável. Lembro de um, por exemplo, que era os dados de turistas no Brasil, turistas que chegavam no Brasil. Era uma série de dados que mostrava desde 1990 feita pela Polícia Federal. Só que ao longo dos anos a metodologia mudava assim absurdamente e eles não explicavam porque. Mas esse a gente até conseguiu usar uma solução alternativa. A gente pegou um ano inteiro que estava bonitinho um e fez um ano só porque a série estava impossível de ser trabalhada. Outro caso também é se é pesquisa de entrevistas (surveys) a gente sempre olha qual é a amostra. Teve uma pesquisa alguma vez que era, sei lá, de práticas sustentáveis, alguma coisa assim, que a gente recebeu. Você olhava a amostra deles eram pessoas com renda mais alta, influencia. Outra pesquisa que vários veículos sempre dão é aquela da associação de professores das universidades federais e sempre fazer um levantamento que é o perfil dos estudantes no ensino superior. Só que é uma pesquisa que não tem nenhum tipo de desenho de amostra. Algumas universidades mandam por e-mail para algumas pessoas responderem então não são representativas isso não dá pra publicar. Eu não consigo dizer com propriedade. Mas eu imagino que as dificuldades talvez sejam fugir do óbvio, sabe? Pensar formas de mostrar, dispor essas informações que tenham algum pensamento de design mais elaborado. O trabalho deles envolve muita coisa manual que é ficar arrastando coisa, diminuindo forma. O nosso trabalho é bem pouco manual e mais programação, né? O deles acho que envolve muito esse processo manual e mais lento, né? Mas eu também não sei dizer.

Data literacy

[autonomia do jornalista] No tratamento [de dados] a maioria das vezes são os jornalistas que fazem, né? Todo mundo lá mexe sim pelo menos basicamente no spreadsheets e consegue fazer uma análise básica. Acho que o que pega mais pra eles é quando algum levantamento que pra eles é um pouco mais demandante, mais difícil, mas que pra gente é besteira, coisa de cinco minutos e

que eles demorariam uma manhã fazendo. Então isso é uma das dificuldades pra eles. Teve por exemplo um que era de taxa de fecundidade dos países, né? Eu não sei se era taxa de fecundidade ou gravidez na adolescência, mas era alguma coisa assim, que era pelo texto e... [28"20'] a gente coletou. Então foi só entrar no site da ONU, puxar os dados, ajeitar a tabela, pegar as variáveis de interesse, pegar alguns gráficos e fazer mapas. Isso foi pra gente bem besta, mapa é super simples de fazer pra gente, mas, não sei, os redatores acho que nem conseguiriam fazer um mapa por conta própria. Então esse tipo de coisa que as vezes envolve um volume maior de dados, né? Uma coisa mais volumosa. Normalmente, assim, muitas coisas que eles fazem são estudos e relatórios que já vem com as tabelas, gráficos, sumários.

Credibility and interest

Minha perspectiva é que, ainda mais num momento assim um tanto, de tanto problema de desinformação, de fake news, eu acho que os dados permitem você se afastar um pouco do senso comum, se afastar muito de situações que as vezes não são muito...que acabam num embate que não tem muita solução, né? Então acho que os dados acabam trazendo evidências mais robustas pra discussão, né? São coisas...eu dizer que o Bolsonaro não tem base no Congresso, isso pode ser uma afirmação de informação solta, ah Bolsonaro não tem base no Congresso, mas se você olha os dados e consegue mostrar isso pelos dados é muito mais forte do que você falar por impressões. Acho que os dados se contrapõem um pouco ao impressionismo. Em um momento de tanto questionamento sobre credibilidade da informação, eu acho que é como se fosse um argumento de autoridade.

Data literacy

Mas ao mesmo tempo você pode produzir fake news com gráficos com dados. Justamente por isso até em outra questão que é a literacia de dados, de gráficos. Que você só consegue perceber esses erros, essas inconsistências quando você tem certo de nível de familiaridade com essas coisas.

Credibility and interest

Eu acho assim, coisas que expõe indivíduos em específico são mais complicadas, independente do assunto, a não ser que a pessoa seja uma figura pública, deputados, né? Mas, assim, coisas que expõem indivíduos acho que é bem complicado de publicar num jornal. Há temas que são sensíveis, tem que ser bem pensado antes de fazer qualquer coisa. A gente fez um gráfico já faz um tempo sobre deficientes no mercado de trabalho. Então esse é um assunto complicado que a gente pode acabar cometendo algum erro, deslize, então é por isso que a gente conversa com especialistas, né? A gente entrou em contato com uma especialista pra ela conseguir dar uma olhada no conteúdo antes da gente publicar. E coisas assim que são recomendações eu acho que são difíceis. Por exemplo, fazer um gráfico sei lá sobre vamos dizer assim como evitar o coronavírus, qualquer coisa que envolva você ter que dar um...você for fazer a pessoa tomar uma atitude com base naquilo que você recomendou. Eu não gosto

muito disso porque assim a gente não é especialista em saúde, né? Pra dizer faça isso, não faça aquilo, né? É complicado um jornalista fazer isso. E passar por algum especialista, sempre ter uma confirmação [production]. Uma coisa que a gente não faz lá também, que entra em nossa barreira ética do jornal é também coisas que... [35"30'] qualquer tipo de gráfico, matérias que sejam, por exemplo, patrocinados por empresas que tenham interesse naquele dado, né? Então acho também que isso é uma coisa ruim que várias empresas pagam matérias em jornais pra mostrar algum dado, alguma informação positiva pra elas. Então isso é uma coisa que a gente nunca faz também, um dos nossos limites éticos. Mas não é uma coisa que a gente se depara com tanta frequência com o problema. Acho também que tem o limite assim que é o bom senso.

Production

Tem uma coisa que é até uma coisa que eu escutava muito na faculdade de alguns professores, que eles falam muito, que você, uma fonte, uma autoridade você pode ouvir a opinião dela, mas não dá pra você montar um, tomar aquilo como a verdade...pra várias situações você tem que ter alguma forma de mensurar o que está acontecendo, mensurar a situação que seja externa aos atores, né? Porque os atores têm discursos que tem motivações. Até mesmo nesse exemplo: o Bolsonaro tem base, não tem base, no Congresso. Eu não sei se entrevistando os líderes partidários seja melhor. Você pode entrevistar, entende? Mas você vai muito mais pra uma questão de análise do discurso. Não que não seja válido, é muito válido também, mas é outra ...não vai chegar a essa conclusão só...você precisa de alguma informação que seja externa aos atores, né? Toda informação não humana ela foi produzida, desenhada, em algum momento por um ser humano, né? Mesmo um Machine Learning é uma coisa que teve um ser humano que desenhou. Nenhuma informação é 100% não humana. Normalmente essas informações que são bancos de dados, tabelas, APIs, coisas do tipo, tem documentação, né? Então a pessoa fez e pensou tentou deixar de uma forma que a pessoa quando utilizar aquele dado não precise fazer um esforço interpretativo tao grande. Então uma boa prática também é isso de quem produz informação. Se você quer que sua informação seja divulgada, né? Vamos dizer assim, por exemplo, um IBGE da vida.

Appendix 3 – Transcription of the interview with Rui Barros

Rui Barros 03/27/2020

data journalist at Público

Sociodemographic characteristics

Sou jornalista há quatro ou cinco [anos] já não me lembro, mas por aí. E basicamente o meu percurso académico baseou-se muito ...nós em Portugal no ano temos que fazer a escolha entre ciências e tecnologias ou humanidades ou económica, ou seja, uma área de foco que já determina um pouco a área de estudos em que tu vais focar, não é? E eu fui pra matemática ...ciências e

tecnologias, portanto tive matemática, física, química, biologia, ou se, as disciplinas que serviriam depois para um curso de engenharia, de bioquímico, não é? Ciências da saúde, algum coisa assim. Eu já na altura gostava de mais até de ciência do que propriamente de matemática, tinha algum interesse em ciência, mas também gostava muito de história, de ler, da literatura. Portanto estava num limbo que não sabia bem pra onde ir. Portanto acabei por decidir por uma área que achei que daria mais portas, não é? Porque sabemos que ciências e tecnologias acabam por ter mais saída. Não gostei muito e, portanto, concluí esses três anos de ensino secundário mas sem grande interesse, percebi que a matemática não me fascinava assim tanto e que eu gostava de... já na altura me tornei um ávido leitor de jornais, gostava da informação, gostava de ... e portanto quando terminei esses três anos de secundário decidi que queria ir pra um curso de comunicação. E, portanto, fui, ingressei na Universidade do Minho no curso de ciências da comunicação. Quando cheguei ao curso de ciências da comunicação, percebi logo, percebi duas coisas. Percebi que... eu já na altura, pronto, gostava, pra além desse meu interesse, eu gostava de computadores, programação, mas não tinha qualquer conhecimento. E tive no meu primeiro ano um cadeira de métodos quantitativos. E, portanto... é... portanto, estatística pura. E quando cheguei a essa cadeira percebi: pera lá, eu afinal, eu não gostava de matemática, mas eu adorava estatística porque estatística finalmente faz sentido pra mim. É uma matemática que eu consigo entender onde que estamos a querer chegar porque estamos a falar de pessoas. E fazia aqui uma ponte muito interessante entre esses dois mundos. Para além disso, mas pronto estava focado em ser jornalista, na altura nem sequer tinha noção do que que era jornalismo de dados, eu queria ser jornalista. Comecei a perceber... eu entro na minha licenciatura em 2012, em 2012 tínhamos tido o WikiLeaks a pouco tempo e tivemos o escândalo do Edward Snowden e na altura eu comecei a perceber que por exemplo o Guardian, e o Guardian foi uma referencia nisso, que tinha uma coisa chamada data team que era uma coisa assim, uns gajos que faziam um jornalismo com estatística e eu fiquei: ok, isso parece-me interessante. E percebi uma segunda coisa que é eu via aqueles trabalhos espetaculares do Guardian, aqueles trabalhos do NY Times e é também nessa altura que aparece o snowfall que aquele primeiro grande interativo. Portanto eu olho pra aqui e digo: Epa, é isso que quero fazer porque eu não quero ser jornalista de rádio, de imprensa de papel, porque eu percebo que isso é muito mais interessante porque permite-me contar uma história de uma forma completamente diferente, mas eu não fazer isto. Eu quero fazer esses gráficos interativos e salta e rola e não sei o quê, mas eu não sei fazer isso, como é que posso fazer isso. Percebi imediatamente que o curso não tinha nada que me oferecesse nessa matéria e, portanto, comecei a aprender sozinho. E, portanto, comecei a tentar copiar um bocadinho o que é que eles faziam. E olhava pra aquilo e dizia: Epa, isso é interessante, deixa-me tentar fazer. Claro que os primeiros projetos foram muito falhados. Tive ainda a sorte de o curso de comunicação na Universidade do Minho no último ano, a cadeira do último semestre do último ano é uma coisa chamada projeto em que nos somos desafiados a ter que produzir um grande dossiê, não é? Com várias partes. E naquele ano o professor disse: este ano nós queremos que vocês aprendam

sobre jornalismo de dados e queremos que vocês façam coisas com dados, sim, na licenciatura. Isto foi a minha praia no sentido de, pá, tive pela primeira vez a oportunidade...porque até então todos os trabalhos que eu tinha que fazer para a universidade, não se ligavam muito a isso que eu andava a estudar fora, não é? E eu estava a estudar isso fora, quer dizer, era porque eu queria fazer aquilo, mas também era numa: eu não sei o que me vai acontecer no final da licenciatura. Sabemos que as vagas no mercado de trabalho em comunicação são muito poucas e, portanto, mas vale aprender outra forma de conhecimento, neste caso, programação, para, na pior das hipóteses, eu sair daqui ou ir fazer novamente um curso na engenharia da informática ou tentar arranjar aqui uma coisa no mercado. Terminei a licenciatura, segui para o mestrado aqui na Universidade do Minho. O mestrado tinha uma vantagem que era pós-laboral. E, portanto, era sempre à noite quatro dias por semana, quatro ou dois, já não me lembro, acho que, não, eram dois, dois dias por semana. E, portanto, eu passei ater muito tempo livre. E só queria fazer o mestrado porque eu achei: 'se eu sair da academia não vou querer voltar. Portanto, é melhor já encaixar tudo, pronto, e a seguir...ainda havia um bónus que é na Universidade do Minho nos temos estágio no mestrado. E, portanto, o segundo ano é o estágio e depois a dissertação. E, portanto, eu passei a ter...pensei 'ok, tenho bastante livre'. O que que eu fiz? Comecei a trabalhar e tentei sempre moldar os projetos do mestrado a este meu interesse por dados e visualização de dados e programação. Por exemplo, tivemos uma cadeira que não era semiótica, porque já tivemos na licenciatura, mas já não me lembro agora o nome, eu penso que era linguagens e qualquer coisa. Mas basicamente eram os fundamentos eram muito da semiótica. E um dos trabalhos finais era um ensaio e, portanto, fiz meu ensaio sobre a semiótica na visualização de dados. E, portanto, tentei sempre adaptar um bocadinho aquilo que era preciso fazer na universidade a este meu interesse. Quando terminei o primeiro ano de mestrado, fui para o estágio. Fui para o estágio na rádio renascença e, pronto, cheguei lá e estagiei durante três meses. E ainda tinha uma coisa que é eu despedi-me daquele trabalho que tinha até então para terminar o estágio. Portanto, eu disse, na altura disse, eu quero vir para cá, mas quero fazer jornalismo de dados que é isso que eu ... porque na minha cabeça eu ía estagiar três meses num jornal e não ía ter mais a oportunidade de fazer aquilo que queria e, portanto, eu pensei, bom já que vou ter de estar três meses a trabalhar de casa, mas vale trabalhar a fazer uma coisa que eu quero. E, portanto, disse que queria fazer jornalismo de dados e três meses depois convidaram-me para ficar lá onde eu fiquei quatro ou três anos, cinco quatro anos, de 2016 até agora.

[sempre com jornalismo de dados] É assim, quase sempre. Por que?

Porque...não havia jornalistas de dados em Portugal, quer dizer havia a Raquel que está no Expresso, mas que vai fazendo, mas também não tinha propriamente...não estava completamente dedicada áquilo, ela tinha saído do Público para ir pro Expresso. E, portanto, quer dizer não havia cultura de jornalismo de dados em Portugal, em 2016. Não havia, quer dizer, havia esta noção de que o New York Times, o Guardian, o El País, quer dizer, os grandes jornais estavam a fazer, mas não havia cultura e, portanto, foi preciso fazer também aqui um trabalho, não é? Eu tive a sorte de na Renascença tive a minha

chefe principal e o diretor que viam este trabalho e tinham noção e pronto e diziam: 'ah, isto é fixe, vamos fazer uma coisa parecida, podemos fazer uma coisa assim'. Mas, quer dizer, eu não estava exclusivamente dedicado aos dados o que eu fazia um pouco era, tinha esses trabalhos de dados quando aparecia uma coisa assim mais numérica diziam-me: 'olha, o Rui, ficas agora a trabalhar nisto'. Mas quando não havia assim nada e quando eu não tinha nenhuma ideia que tinha proposto para trabalhar, fazia o trabalho normal. Claro que isto foi mudando um pouco ao longo do tempo, ou seja, no início eu fazia mais, como disse, apuração clássica, ou seja, jornalista normal, de texto. Sendo que acabava por cobrir um bocadinho tudo que metia estatística, não é? Diziam-me: 'Oh Rui, podes olhar e não sei o quê'. E acabava por fazer coisas sobre internet, redes sociais, tecnologia, coisas assim eu acabava por fazer um pouco. E depois, quando não estava, ou seja, isso foi mais na fase inicial, na fase final eu já estava quase sempre nesses grandes trabalhos e nesses trabalhos com dados e, portanto, eles iam se encadeando um pouco eu já estava quase exclusivamente [dedicado aos dados], sim. E, pronto, depois, agora no fim de dezembro, entretanto o Público tinha uma jornalista de dados que era a Rita, não se já falaste com ela. A Rita saiu do Público e foi trabalhar para uma startup e o Público lançou-me o convite. Sendo que, pronto, [15'40"] quer pra mim na altura era um luxo que era estar 100% dedicado a esse tipo de produção e pronto, com uma estrutura relativamente maior. Pronto, e foi esse um bocadinho o desafio que surgiu. E acho que é um bocadinho isto.

Big Data - Volume

Então. Há logo aqui uma coisa que é, eu acho que pouco jornalismo de dados é ainda com big data. Por que? Porque quando falamos de Big Data significa que é mesmo muitos dados. Nem é o fato de ser muito complexo, é o fato de ser preciso uma estrutura, é preciso uma coisa base que é a matéria prima disto tudo que são dados. E muito vezes não há, ou quase sempre não há. E, portanto, Portugal tem ainda um problema grave de abertura de dados. E, portanto, quer dizer, as coisas estão sempre muito...os dados nunca são tantos quanto gostaríamos e, portanto, o trabalho nunca é de Big Data por assim dizer, é mais *small* data. Agora, o que eu uso no meu trabalho. Eu tenho esta polivalência que por acabo de fazer a parte de análise de dados, depois o desenvolvimento web, não é, dos trabalhos interativos. E, portanto, eu uso acima de tudo o R como linguagem de programação para a parte de análise de dados e visualização pra análise não propriamente para o produto final, uso o R constantemente. E depois uso mais JavaScript, CSS, HTML para a parte de criar [os produtos]. Inicialmente eu comecei por usar o Excel pra fazer aquelas análises muito básicas, rapidamente percebi que havia muito coisa que não dava pra fazer no Excel e que o Excel era muito limitado, e que escrever o script é a melhor coisa do mundo porque permite que as coisas fiquem todas automatizadas e que eu possa correr aqueles duas vezes. Por exemplo, se houver uma atualizar nos dados, no Excel eu tenho que fazer tudo novo e no R eu já tenho aquilo programado é só voltar e clicar e aquilo corre novamente. Comecei a perceber a vantagem de usar a linguagem de programação para análise de dados. Pronto, é um bocadinho isso. Eu acabo por

trabalhar quase tudo com CSVs, as vezes tenho que usar umas APIs que já estão disponíveis, mas é isso.

Production

Eu diria que não há grande, ou seja, a vantagem de ser jornalista de dados é que és um jornalista especializado sem ter uma editoria. Isso é bom porque um dia estou ...eu como não gosto de ser monotemático e só trabalhar por tema, é muito gratificante pra mim porque posso... eu sei bastante de dados de educação porque faco ranking das escolas há quatro anos e agora vou fazer no Público também. Por exemplo, saúde é uma área que tem muitos dados, principalmente porque há uma iniciativa do governo português que é o portal de transparência do serviço nacional de saúde que tem muitos dados para poderem ser trabalhados. Eu acho que não há propriamente [uma segmentação]. Acaba por ser, lá está, ou seja, eu também fiz agora uma mudança completamente diferente. Na Renascença, nós não seções de editorias, ou seja, tínhamos pessoas que eram especializadas em determinados temas, mas não havia uma editoria de economia...e, portanto, o que acabava por acontecer era que os jornalistas que percebiam melhor aquilo que eu fazia eram aqueles que tinham as melhores propostas para trabalharmos, ou seja, se é uma pessoa que não tem sequer noção do que eu faço tem mais dificuldade em chegar a minha beira e dizer: 'Olha o Rui, era engraçado fazer um trabalho sobre isto e não sei o quê'. Quando eu já tenho pessoas que estão habituadas a trabalhar comigo, que já trabalharam comigo em outros projetos, que têm noção do que eu faço, é mais fácil para eles chegarem a minha beira e dizer: 'Olha, não sei há dados disto, mas era interessante fazer uma coisa que fizesse isto, isto e isto'. E, portanto, eu acho que não editoria que acabam por pedir mais do que outras. A única que eu diria que trabalhei pouco é cultura. Acho que nunca fiz assim nada...fiz alguma coisa com dados do Spotfy para o Eurovisão, mas é coisa assim mais cultural pop, mas, sim, pelo menos da minha experiência da Renascença cultura acabava por ser a área que menos trabalhei.

Production

Eu basicamente tive três semanas no Público e depois fomos todas para casa. Agora, há uma diferença a nível de estrutura entre o que a Rita fazia e o que eu faço é que a Rita estava inserida na seção de sociedade e eu estou na seção do online e, portanto, os meus editores são editores do Online e têm um pensamento muito mais voltado para aquilo que é o online, o interativo. Mas o meu workflow, pelo menos por agora, o que tem sido tem...ou tenho uma base um dados, ou alguém surge com uma ideia, é mais ou menos isto, e quando alguém surge com uma ideia diz-me: 'O Rui podes ver se há dados sobre isto ou podes tentar arranjar uma coisa sobre isto'. Depois faço minhas análises de dados, as minhas conclusões, escrevo...se for um texto mais pequeno, escrevo meu texto, não é? E quando não propriamente uma área que eu domino eu falo com alguém que cobre aquela área e digo: 'Olha, tenho essa conclusão, achas que isto faz sentido?'. Quando são trabalhos interativos, especiais, tenho trabalhado com a equipe de

design e desenvolvimento do publico, como esta vantagem de, como eu também programo permitam-me a estar a trabalhar com eles diretamente no código e consigo, de alguma forma, quer dizer, eles dizem: 'Olha, o Rui, era fixe que os dados já viessem assim, não é, pra nós conseguirmos fazer isto assim, assim e assim'. E eu já faço isto no R, ou seja, consigo fazer esta ponte, é muito mais fácil porque eu consigo entender as necessidades de todos. Tem sido um bocadinho este o processo. Sou o único jornalista de dados do Público. [Grade fixa] Não porque, lá está, as coisas vão sugerindo de uma forma mais a nível de pedidos ou de ideias que vão surgindo e vamos sempre fazendo uma calendarização em cima do...mas olha, por exemplo, agora eu tenho um trabalho que será pra sair pra semana, e, portanto, tenho aqui o deadline na cabeça, não é? E quando terminar vou analisar e ver se há aqui alguma coisa que vale a pena pegar e depois tenho aqui estoque de ideias e de coisas que eu quero trabalhar, mas desde o Covid-19 que eu não tenho tempo para fazer mais nada. É um bocadinho essa a estrutura. Vou tendo as conversas diárias com o meu editor em que vamos estratificando um bocadinho isto e dizendo: 'olhar, agora vamos por aqui, vamos tentar pegar nisso'.

Production

Eu acho que não há diferença nenhuma, no sentido, e eu esclareço o que eu quero dizer. Como é que normalmente...se eu tivesse que programar um jornalista para trabalhar, qual eram os passos que ele segue. O jornalista tem uma ideia, ou há uma coisa que é preciso trabalhar, ou seja, uma investigação, uma bomba que arrebentou em um sítio, é preciso fazer alguma coisa sobre aquilo. O que o jornalista faz? O jornalista prepara-se para aquilo que vai fazer, estuda sobre o tema, e a seguir passa por um período de recolha de informação, o que ela faz? Entrevistas e não sei o quê. Antes disso ele, em teoria, prepara a entrevista, não é? Aponta já umas ideias que quer abordar, uns temas em que quer tocar. A seguir ele faz a entrevista, volta pra redação, transcreve aquilo que foi dito, matura um bocadinho aquilo que foi dito, escreve um texto, ou uma peça de televisão, ou o que quer seja, e publica. Este é o processo normal de um trabalho de jornalismo tradicional, não é? Para usar aqui a expressão, vá. Como é que é o trabalho de um jornalista de dados, há uma ideia, ou há um fato que vale a pena ser investigado. Preparo-me para investigar isso, leio, vou procurar que tipo de dados que há, que informação é que há, como é que eu tenho que recolher aquilo. Depois há aqui um problema grave, lá está, não há dados abertos, muitas das vezes temos que fazer scraper, whatever, whatever. A seguir, limpa-se aqueles dados e o que eu faço a seguir? Eu sento-me e preparo a minha entrevista, ou seja, preparo questões que quero colocar aos meus dados: qual é o valor mais alto, qual é o valor mais baixo, qual é a tendência, há coisas que se pode ver aqui? O que se pode ver aqui? Agora a grande diferença é que, claro que não estamos a entrevistar pessoas, estamos a entrevistar base de dados, mas os jornalistas antes também usavam documentos oficiais e também faziam o mesmo tipo de abordagem, liam os documentos e perguntavam: tá ok, o que eu posso tentar saber aqui? Até onde que eu posso ir mais pra além disso? É mais

ou menos o mesmo processo, ou seja, eu aponto um conjunto de questões que eu quero colocar aos dados, e aqui há um parentese importante que é o teu conhecimento de estatística, ao nível de programação, influencia as questões que tu podes colocar, não é? Mas isso é a mesma coisa no jornalismo tradicional porque eu se vou...por exemplo, eu não tenho grandes conhecimentos de economia, sei algumas coisas, leio algumas, mas não tenho noção como um jornalista em economia. A entrevista que eu faço ao ministro da economia português vai ser completamente diferente o que eu pergunto do que um jornalista especialista em economia, não é? E aqui é a mesma coisa, ou seja, as perguntas que eu coloco à base de dados dependem dos meus conhecimentos de estatística e por isso que eu faço questão de ler constantemente sobre estatística, sobre novas coisas, ver exemplos: 'Olha esses gajos fizeram isso. Ok. Dá pra fazer isto', não é? Pronto. A seguir eu faço essas questões e chego às minhas conclusões. Normalmente há uma história ali é por isso que eu fui atrás. Havendo uma história, não é, faço escrever, a mesma coisa, desta vez tem que citar os dados e as minhas conclusões. Há aqui uma coisa também importante que é: o trabalho em dados não exclui o chamado jornalismo tradicional, pelo contrário, ou seja, eu tenho ali a base de dados, mas eu tenho na mesma que fazer perguntas a pessoas, eu tenho que ligar a pessoas.

Production

[fontes complementares] Claro, quer dizer, as coisas não...agora dá-te às vezes uma vantagem que é: uma coisa é eu ligar pra um assessor e dizer: 'olha lá, os meus dados dizem que o estado triplicou o custo que tem com médicos'. Uma coisa é isto. Outra coisa é eu questionar o assessor a dizer: 'olha, quanto [qual é o custo]?' Não é? Ou seja, já estás a depender ali de quem quer controlar a informação, não é? Porque é o trabalho e é legítimo e não estou a criticá-lo. Mas estás a depender da resposta daquela pessoa. Por exemplo, a peça de hoje do Público que eu escrevi foi sobre o número, a queda de idas às urgências em Portugal, porque eles deixaram de ir às urgências dentro deste contexto. Eu tinha os dados...eu tive de recolher à unha, quase, porque não dava pra fazer scraper aquilo, recolhi os dados e cheguei às conclusões. Se eu fosse [33'17] até que podia publicar aquilo assim. Agora havia ali uma grande questão que é: esses dados estão corretos? Será que não houve um problema de *report* dos hospitais e eu estou a fazer uma história em cima de nada. E, portanto, eu tive de fazer a seguir a segunda verificação que é ligar para as fontes e dizer: 'olha eu tenho isto. É verdade esses dados? Confirmam-se? Ou seja, o processo continua a ser igual. Aqui que te dá é permite chegar a história completamente diferentes porque não estão, ou seja, não dependes da fonte que te [que regula a informação] exatamente, dependes dela para confirmar, justificar, mas é completamente diferente o processo. Uma coisa é eu dizer: quanto é que é? Outra coisa é eu dizer: Os meus dados dizem-me isto, é verdade? Por que isto está a acontecer? Vocês têm alguma justificação para isto? É completamente diferente do processo, mas é a única coisa.

Data literacy

Tem uma cadeira que participei aliás na semana passada em videoconferência que é de infografia e fotojornalismo, acho eu, em que eles tocam o bocadinho a questão do jornalismo de dados. E há uma cadeira de jornalismo especializado que eu não sei se tocam este tema. Eu não terminei meu mestrado porque eu não fiz a minha tese, comecei a trabalhar e, pronto. É uma coisa que eu tenho atravessada...mas está muito complicado. Agora, sinceramente a maior parte da educação que eu tenho feito não é uma educação formal, é uma educação muito informal, tenho lido muito sobre estatística e faço uma coisa que não é propriamente o melhor processo de aprendizagem mas tem me servido que é à medida que vou precisando, vou aprendendo, ou seja, eu agora quero fazer isto, mas eu não sei fazer isto. Pronto, eu tenho que aprender a fazer isto, tenho que estudar sobre isto, tenho que escrever mais sobre isto. E, depois, faço questão de seguir os blogs ou coisas sobre o data science, não é, para tentar perceber e, ah, há outra coisa muito importante aqui que é: como as grandes organizações jornalísticas tem este hábito muito saudável de publicar o código de suas análises, e de seus processos, eu faço uma educação informal de ler o que eles escreveram, olhar para aquilo e dizer: 'Olha que interessante. Isto dava pra fazer, eu nunca tinha pensado nisto. Ok! Isto pode ser um bom ângulo de análise, não é?'. E, portanto, é um bocadinho assim que eu me vou educando na área. E depois há uma coisa fundamental que é falar com pessoas que já trabalham na área e elas dizem: 'Olha, o Rui, já experimentaste ler sobre isto? O que acha sobre isto?'. Isso acaba por também ser muito útil porque vou aprendendo bastante assim. [Pessoas de referência] Amigos, pessoas que acabei por conhecer precisamente porque me apresentava como jornalista de dados, e fui falando 'sou jornalista de dados, é isto que eu faço'. E, portanto, isso também levou a que um pouco a comunidade de cientistas, pessoas com interesse em dados, acabaram por me seguir no Twitter, e vão me sugerindo, vão falando comigo, ou seja, acaba por ser aqui uma partilha informal de informação, mas que me tem servido muito porque tenho aprendido muito assim.

Data literacy

[o que não poderia faltar numa especialização formal ideal] Há uma coisa importantíssima que é literacia numérica e de estatística. Isso é mesmo muito importante para perceber a diferença de uma moda e uma mediana, perceber o que é uma regressão linear, coisas assim que são importantes porque depois vais precisar, lá está, pra poderes fazer questões precisas desses conhecimentos para os colocar. Sendo que, a experiência diz-me que essas coisas só se aprendem mexendo com dados, e não propriamente estar numa aula e perceber 'ah! Ok, sim, a média é isto a dividir por isto'. Certo, está bem, mas pra que depois isto serve, não é? É um bocadinho isto. Eu, isto pode ser uma opinião polémica, mas eu acho que aprender programação é essencial. Eu costumo usar um bocadinho esta analogia: um fotojornalista tem que conhecer a máquina de fotografar, tem que saber como ela funciona, tem que saber o que é exposição. Um fotojornalista não vai fotografar com uma máquina daquelas descartáveis, até pode fazer um excelente trabalho com isso, não é? Mas idealmente tem que

conhecer o material com o que está a trabalhar. Aqui, o material, o objeto de trabalho é o computador? Por que? Porque toda a análise estatística já não é feita a mão, não é? Já não é feita com uma calculadora. É toda feita com o computador e, portanto, eu tenho que perceber como funciona o computador, como é que funciona a Internet, o que é uma API, para que serve, o que eu preciso fazer com isto, como é que eu posso aprender. Portanto, ter noções básicas de programação. Eu não digo, porque também depois há esta coisa que é ter de aprender uma linguagem, aprender uma linguagem é muito bom. Mas eu não digo ter que aprender esta linguagem em vez daquela, mas ter a noção do que é um algoritmo de como é que se programa algo dá-te aqui uma vantagem essencial, muito importante para poderes, não é? Repara, eu estou atrás do dados de uma coisa. Os dados não estão abertos, o governo não os quer disponibilizar, mas eles estão numa página ali. Eu penso: 'Ok. Eu tenho que retirar esses dados daqui. Então eu tenho que fazer um scraper. Como é que eu faço isto? Como é que eu retiro esses dados daqui?'. Por exemplo, nesta semana num dos trabalhos que estou a trabalhar eu tinha uma base de dados que era atualizada diariamente, de hora a hora, aquela base de dados era atualizada e os dados pra trás eram esquecidos. E, portanto, eu tinha um problema, não é? Porque não tinha ali como descarregar. O que eu fiz? Eu fiz um script que de hora em hora vai lá, tira os dados e guarda. E está aqui há cinco dias a fazer isto. E, portanto, isso ...é preciso perceber como é que o computador funciona, como é que a internet funciona para poderes pensar dessa forma, para poder pensar: 'Ok, meto uma coisa vai ali automaticamente todos os dias e tal e eu consigo ter os dados disso. E, portanto, saber programar pra mim é essencial. Dito isto. Isto não é uma profiss...tu não és data Scientist, tu és jornalista. E, portanto, ter noção do que é jornalismo é mais importante do que tudo isto. Ou seja, isto é importantíssimo, mas ter noção de que estás comprometido com a verdade, de que não podes fazer coisas porque essas coisas podem ter consequências gravíssimas por um erro teu. Isso dá te...é completamente diferente, és jornalista, continuas a ser jornalista, és só um jornalista especializado, que sabe muito sobre uma coisa, mas continua sendo jornalista. E, portanto, eu não acho que um curso de jornalismo de dados sem cadeiras de jornalismo funcione. Pra isso, vais tirar uma licenciatura em engenharia informática, em data science, ou qualquer coisa. E, portanto, acho isso muito importante, acho importante continuares a ter noção do que é jornalismo. E, depois, por último, era aquilo que eu estava a dizer, é ter, é ler o que é feito lá fora, ver o que os grandes estão a fazer porque eles estão no mercado [43'58]...claro que descobri aqui uma frustração que tive durante muito tempo e ainda tenho que é: tu não és o NY Times, não tens a capacidade de produzir o que eles estão a produzir porque eles têm muito mais meios. Mas ambicionar querer fazer aquilo é muito bom porque é isso que vai obrigar a querer fazer mais, querer melhorar. Porque se tu também diz: 'Epa, yeah, tô aqui sozinho e não consigo fazer melhor'. Isto também não é saudável. Mas pra isto acontecer tem que acontecer uma coisa muito básica que é todos os dias ler aquilo, todos os dias olhar pra aquilo, estar atento e ver: 'Epá, esses gajos fizeram aqui uma coisa do caraças. Isto está muito interessante'. Isto é essencial. E, lá está, pra isto

tens que ser jornalista, pra teres esse interesse em consumir informação desta forma.

Production

[colaboração por aversão a programação] Isso acontece, mas também acontece por uma coisa que é. O jornalista de dados também agradece isso [parceria] por duas razões. Uma delas é: há coisas que eu não sei e se tu és especializado em saúde eu prefiro que trabalhes comigo porque tu vais ser capaz de dar o ângulo certo e dizer: 'olha, eu já li, já me disseram que isto era importante olhar pra aqui, vamos olhar pra aqui'. Isso aí ninguém substitui. Tu, se não és especializado numa área, não tens isso. A segunda é que é: esses trabalhos demoram muito e exigem muito de ti mais do que estamos normalmente habituados.

Porque...explicar a uma pessoa o que é limpar dados e que tu tens de estar as vezes há uma semana a limpar uma base de dados. Para um jornalista normal o que isto significa? E, portanto, ter duas pessoas a trabalhar numa coisa é importantíssimo porque te ajuda, se tu estás a limpar os dados...quando acabas uma análise de dados estás todo roto, já não consegue pensar...e ter alguém que está ali disponível para estar contigo a escrever e não sei o que mais, isso é muito importante. E, portanto, sim acabo por assinar muitas peças em conjunto.

Credibility and interest

Eu não posso queixar. Pelo contrário. Eu sinto que, se calhar, por ser novidade, ou uma coisa assim, mas sinto que os leitores gostam muito do meu tipo de conteúdo. Com uma coisa que é, as pessoas que são especializadas em ciências de dados, em estatística, normalmente nunca são muito críticas daquilo que eu faço, ao contrário, as vezes dizem-me: 'olha, o rui, se calhar era melhor teres feito isto, em vez disto'. Isto eu assumo e gosto desse feedback. Dentro do grande leque do que nós chamamos o leitor, ou seja, que não é esse leitor especializado, tendencialmente, os leitores gostam muito disso, mas por uma coisa: porque a maior parte dos trabalhos de dados são interativos e permitem coisas que um trabalho normal não permite, permite personalizar os dados para o leitor, permite ao leitor que explore ele os dados, que brinque não é, de alguma forma, com aquilo. E isso tendencialmente faz com o que esses trabalhos tenham grande visibilidade, sejam muito partilhados nas redes sociais. [Exemplo] Tenho assim dois que são os dois que mais renderam nas redes sociais. Um deles foi na Renascença, foi esse trabalho nomeado para o Data Journalism Awards. Que é basicamente uma calculadora do custo de vida em que eu chego cá, ponho o meu nome, pra onde é que eu quero viver, quanto é que eu ganho. E basicamente, sempre que vou dando mais informação sobre mim ao artigo, o artigo adapta-se a mim. Por exemplo, eu lá em cima digo que moro com zero pessoas e cá embaixo ele diz-me: 'Se calhar, queres um T1 pra viver, não é?'. Eu, depois posso ir lá e dizer: 'Eu quero um T3'. E o texto adapta-se logo ao leitor. E este aqui foi muito partilhado, foi uma coisa surreal. Outro que rendeu mesmo muito já no público foi das curvas do coronavírus. Porque, lá está, era interativo, era fácil de perceber, explicava ali uma série de coisas que as pessoas na altura estavam preocupadas e este também foi muito partilhado. Estes dois casos eu tenho consciência de que

resultaram muito bem. Sobre o que dizias da credibilidade, então, eu acho que os números nos dão uma confiança falsa sobre a sua fiabilidade e o quão perfeitos são. Claro que eu tendo a achar que o leitor valoriza isso demasiado e é por isso que faço mesmo questão de explicar as fragilidades que aqueles números tem, que podem ser incompletos, fazer questão de explicar muito bem a metodologia: foi feito isto, foi excluído isto, ou seja, faço muita questão disso. E era por causa disso que na Renascença publicava todo o código das análises também pra isso. No Público espero que isso vá acontecer, ainda estamos a decidir isso e com o coronavírus também...mas é uma coisa muito importante, lá está, para o leitor mais interessado do que só ler a metodologia, pode ele mesmo correr a análise que eu fiz e ver onde é que eu posso ter feito asneira. Agora, por exemplo hoje eu tive...ou seja, eu tendo a achar isso e acho que as pessoas até valorizam demasiado os números como sendo científicos e não sei o quê, mas hoje nesse trabalho que eu fiz das urgências teve um leitor que nos comentários disse qualquer do coisa tipo: 'Ah! Olhar para os números eu também sei olhar. Agora era ir falar com pessoas e não sei o quê...'. Portanto, sim, quer dizer, também há um bocadinho essa coisa de que as vezes o jornalista de dados foca-se demasiado nos números e pode se esquecer de fazer esse trabalho mais humano e mais ...não é?

Credibility and interest

[audiência de jornalismo de dados é diferenciada?] É porque tens um conjunto de pessoas, ou seja, como é que posso dizer isto, a grande fatia vai ser sempre o público em geral, as pessoas que não vão lá porque é jornalismo de dados, vão lá porque o conteúdo lhes interesse, ou alguém partilhou nas redes sociais, essa vai ser sempre a grande maioria, será sempre. O que notas é que tens um público fiel que tem interesse neste ...que é a malta que tem interesse em estatística, em ciências de dados, em programação, e, portanto, é um público que lê-te não por causa do conteúdo, no sentido não por causa do tema que abordaste, mas lê-te porque sabe que é jornalismo de dados. Isso acaba sendo mais a minha bolha no Twitter, as pessoas que me seguem, não é? Porque já sabem que 'aquele tipo no Público faz esse género de trabalhos e eu tenho interesse nesse género de trabalhos'. E, portanto, é essa audiência que eu digo que é diferenciada porque tem interesse específico naquele género de coisas.

Appendix 4 – Transcription of the interview with Fabio Takahashi

Fábio Takahashi – 03/11/2020

Data Journalism Editor at Folha de S. Paulo

Sociodemographic characteristics

Trinta e nove anos, três nove. Me formei em jornalismo na Metodista, em São Paulo, em jornalismo. Comecei como jornalista esportivo em alguns sites, o mais conhecido foi o Terra. Em 2003, eu vim para a Folha como repórter de educação, daí então Folha Online que antes era separado o impresso do digital, então eu estou no Grupo Folha desde 2003. De 2003 a 2015 eu fui repórter de educação,

depois juntou o impresso com o digital e tal, então [1'56"]. Estive né também envolvido em outros projetos na Folha como o ranking universitário da Folha o book, alguns cadernos especiais de primeira infância e tal, mas primordialmente era repórter de educação de 2003 a 2015. Em 2015 eu fui editor do trainee, o programa de trainee da Folha por duas turmas só porque em 2016 eu fui para Universidade de Columbia em Nova York para fazer uma pesquisa por meio de uma fellowship de jornalismo de educação então fiquei lá de 2016 a 2017. Em 2017 eu voltei no meio do ano, um pouquinho depois do meio do ano, para o recém-criado núcleo de jornalismo de dados da Folha e aí eu sou editor desde então, desde agosto de 2017 até agora. Eu sou o primeiro editor desse grupo de jornalismo de dados da Folha começou como núcleo de inteligência da Folha e o ano passado rebatizou para Delta Folha é o primeiro grupo de jornalismo de exclusivamente para produção de conteúdo da Folha porque antes disso, alguns anos dessa criação desse grupo, teve um outro grupo que criaram mas juntava audiência com jornalismo de dados e produziu pouco dado e ficou muito em cima da audiência. Então, desde 2017 que eu tenho esse grupo específico de jornalismo de dados na Folha e eu trabalho desde então. Calhou...algumas coisas convergiram. Primeiro como eu fui tempo repórter de educação eu peguei bem uma fase que o uso de dados da educação explodiu. Então mesmo que eu não programo hoje, não programava até então, mas mexer com os dados era uma coisa que acabou virando parte do meu dia a dia na educação. De forma muito acessória a isso foi quando fui acessória a isso quando eu fui pra fazer a fellowship de jornalismo de educação na Columbia e lá fiz uma aula só uma disciplina de análise de dados que me ajudou também um pouco, mais na teoria. Juntando isso, quando eu estava voltando dessaeu tinha uma licença não remunerada da Folha, ou seja, eu tinha o meu emprego aqui ainda, mas a Folha tem como tradição, depois de um tempo fora eles tendem a oferecer uma outra coisa, para você fazer uma coisa diferente e calhou bem na hora que eles estava criando esse grupo de jornalismo de dados então eu estava no fim do meu período em Nova York e eles já me convidaram para esse grupo, também está dentro dessa convergência de fatores que eu já tinha também experiência com dados pelo período da educação e muito acessoriamente tinha feito esse curso na Columbia.

Production

Eu sou mais o gestor, eu sou o editor. Então, a gente tem muitos profissionais que na formação deles nem são jornalistas na verdade. E aí cada um tem um pouco essa estratégia ... a gente não tem uma coisa muito uniforme. Cada um tem seus métodos, ferramentas...alguns usam mais o R e outros mais Python ...eles são bem autônomos na verdade.

O [9'12"] de projetos pode começar de duas formas: ou a gente tem uma ideia dentro do grupo já dos dados que é o Delta Folha aí a gente apresenta ou pra direção do jornal, para secretaria da redação da Folha ou para as editorias, então esse é um caminho, nasce da gente. Um outro caminho também é que nasce das editorias ou da direção do jornal, eles acham que a gente pode entrar, a gente conversa e vê se a gente pode entrar mesmo, se faz sentido. A gente não tem

separado, assim, jornalista de dados com cientista de dados, com analistas de dados, com estatístico é... as pessoas são tudo ao mesmo tempo, mas com diferentes características. Eu tenho algumas pessoas que não são jornalistas, sei lá, eu tenho dois da Biologia, são biólogos, então trabalhavam com genética e tal. Então eles sabem, eles conhecem uma boa capacidade de programação, uma boa análise estatística também, mas eles não entregam o produto final, não escrevem o texto. Com o tempo aqui, eles já vão pegando o tino jornalístico, eles sabem que há história, o que rende mais história jornalística ou não. Mas ainda assim tem uma limitação, não vão chegar e entregar tudo. Não vão, sei lá, fazer uma raspagem, fazer uma limpeza, analisar os dados, escrever o texto. Esses profissionais não fazem a cadeia toda. Aí eu tenho um jornalista, que é jornalista, que programa um pouco, mas, então, é ao contrário são pessoas que podem entregar um produto final, mas são coisas um pouco mais simples, eles não são tão avançados na programação. Então eu tenho esse mix de pessoas na minha equipe e além disso eu tenho uma infografista que cuida da visualização dos nossos projetos. Informalmente, assim, não é uma camisa de força, assim, mas a gente tenta ter duas publicações por semana, mas, assim, não é uma camisa de força isso. Tem semana que a gente vai fazer coisas menores para publicar mais e às vezes fica meses num conteúdo só daí a gente vai publicar menos. Acho que, diferente de outros grupos de jornalismo de dados que o grupo acumula também essa função de fazer pedidos rotineiros de LAI, a gente não costuma usar muito não é, a gente vai mais no dado que já está público. Eventualmente, assim, o que acontece é: às vezes um repórter faz uma LAI e vem um artigo gigantesco e a gente ajuda a fazer, mas não nasce da gente. Dificilmente um projeto que nasce no meu grupo que seja pensado esperando uma resposta da Lei de Acesso de Informação.

Eu acho que o quê coloca no eixo aqui pra gente é ter uma hipótese bem clara e essa hipótese é jornalística, sai uma pergunta que um repórter deveria fazer, que um editor poderia fazer só que em vez de entrevistar uma fonte a gente vai atrás de dados para responder essa pergunta. Então, acho que a gente coloca no prumo jornalístico acho que escolhendo bem essa pergunta, essa pergunta é uma pergunta jornalística, que o repórter poderia estar fazendo para uma fonte oficial, por exemplo.

Credibility and interest

A gente não comete crime, assim, a gente não é hacker, a gente vai em dados públicos, isso é uma coisa. A gente avalia também o quanto está expondo o indivíduo. Se a gente consegue dados granulares a ponto de identificar pessoas a gente tem uma discussão se aquilo faz sentido. E se faz sentido a gente segue toda a cartilha jornalística, se a pessoa for citada, ela vai ter de ser ouvida para dar o outro lado.

Production

Os nossos conteúdos...a gente trabalha para que eles sejam o menos diferenciados possível do resto do jornal. Mesmo que uma matéria seja, uma reportagem do nosso grupo que envolveu um trilhão de linhas de análise e 200 mil colunas, no fim a gente quer que essa reportagem siga os outros padrões da

Folha. Se tiver personagem...às vezes não é nosso grupo que faz isso, né? Então a gente aproveita...muita gente trabalha em parceria com os repórteres das editorias, mas em última instância o produto que a gente faz, que está no nosso selo lá [18']...a gente fez uma boa reportagem da Folha, então a gente não fica muito ...a visualização nossa é importante, é uma parte importante do nosso trabalho, de como expor nossos trabalhos, mas é tão importante quanto um texto que tenha bons personagens, boas entrevistas. Acho que isso eu sinto um pouco diferença no nosso trabalho para outros grupos de jornalismo de dados que são muito calcados na visualização, de grandes interativos e tal. A gente que fazer ótimas reportagens da Folha de S. Paulo, só que no nosso caso, a gente está partindo de grandes bases de dados.

Big Data - Volume

São esses dois [Python a linguagem R]. Principalmente a gente trabalha em R e Python, geralmente em R. [Análise estatística] eles fazem tudo no R mesmo.

Production

Nossos matérias são uma reportagem da Folha de S. Paulo como qualquer outra. Se tiver uma acusação uma acusação contra o ministério da saúde a gente vai ouvir, provavelmente a gente vai ouvir especialistas de saúde, achar personagens.

É raríssimo conteúdo que a gente gaste menos de três dias, muito raro. A gente já teve projeto que a gente ficou três meses. Alguns projetos, a gente ficou três meses fazendo cada um. O último que a gente fez grande, os últimos dois grandes. Um foi muito grande, o GPS Ideológico é uma coisa que a gente faz, uma análise do posicionamento ideológico das pessoas no Twitter, se a pessoa é da esquerda, direita e tal, tanto parte de dados, quanto a parte da redação. No ano passado a gente também fez um, não sei a gente a ficou três meses, mas a gente ficou algumas semanas, talvez dois meses, fazendo um especial de música, chamou de Música sem Fronteira, ficou comparando o que é sucesso no Brasil com os outros países. E outro que a gente demorou três meses também, mais ou menos, foi um estudo estatístico em que a gente levantou quais provas possivelmente foram fraudadas do Enem, foi em 2018, a gente ficou três meses fazendo.

Não vejo nenhuma área que ...acho que depende do tema, da estrutura do dado. Não sei, não tem uma área...por exemplo, política pode parecer difícil, mas tem muito dado razoavelmente bom, assim, então o TSE por exemplo é...em geral os dados do TSE são bons, assim. Uma área que eu poderia responder política não vai ser, assim. Por exemplo o TSE ajuda bastante nessa questão eleitoral e tal, também votações do Congresso, muito dado interessante para trabalhar.

Eu tenho muita confiança pela minha equipe, mas mesmo os melhores, é uma equipe que eu já acho boa, pelo menos mais duas vezes veem cada o trabalho das pessoas. Em geral a gente discute...se a pessoa tem uma ideia e está tocando um projeto, eu, e em geral mais uma pessoa, vai dar uma olhada no resultado. E a gente aqui já salvou incontáveis erros, mesmo de pessoas muito boas, sabe? Muita capacidade de programação, bom conhecimento estatístico,

mas é, assim, é um projeto tão grande, é tanto dado que a pessoa já não está mais conseguindo processar aquilo no próprio cérebro. Então, assim, a gente faz essas checagens, digamos externas, pessoa de fora da... não só a pessoa fora [...] ela nunca vai ter uma autonomia completa, pelo menos eu vou ver exatamente o que está acontecendo e em geral mais uma pessoa da equipe também vai ver. Fora que nossos conteúdos são publicados nas editorias, então são publicados na editoria de política, de economia e tal. Então ainda tem mais um filtro. Como eu te falei, a gente faz projeto de tudo, né? De música, de política, de esportes e tal. É bom e é ruim, né? [...] Exatamente vai conhecendo aquela base de dados, a gente não tem esse domínio de tanta coisa. Mas acho que o próprio processo dentro do jornal consegue ter barreiras para evitar grandes erros. Para gente faz super sentido um dado, bate no editor de cotidiano, por exemplo, o cara: 'Pô, mas espera aí, né? Isso não foi em 2014?'. Então acho que são vários filtros que vão mitigando eventuais erros.

Production

Eu acho que sem o tino do jornalístico, não adianta ser um super programador. Até quando eu estava montando a equipe aqui é... eu estava procurando [28'22"] que tipo de profissional que eu precisaria ter. Eu achei [28' 25"]. Exatamente, tem gente da ciência política, ciências sociais, tem o pessoal da biologia também que eu te falei. Digamos, eles são profissionais, usam programação e eles conseguem enxergar um problema, sabe? Pra responder uma pergunta. Em geral esse pessoal está na academia, acho que isso é uma boa convergência. Por que eu estou dizendo isso? Porque eu também fui, por exemplo, nas Ciências da Computação, Sistema de Informação. Eu também fui nessas para ver profissionais quando eu estava procurando aqui gente para minha equipe. E acho que não funciona por duas coisas, a primeira é que eles ganham 20 vezes mais do que a gente ganha, e a outra é de fato que eles programam demais, muito mais do que o pessoal que trabalha com a gente aqui, e eles não tem a pergunta, sabe? Não interessa pra gente. Eles respondem outras coisas. Eles vão desenvolver um software para melhorar as transações bancárias. Isso é legal. Alguém vai ficar rico com isso, mas... Então o profissional... o que vai funcionar pra gente aqui [29'50"] pra gente aqui é está no meio do caminho, ter uma grande capacidade de programação, mas ter muito claro qual que é a pergunta e a resposta a que vai chegar, né? E ainda assim, né? Profissionais tem que ... alguns com mais facilidade, outros menos facilidade, se eles não são jornalistas, vai demorar pra pegar o que é uma notícia. Porque uma coisa é se você faz um software que vai melhorar 20% das transações bancárias se sabe o que é. Agora o que é uma notícia? O que vira uma manchete da Folha ou o que nem interessa para a Folha? Essa é uma coisa também um pouco de treino. Agora até a gente pretende um deslocamento dentro da própria da redação, vai vir uma jornalista que estava em outra editoria pra minha equipe. Ela participou de uma iniciativa bacana aqui na Folha. A gente teve um trainee só de jornalismo de dados, foi uma parceria com o Google e aí foram quatro meses intensivos de programação para profissionais da casa e a gente abriu algumas vagas também para público externo. Então os profissionais da casa ficaram fora da rotina de

trabalho deles para aprender programação, não só aprender porque alguns já sabiam e precisavam se aprimorar. Mas além disso, em geral eles não param de fazer cursos e tal, cursos livres, books.

Credibility and interest

No limite, como te falei, a gente espera que esse público do Delta Folha não se diferencie do público da Folha, mas eu acho que ... há uma parte do nosso público que de fato é um pouco mais específico que o público da Folha, não tem jeito, ainda que eu tenha dito ao longo dessa conversa que a gente busca sempre aproximar o máximo possível os nossos materiais aos da Folha. Mas é diferente o material nosso, do material de um repórter que descreve um bombeiro que salvo [...] é diferente. O nosso é um pouco menos humano digamos a forma de contar história e um pouco mais técnico. Eu acho que aquele público que busca um pouco mais de evidencia das notícias, sabe? Fatos concretos, menos emoção. Ainda que acho que a gente consegue navegar por todos os públicos aí, mas acima de tudo um publico assim mais...pelas respostas, sabe? Quem comentar no Twitter, sei lá, professor de economia da USP, sabe? Parece uma coisa mais...uma pegada mais acadêmica.

O que mais tem repercussão é no Twitter que eu acho que a coisa pega mais pra nós aqui, assim, tem mais tração. A Folha não atualiza o Facebook há alguns anos, a Folha inteira, então já é uma plataforma que está descartada pra gente. A Folha tem um canal no Instagram, que é bem... perfil, né? Perfil no Instagram que é bem potente. Mas a gente acha que os nossos materiais não têm muito nossa cara lá no Instagram. Então no Twitter a gente tem uma conta do Delta Folha em que a gente publica nossas coisas, responde algumas perguntas, alguns comentários e tal. Ficou surpreso muitas vezes pela qualidade dos debates, não que seja fantástico. Mas tem momentos que acho são interessantes, assim. A gente até pesca umas coisas ali para usar numa outra matéria, pra complementar até a nossa. Claro, se for quantificar a maioria é...são os ataques. Se é uma matéria contra o Lula, a gente é Bolsonaro, por aí. Mas acho que no meio disso tem coisas interessantes assim. Tanto no Twitter e obviamente mais nos comentários das próprias matérias que daí é só para assinante então já tem um filtro. Mas tem coisas interessantes. Vejo o pessoal discutindo coisas bacanas. [Exemplo] Posso te mandar, assim, não lembro de cabeça. Quase toda matéria nossa tem um dois comentários que você fala: 'Olha! Legal. É um bom comentário'. Eu vou procurar alguma coisa e te mando depois.

[Dados usados por instituições] Sim, várias vezes. E acho que o que foi mais usado nosso foi esse GPS ideológico, a gente foi até chamado pra falar em grupos de Ciência Política da USP. A gente como painalista ali no meio de um monte de cientista político. Se eles não gostaram, foram educados. Mas foi bem legal, parece que a gente trouxe mesmo interessante até para academia mesmo. Outros projetos que a gente fez também. Esse de música também vejo que tem uma boa discussão. Recebi algumas vezes pedido de entrevista ou pra falar mais sobre os dados para dissertações, pra teses e tal. A gente tem uma boa entrada nesse meio acadêmico.

Appendix 5 – Transcription of the interview with Sergio Spagnuolo

Sergio Spagnuolo – 03/04/2020

Founder of Volt Data Lab and Núcleo.jor.br

Sociodemographic characteristics

Eu tenho 35 anos. Formei em jornalismo pelo Mackenzie em 2007, tenho mestrado pela PUC em ciências sociais, 2010 a 2012. Motivacional...eu nem sei dizer o que me motiva. Acho que eu gosto da profissão, eu sou um entusiasta do jornalismo. Ao contrário do que muita gente acha, pelo menos na minha experiência, assim, tipo, não tô rico, mas eu sei que há salários bem condizentes com...o jornalista acha que ganha mal, mas se você compara com outras profissões o jornalista ganha até que bem, assim, sabe? Obviamente tem quem ganha muito mal e tem quem ganha muito dinheiro. Mas a média não sei se é tão ruim quanto muito jornalista gosta de reclamar. A motivação mesmo é, enfim, querer produzir bom conteúdo, fazer coisa de relevância para as pessoas que vai orientar decisões diversas e tal, eu sou muito motivado por tentar produzir alguma coisa para ser útil.

Todo jornalista, num nível ou no outro, trabalha muito com dado. O que acontece é que a trajetória do jornalista, a trajetória de trabalhar com dado do jornalista foi trabalhar já com dado mastigado, aquele dado que já vem pronto, do IBGE, da ONU, de empresas, ONGs e tal. Eu comecei a focar, trabalhar exclusivamente com dados, não exclusivamente, né? Mas orientado a trabalho principalmente por análise de dados e tal, quando eu percebi que tem muito dado que pode ser explorado do começo, né? Eu pegar dado bruto, não mastigados, microdados que eles chamam. Isso dá uma perspectiva nova e a gente consegue fazer coisas novas que antes o jornalista não conseguia. O jornalista dependia da fonte, dos outros analisarem isso aí. E com tecnologias novas, ferramentas open source principalmente, foi possível para jornalistas trabalharem com dados num nível muito mais sofisticados que faziam antes. Então eu vi essa janela se abrindo no Brasil em 2014 e montei o Volt pra ajudar redações e ONGs a produzirem conteúdos com viés jornalístico a partir de dados.

Production

Por enquanto tem, enfim, né? Por enquanto o jornalismo de dados está aí para ajudar o jornalismo tradicional, digamos assim, a lidar com esses dados. O jornalista ainda tem uma cabeça muito voltada para o jornalismo declaratório, o jornalismo baseada em fontes não quantitativas, né? E aí está começando a ter um reconhecimento de valor que o jornalismo de dados fornece pro noticiário em geral, né? Assim, acho que precisa melhorar muito ainda. Tem uma coisa que, enfim, que é inerente ao jornalismo e eu acho que precisa mudar ...enfim. Estávamos tendo esse debate, né? O que é jornalismo? O que não é? Porque hoje em dia você pode fazer um aplicativo e chamar de jornalismo, entendeu? O jornalismo não precisa ter aquela fórmula quadrada, é, de lide, gráfico, vídeo, declaração...você pode ter vários formatos de jornalismo. Então, o Núcleo é um

site que a gente lançou esses dias aí, em janeiro. A gente não tem uma necessidade urgente de chegar a uma conclusão com uma matéria, entendeu? Isso causa desconforto em muita gente, inclusive, quantas pessoas que fazem trabalhos pra mim e tal. O Alexandre que é um dos nossos editores, até hoje ele fala: 'Não, legal! Vamos fazer diferente, vamos tentar ver até onde vai'. Mas até hoje eu percebo que ele é resistente porque falta aquela conclusão. E daí? O que eu tiro com isso. Mas a resposta que a gente tem para as coisas que a gente faz é incrível, entendeu? Eu tenho gente do governo me mandando mensagem, falando que vai utilizar em CPI, ou falando que o trabalho que a gente tá fazendo é muito legal. Basicamente...uma matéria que a gente publicou ontem foi 'isso é o que sabemos sobre o enxugamento do bolsa família do Brasil'. É isso. Analisamos dados e constatamos o que o dado está falando para gente. A gente não tá fazendo afirmações muito forte. O ideal é que a gente tenha uma conclusão legal, ou uma afirmação forte pra fazer. Isso a gente só faz quando a gente tem certeza. Então, assim, o bolsa família é um programa muito...tem muitas nuances, é um programa muito amplo, tem muitas famílias. Tem gente que fraudava, tem auditoria, tem gente que precisa pra caramba, tem fila. Então, em vez de a gente falar assim: 'Bolsa família foi reduzido no governo Bolsonaro em tantos por cento, a gente fala assim: 'Olha, isso é o que a gente sabe. Os dados até novembro mostram uma queda de 4% no ano' ... Então, assim, a gente não precisa cravar com certeza para depois chegar alguém e falar: 'porra, mas esse dado tem uma nuance melhor'...ficar uma guerra de versões. Eu acho muito bacana quando o jornalismo tira conclusões e acho que o objetivo principal tem que ser você chegar realmente numa conclusão e ter alguma coisa para falar. Mas você também pode apresentar informações de uma maneira que ... não é todo mundo que consegue pegar dados do bolsa família, criar um [10'40"] pegar a API e fazer...então não é todo mundo que consegue. Então a gente vai lá e faz isso pelos outros e os outros vão se informar a partir disso, entendeu? Pode servir para pautar o jornalismo, pode servir para o usuário final. O nosso objetivo no núcleo é mais pautar ONGs, jornalistas, terceiro setor do que realmente chegar na ponta. O nosso escopo são advogados, gestores de terceiro setor, coordenadores de comunicação. A gente meio que escreve para esse pessoal. É válido. O retorno que a gente tem das pessoas é muito grande, muito positivo. Lógico que tem gente que fala assim: 'isso que você faz não é jornalismo'. Como o cara falou ontem no grupo: isso que você faz não é jornalismo de *dashboard*'. Tá bom. Beleza. Pode pensar que não é jornalismo, mas tem um monte de gente que tá usando isso como jornalismo. E o meu propósito é fazer jornalismo com isso.

Production

Por exemplo, no Núcleo, que é o veículo do Volt aí, a gente teve essa matéria do Ibama, que uma representante da consultoria legislativa da Câmara dos Deputados mandou uma mensagem pra mim falando que iria usar esse material na CPI do óleo do nordeste. Pedi, queria saber se eu podia mandar os PDFs. Tava tudo clicado na matéria, né? Ela só não clicou no link, mas ... eu mandei pra ela o link que tava os documentos e tal. Imagino que vá usar ou não. Não sei o que aconteceu. Por exemplo, a gente construiu a API do Fogo Cruzado que é um

aplicativo que mapeia o tiroteio no Rio de Janeiro. E a gente fez a API pra eles fez o pacote em R para os pesquisadores poderem utilizar e tal. Aí você fala assim: 'Putz! Sérgio. Isso não é jornalismo. Você desenvolveu um sistema um software'. Mas esse software que a gente desenvolveu está sendo pela secretaria de segurança do governo do Rio de Janeiro, Entendeu? Está sendo usado por pesquisadores, está usando pela imprensa. Assim, o propósito com o qual a gente fez é jornalístico. Olha só o impacto que isso tem. Tem o Atlas da notícia que é usado para pesquisa, é usado para reportagem é usado por assessoria de imprensa para achar as cidades do Brasil, para tentar anunciar... Tem vários exemplos. Tem uma matéria do Daniel Mariane da Folha que o cara achou a possibilidade de fraudes em provas do ENEM. Isso levou a uma investigação do MP. Tem a matéria do G1, projeto grande do G1, sobre o mapa da violência que é um esforço coletivo gigantesco no G1 para mapear os assassinatos no Brasil. E os dados são muito parecidos com o atlas da violência que estuda boletim de ocorrência. Tem bastante exemplo de coisa que a gente coisa tirar, que o jornalismo consegue ter um impacto.

Credibility and interest

Eu nunca vi pesquisa sobre para ser sincero. O que eu acho é que credibilidade está associada diretamente a transparência, né? Quando você divulga os seus dados, você diz de onde você tirou, você cita a fonte, quando você usa uma fonte em off e você explica porque ela está em off, e dá o máximo de informação que você puder sobre essa fonte. E que é difícil na prática porque enfim as fontes não querem se identificar. O legal de trabalhar com jornalismo de dados é que não existe, ou pelo menos a maior parte, 95% das vezes, não existe fonte off, as fontes são abertas. Então eu acho que o jornalismo de dados um potencial de credibilidade gigantesco. O que acontece é que muitos veículos não tem a cultura ou o interesse de documentar o trabalho que eles fazem e mostrar essa documentação. Se você entrar no nucleo.jor.br, você vai ver que todas as nossas matérias têm link para os dados e no final da matéria tem a metodologia que a gente usou, o link para a fonte original, tem tudo que você imaginar de transparência tem aí. Tem até o negócio se o jornalista for útil, conte para gente, se tiver sugestões, correções. Então assim, a gente se torna acessível pro leitor, a gente mostra as informações. Mesmo quem nunca ouviu falar da gente vai pegar e vai falar assim: 'Putz. Os dados estão aqui, os códigos estão aqui, sabe? A informação está aqui, o cara está aberto a ouvir o meu feedback, se eu quiser dar pra ele. Então é uma coisa que falta no jornalismo. Por exemplo, no lide da minha matéria do bolsa família, eu cito El País, Folha de S. Paulo e o Estado de S. Paulo, um pouco mais pra baixo eu coloco link para todo mundo, para todas as matérias. Está tudo documentando. Então, assim, o retorno que isso dá pra gente em termos de transparência, em termos de credibilidade, é muito grande, entendeu? Tem gente do governo que manda mensagem, tem um cara no nosso grupo de dados aqui que trabalha no Tesouro Nacional que fala: 'Cara, o trabalho que vocês fazem é excelente'. E, assim, a gente tem muito que melhorar em termos de conteúdo. Temos que fazer mais e melhor, sabe? Tem que sempre ir melhorando. Mas a transparência consegue a gente uma lisura para aquilo que a

gente tá fazendo, entendeu? Eu acredito muito fortemente, que quanto mais você for transparente com o jeito que você faz reportagem...a gente fala aqui: 'Por que o assunto é importante? E se é importante, porquê, né?'. Assim, a gente tá tentando acrescentar uma linguagem nova para o que a gente tá fazendo no jornalismo. Não que os outros veículos têm que copiar a gente ou não. Eu acharia ótimo se eles copiassem. Mas, assim, eles têm que achar fórmulas de serem transparentes com o leitor deles. NY Times faz isso, o Guardian faz isso, o Washington Post faz isso, o Marcamp (semana passada), né? Que essa iniciativa nova que é mais ou menos o Núcleo dos estados unidos, eu digo o Núcleo dos Estados Unidos porque nós lançamos antes. Por que o jornal? 'Ah! Mas gasta tempo fazendo, mas é muito necessário. A Folha é extremamente atacada por não revelar nada sobre o jeito que ela faz o jornalismo dela, entendeu? Tem um puta mérito. Eu confio na Folha. Quando eu leio na Folha eu sei que eu estou lendo um negócio que foi apurado, entendeu? Mas as pessoas não sabem disso. A gente conhece os jornalistas lá. Então eu confio no que está sendo publicado ali, mas as pessoas não sabem, entendeu? Então. É, enfim. Seria muito legal ter isso em outros veículos.

Production

Olha, o Volt nunca foi um veículo jornalístico, né? A gente é um prestador de serviço e consultoria para ONGs e redações. Então, assim, a gente já fez conteúdo, já desenvolvemos pautas, mas não era espontânea, é muito reativo e quando a gente tinha alguma coisa a gente chegava e vendia em parceria e tal. Nunca foi que nem o Núcleo que tem uma produção própria de conteúdo, né? Então é bom deixar isso claro. E hoje com o Núcleo o Volt é basicamente uma consultoria. Eu faço consultoria e cursos para redações. Dito isso, no Núcleo, aqui no site a gente fala, a gente cobre alguns setores principais que é transparência de recursos públicos, transparência de políticas públicas, mídia, tecnologia e segurança da informação, meio-ambiente e cidades inteligentes, políticas de direitos humanos. Então a gente meio que escolhe aí dentro. Não tem um processo muito. Como a gente está bem no começo também, com pouco recurso e tal, então a gente vem com algumas pautas que a gente já estava vendo antes, quando o Vortex existia ainda...enfim, o Vortex é um outro projeto que a gente participou que não deu muito certo. Enfim, a gente o que está acontecendo no Brasil agora e vai atrás de certos assuntos. Não tem muito um brainstorming pra essa coisa de pauta. Basicamente eu e o outro editor, o Alexandre Eurico, a gente fala 'Pô! Essa pauta vale a pena fazer então vamos atrás dela. Às vezes a gente comissiona alguém, paga alguém para fazer. Às vezes eu mesmo faço. Se o negócio andar mesmo e tal, suponho que vai ter um processo um pouco mais organizado de definição de pautas e tal. Por enquanto está muito tentativa, muito experimental.

[Complementary sources of information] Depende do formato, se você entrar em qualquer matéria do Núcleo, você vai ver, em baixo do título, ele tem ali essa é uma descrição, por exemplo, e explica mais. Por exemplo, numa análise ou numa

investigação, provavelmente a gente vai consultar outra pessoa, uma explicação. Numa descrição a gente não vai consultar, ninguém, a gente vai descrever os dados como eles estão ali. Então depende do formato que a gente está fazendo, né? Se você ver nessa matéria do Ibama por exemplo, a gente entrevistou uma especialista. O que a gente não quer fazer é fazer aquele jornalismo: “Ah! Eu vou buscar o contraditório. Igual, por exemplo, a CNN faz. A CNN tem os talking heads. A CNN literalmente paga pessoas para discordarem uma das outras. Então você o talking head do Partido Democrata e o talking head do Partido Republicano então eles nunca vão concordar. Eles são pagos para defender um lado, para discordar. O que eu não quero é isso: guerra de versões. Também não é que eu quero uma pessoa para corroborar o meu dado, eu quero uma pessoa isenta que de alguma forma: ‘Olha, isso aqui é assim, assim e assado’. Tem um monte de gente, por exemplo, que é mais de esquerda e tal, e se você perguntar: ‘Ah, mas e o enxugamento do bolsa família?’ Ela vai falar assim: ‘Não, mas houve no governo Dilma, houve no governo Temer’. Não é uma exclusividade do governo Bolsonaro. Então, assim, se for para consultar especialista, tem que ser um especialista que não vai ser enviesado. Não vou falar com alguém, por exemplo, que eu sei que vai defender um lado, não interessa o motivo. Então tem que ser alguém muito mais para o lado da ciência que para o lado da política.

Production

Alice, é assim. A gente não faz nada de um dia para o outro. No jornalismo de dados não existe uma matéria de um negócio que precisa para amanhã. Isso não vai acontecer. A não ser que você queira um negócio bem porco ou bem raso. Então, por exemplo, esse caso do bolsa família, eu sentei nesses dados por um tempo. Rodei as mesmas análises várias vezes, busquei as fontes. Tentei entender o que era. E, assim, o próprio Jonathan Stray, ele fala no livro dele lá no Curious Guide for Journalists, que o jornalismo ele basicamente conta coisas. A gente está contando coisas e tirando proporções delas. Nesse caso aqui, que eu fiz, no Bolsa Família, por exemplo, vou continuar no exemplo. Eu peguei todos os dados do Bolsa Família que eu consegui por cidade e somei, fiz uma média anula, por exemplo. Não tem nenhum segredo, nenhuma regressão, não tem nenhuma fórmula mágica. E obviamente, à medida que você vai entendendo melhor os dados, você pode ir ampliando o tipo de análise estatística, de sofisticação de análise que você está fazendo. Nesse caso aqui eu basicamente agrupei e somei. Então, só pra entender o que está acontecendo nesse contexto. Então, enfim, é meio que isso assim. A gente não, eu não estou fingindo que eu sou um especialista em bolsa família e tô fazendo as melhores análises do mundo. Eu estou falando assim: ‘Olha, é isso que esses dados têm. No nível mais básico é o que eu consigo te mostrar agora’. E à medida que a gente vai entendendo esses dados melhor, a gente pode ir aumentando a sofisticação, por exemplo, essa outra matéria que eu fiz aqui dos desertos de notícia, à medida que a gente... esse é um tema que já entendo muito melhor, consigo fazer cruzamentos muito mais sofisticados do que a análise de dados do Bolsa Família, entendeu? Então, você vê que essa análise está bem diferente. Consigo fazer análises mais avançadas. Então é meio que isso, sabe? A capacidade que a gente tem de fazer

aquela coisa com o que gente tem na mão. Correção: eu quero fazer matérias com conclusão. Mas às vezes a gente...eu não quero ser obrigado a ter uma conclusão 100% das vezes, né?

Data literacy

O que é mais difícil, em geral, a parte técnica é a mais fácil, uma vez que você domina ela. Não que eu seja o maior programador do mundo, o maior especialista e tal. Mas sempre vai ter alguém que vai conseguir tirar aquele dado, formalizar, vai dar trabalho, enfim, a parte técnica ela é contornável na maioria das vezes. Obviamente que tem vezes que é impossível...estava tentando tirar dados do Banco Nacional de mandado de busca, né? E os caras colocaram o captcha lá, ferramenta para limitar o número de busca, que é quase impossível superar aquilo lá. Então é...enfim. Mas a parte técnica em geral é mais fácil. A parte mais difícil de todas é você tirar sentido dos dados, entendeu? Essa é a parte mais foda de todas. Porque, enfim, as vezes você não entendeu aquele dado, publicou alguma coisa errada, acontece com tudo mundo, eu acho. Você esquecer alguma coisa, por exemplo, assim, eu li bastante sobre o Bolsa Família para fazer essa análise, que é uma análise simples, tá? É, no começo, se eu tivesse publicado na primeira semana de ter analisado esses dados, eu teria cometido um monte de erro, teria falado um monte de bobagem, entendeu? Vai saber se eu não falei agora, inclusive, né? Eu acredito que não porque eu estava seguro do que eu estava fazendo. Mas, porra, vai saber. Então, assim, é, você fazer o negócio certo, fazer a análise certa e focar no dado e entender esse dado é a parte mais difícil de todas. Dá trabalho, leva tempo, é frustrante. É repetitivo. A parte técnica que todo mundo fala, vamos aprender a programar e tal isso aí é muito importante é legal. Recomendo que todo mundo aprenda a programar. É útil para caramba, maravilhoso, abre muitas portas, mas uma vez que você domina isso não é a parte mais difícil. Até se você for intermediário não vai ser a parte mais difícil. O mais difícil é você tirar sentido de todo o barulho que tem nos dados. [Tempo de trabalho do Bolsa Família] Um mês e meio mais ou menos. Assim, não é que eu fico também todo dia, mas, sei lá, duas, três vezes por semana eu pegava e passava várias horas analisando. Analisava, via se tinha atualizado os dados, refazia, li a respeito e fui construindo a narrativa em cima disso aí. Se eu tivesse pegado para fazer pra valer, quem sabe em duas semanas eu tinha matado essa matéria. Mas fui fazendo outras coisas no meio.

Credibility and interest

Limite ético para mim é mentir, né? Eu não mentiria. Assim se o dado for um dado privado, pessoal, não público. Aí eu não usaria também, né? Essa matéria dos twits que a gente fez a gente divulgou os dados de forma anonimizada, né? A matéria dos políticos, por exemplo, com dívida ativa está no nosso banco de dados, mas não está na matéria. A gente não nomeou nenhum político porque não é proibido você dever pra União, né? Se você não pagar e tal pode ser um 34'09" fiscal. Mas, oficialmente você pode, você tem o direito de dever, se você é um devedor, você vai ter que pagar depois. Mas não é proibido, não é um crime.

Dever por si só, não é um crime. Então a gente acabou não nomeando ninguém. Então, assim, a privacidade é uma coisa que a gente leva muito em conta, e tem que ser levada, o limite ético. Obviamente ser honesto, não mentir, ou se você errar corrigir. Enfim, é caso a caso. É difícil avaliar muitas das coisas antes delas acontecerem, principalmente num cenário que muda tanto.

Appendix 6 – Transcription of the interview with Dr Suzana Barbosa

Suzana Barbosa – 05/04/2020

Director of the bachelor's degree course in Journalism at the Federal University of Bahia

Sociodemographic characteristics

Suzana Oliveira Barbosa. Eu sou jornalista de formação, minha graduação em jornalismo é claro no momento que não falava disso nem havia ainda uso expandido de informática e o jornalismo não era feito desse modo, nós não aprendemos com tecnologias digitais. Eu sou mestre e doutora pelo programa de pós-graduação em comunicação e culturas contemporâneas na faculdade de comunicação aqui da Universidade Federal da Bahia. Fiz o estágio de doutoramento, como se fala em Portugal, aí na Universidade da Beira Interior com o professor António Fidalgo, que hoje é reitor da UBI isso em 2005, meu doutorado em 2007, e logo em seguida fiz um pós doutorado na Universidade de Santiago de Compostela, com o professor José Lopes Garcia, que é o diretor do grupo novos medios. Bom, isso era parte de um convênio entre Brasil e Espanha para estudo do ciberjornalismo e eu fui então uma das que primeiro foi, na verdade, para esse pós-doutorado dentro desse convênio, no âmbito deste convênio. Tenho 51 anos, sou professora associada 1 no departamento de comunicação da Faculdade de Comunicação e sou professora permanente no programa de pós-graduação em comunicação e culturas contemporâneas.

Sociodemographic characteristics

O estudo de jornalismo em base de dados vem desde o doutorado. Então foi ali o resultado obviamente o produto da tese foi sobre esse tema que na verdade o título da tese é Jornalismo Digital em Base de Dados – Um paradigma para produtos digitais jornalísticos digitais dinâmicos, e claro que passado esse tempo é o jornalismo em base de dados, o jornalismo de dados ganhou uma outra como eu posso dizer, saiu de algo que poucas pessoas compreendiam na quele momento em 2007 quando eu defendi a tese eram poucos pesquisadores que já estavam atentos a essas mudanças e na verdade dentro do que eu defendi como a tese o que nós vimos depois sendo desenvolvido de maneira mais forte foi essas técnicas que permitiram o que alguns autores chamam de jornalismo guiado por dados, data driven journalism, ou data journalism, sucintamente falando com essa terminologia e eu, na verdade a minha compreensão é muito mais ampla, hoje eu acho que o jornalismo de dados ele denomina mais especificamente aquilo que é das técnicas que são trabalhadas para que se

produzam conteúdos baseados em dados, principalmente relacionando isso ao que nós vivenciamos, mas pra nos América do Sul, Brasil, e também a realidade da Espanha, por exemplo, é em relação a abertura dos dados, ou seja, as leis de transparência, tem essa relação, ou seja maior divulgação dos dados de informações dos governos dos países dos estados enfim, isso levou a essa expansão também, né? Das técnicas de dados, de como se extrair dados, raspar dados, limpar dados, incorporar isso no uso, ma produção de conteúdos jornalísticos. Muitos relacionam o jornalismo de dados a informação numérica, àquilo que é estatístico, é uma forma de compreender. Pra mim, e na verdade o que eu defendi chamando de um paradigma era muito mais amplo. Então creio o jornalismo de dados está incorporado dentro disso que eu considero o paradigma maior porque eu estou falando de algo que diz respeito não apenas ao uso de técnicas para determinado produção de um conteúdo que vai nessa linha, mas que é um jornalismo todo estruturado desse modo que permite com que se tenha muito mais aquilo que tá la dito e pra mim isso é valido hoje o que se tornaram os sites jornalísticos, obviamente que isso se deu a partir do desenvolvimento todo, das tecnologias, das possibilidades, do emprego e desenvolvimento de sites, do ponto de visto do design interativo, das próprias linguagens de programação, daquilo que foi sendo agregado, do uso dos algoritmos, enfim, hoje a gente fala de máquina, do Machine Learning. Então esse contexto pra mim ele é muito mais amplo então se encaixa num paradigma que pra mim teve esse delineamento ainda no início da década de 2000 nessa primeira metade, minha tese é de 2007, entrei no doutorado em 2003, ela reflete obviamente aquele contexto dos primeiros anos da década de 2000 que marcam um diferencial muito forte da maneira como se fazia jornalismo digital, é o que eu to olhando, e muito do que a gente viu de evolução veio muito mais rapidamente e isso se expandiu fortemente e acabou gerando uma denominação muito forte a partir dessa que acabou sendo muito mais popular do que se tornou jornalismo de dados.

Data Literacy - Evaluation

Então, se é dessa forma que a gente pode compreender melhor o que é que, desde lá de trás...pra dizer o que eu disse obviamente foi necessário um histórico. Há pesquisadores que são de uma geração que viram florescer o jornalismo de precisão, ou aquilo que se chama de jornalismo computacional, acreditam que não há porque chamar isso de outro nome, já que é pra eles o mesmo jornalismo que se faz. É usar o poder computacional, as técnicas da reportagem assistida por computador, a hack, pra mim tá nesse histórico, nessa linha de tempo. Obviamente pra eu fazer uma linha de tempo hoje a gente teria que abarcar muito mais, já temos aí muito mais desenvolvimento. Só que eu acho que houve nesse período todo a confirmação, e aí eu to olhando pro meu trabalho, mas não só, vendo de outros colegas pesquisadores tanto aqui do Brasil, como na Espanha, Portugal e outros países, e que reforçam pra mim, e isso é curioso perceber como a gente teve... você tá olhando para os currículos dos curso de jornalismo e dizendo que as técnicas de jornalismo de dados tem que star inseridas porque precisa informar, precisa ensinar os alunos a trabalharem com isso. Ok, mas se a gente olha um pouquinho pra trás a gente vai

dizer: e as técnicas de reportagem assistida por computador, elas foram ensinadas? Nos anos 90, também não, nos anos 2000 muito pouco, veio para os currículos como informática, comunicação multimídia, conteúdo audiovisual. Como a gente agora também tá no contexto do desenvolvimento muito forte daquilo que se chamam ciência de dados aplicado e com esse outro fenômeno social do Big Data isso tudo se junta. Então acho que reforça, dá uma ênfase muito mais forte a esse necessidade de que há muitos anos, se a gente olha em 1970, o livro do jornalismo de precisão do Mayer é dos anos 1970, obviamente isso evoluiu, foi sendo bem desenvolvido. E hoje na verdade em qualquer curso de jornalismo, o reforço disso desde a segunda metade dos anos 1990 com a web, expansão, Internet comercial, tudo isso que passou realmente a ter um peso muito maior os currículos tiveram de ser atualizados. Uma nova disciplina que foi inserida como jornalismo digital, jornalismo online em alguns casos, depois o web jornalismo, que também foi incorporado em alguns currículos como disciplina optativa, mas foi a entrada primeiro por ali, também essas de jornalismo e tecnologia, comunicação e tecnologia, que em muitos currículos tá presente até hoje.

Sociodemographic characteristics

Nos aqui, no caso da faculdade de comunicação, não tive uma disciplina de jornalismo online, por exemplo, mas tive a oportunidade de fazer tirocínio docente que é uma prática que se pede no doutorado e também agora no mestrado que é um doutorando fazia um estágio, esse tirocínio na verdade acompanhava uma disciplina durante um semestre, que era uma forma de vc desenvolver suas aptidões já que a gente tá se formando num doutoramento ele é acadêmico, né? Pressupõe que você ao se formar vai dar aulas como professor. Então eu fiz tirocínio docente na disciplina de jornalismo digital, cujo professor responsável era o professor marcos palacios, criador junto com Elias Machado do GJOL, e também foi meu orientado tanto no mestrado quanto no doutorado e eu membro do gjol desde 2000. Continuo, hoje sou uma das coordenadoras junto com o Marcos.

Data literacy – UFBA

Então a Facom teve esse início no Brasil, to fazendo esse histórico só pra gente entender como nos estamos. Então começou como uma disciplina optativa justamente em função da pesquisa de marcos e Elias conjuntamente com a criação do GJOL e isso foi incorporado no currículo da habilitação jornalismo. Depois um pouco mais pra frente se tornou uma disciplina obrigatória, uma oficina, era oficina de jornalismo digital. Hoje ainda ela é conformada desse modo.

Sociodemographic characteristics

Eu fui professora por dois anos na universidade federal fluminense, sai da Espanha pro rio, fiquei dois anos na federal fluminense (08-10). Em 2010 fiz outro concurso que foi aqui na Federal da Bahia justamente nessa área de conhecimento: jornalismo digital para oficina de jornalismo digital.

Data literacy – UFBA

E quando vim pra cá em 2010, deixei a UFI, fiquei como professora dessa disciplina [jornalismo digital]. E nessa disciplina como ela é uma disciplina com uma carga horário de 130 horas semanais, ela é mais robusta, as oficinas todas de jornalismo na Facom elas têm 136 horas, então nessa disciplina é que a gente de fato trabalha muito mais fortemente porque tem esse tempo, né? São quatro, oito horas semanais. Mas o aluno de jornalismo ele desde o início do curso ele vai compreendendo. Na verdade, essa geração ela já vem com uma formação, digamos, já tem uma aptidão muito maior para tecnologias digitais.

Data literacy – New syllabus

Mas ocorre que nós estamos com uma outra necessidade evidentemente para além do que temos de disciplinas obrigatórias que trabalham comunicação e tecnologia, trabalha jornalismo digital, comunicação multimídia, como eu falei, nós estamos fazendo a atualização do projeto pedagógico. Jornalismo pelas diretrizes curriculares aprovadas no Brasil deixa de ser habilitação, passa a ser um curso. Então nós vamos passar a ser curso de jornalismo. Esse projeto pedagógico está em fase quase final de elaboração, nesse momento com o problema da pandemia nos estamos com o calendário acadêmico suspenso, isso atrasou a tramitação do projeto. Nesse projeto já previmos algumas disciplinas que tenham mais fortemente a interface com a computação e isso ministrado com disciplinas que já hoje existem no curso de computação. Porque pra ensinar programação, a gente, é bom que o professor de jornalismo compreenda, mas eu acho que essa é uma área que, como a gente sempre vê acontecer na organizações jornalísticas de modo geral, se trabalha com equipe interdisciplinares.

Data literacy - Evaluation

Eu acredito que essa interdisciplinaridade também é importante que ocorra na formação. Porque o professor de comunicação, o professor de jornalismo, ele deve conhecer, entender, compreender, mas também deve estar em conjunto com essa outra área também de programação, de computação, porque isso é importante pros estudantes, né? E claro, não é todo aluno que vai querer ser programador, jornalista programador. Acho importante que alguns que queiram ser invistam nisso, obviamente, mas o importante pra mim é que o estudante deva saber o que isso importa na atuação profissional de um jornalista, o que ele precisa saber dessa área pra executar o trabalho, pra formulação das pautas, de saber como posso apurar usando esses recursos todos que estão disponíveis. Porque o uso da internet é mais expandido, muita gente sabe pesquisar, mas não domina as técnicas, muitas vezes.

Data Literacy -UFBA

E quanto mais no caso do jornalismo de dados o que se pode fazer, eu vou pesquisar, vou ter acesso a grandes massas de dados, como é que vou fazer isso? De que maneira se pode atuar com a automatização, como posso fazer limpeza de dados, enfim. Todas essas questões específicas que são importantes e devem constar na formação do estudante de jornalismo em qualquer curso, seja

no Brasil, seja em qualquer país. Mas no nosso caso específico, estamos prevendo isso pra ocorrer nessa reformulação como disciplinas que estarão colocadas como disciplinas optativas, num primeiro momento, mas optativa não significa dizer que ela é menos importante, porque precisamos ter carga horária de obrigatória, carga horária de disciplinas optativas que forma o conjunto daquilo que se deve ter como grade curricular para a formação de um estudante na graduação. Então não é mais possível, não é só jornalismo de dados, né? Porque isso hoje está [20'] nessa realidade. Então, como eu disse, tem todo esse contexto. Quando a gente fala de *plataformização*, fala da Inteligência Artificial no jornalismo, tudo isso é mais amplo. Então técnicas de jornalismo de dados elas também precisam ser aprendidas pelos estudantes. Mas é mais, é um trabalho com esse lado da tecnologia mais forte pra eu entender como é que isso faz diferença num site, como eu identifico isso. Porque o jornalismo em si ele já está sendo produzido há muito com base nas tecnologias digitais. Então a maneira que o produto é distribuído ela pode diferenciar: pode ser uma revista impressa, pode ser jornal impresso, pode ser site, pode ser aplicativo, enfim, são multiplataformas pelas quais esses produtos circularem, mas a base é de como se produz é toda digital.

Data literacy – critical observation

[mais relevante o aluno entender em relação a fonte] Primeiro, antes de tudo, e como sempre, o aspecto da visão crítica ela é essencial acima de tudo e muito mais. Porque às vezes se pensa que a gente tem uma facilidade maior de eu encontrar essas fontes e eu lidar com esse tipo de fontes não significa que eu não tenha que seguir critérios, aqueles que são da validação pra essas informações e que são aqueles critérios os valores jornalísticos do que se tem como princípio básico do jornalismo. Então principalmente o aspecto crítico no uso e no emprego, eu posso ter acesso a muitas informações, mas eu preciso identificar, ou seja, eu preciso, uma vez que tenho informação, as fontes são fontes documentais, podem ser fontes de distintas naturezas, primárias, oficiais, especialistas, eles estão todos disponíveis através das redes, só que claro, como eu realizo um conteúdo jornalístico ele vai ter mais credibilidade, valor jornalístico, valor informativo, valor inclusive informativo jornalístico social, porque é essa função do jornalismo, ele é a instituição que tem essa validação pra produção desse conhecimento na sociedade, é a produção do conhecimento da atualidade é o jornalismo. Então pra eu fazer isso eu preciso dominar não só as técnicas, mas aquilo que jornalisticamente falando eu vou poder identificar e diferenciar o tipo de pauta que eu vou ter e aquela informação que é informação mais imediata com menos apuração, menos trabalho de investigação. Então, eu creio que essa fontes complexificam muito mais e permitem que o jornalismo possa de modo geral ter um conjunto de fontes e dados que podem trabalhar pra gerar um conteúdo mais qualificado inclusive. Agora, não vai se fazer isso, sem ser bom profissional, bom profissional entendendo qual é a função do jornalismo e aquilo que eu tenho conteúdos. Inclusive, a gente vê sempre, toda vez que vai se falar, ou referir a meios jornalísticos, quais são aqueles de maior relevância, quais são aqueles meios que fazer o seu trabalho de maneira diferencial e, portanto, ocupam seu

espaço muito bem. Basta ver mundialmente falando aqueles que são sempre citados. E os que coincidentemente também investem muito na formação e também no desenvolvimento de soluções tecnológicas, naquilo que a gente chama de inovação, com seus labs de inovação que colaboram pra que sejam esses meios, como o Guardian, NY Times, como El País, a gente sempre cita os mesmo, mas é que estes grandes teu seu espaço, seu lugar, o Washington Post, evidentemente tem outros jornais, outros meios locais que se destacam por fazerem jornalismo, por usarem bem essas tecnologias, aquilo que é mais de novidade pra agregar naquilo que a gente chama de novos formatos, lançar formatos novos. Quando o NY Times vai e lança um formato que ele nem sabia que passa até a denominar uma nova forma de jornalismo como o snowfall. Como é que ele chega em snowfall? Mas ali tem o peso da narrativa jornalística forte. Significa o que? Significa que tem que ter planejamento, tem que ter equipes boas de profissionais, porque é aí que vai diferenciar o tipo de pauta que vai ter e o tipo de conteúdo que vai produzir. Então, a base da formação eu acho que ela vai se complexificando muito mais e o que o estudante de jornalismo precisar perceber é que ele tem que expandir sempre o olhar dele e perceber mais: ele está se formando, mas, como gosta de falar esse meu colega professor pra mim uma referência que é o professor m. Palácios, formar é um verbo reflexivo, é formar-se. Então na universidade os estudantes aprendem, eles estão ali cursando uma habilitação, um graduação, mas eles precisam perceber a maneira com eles estão inseridos nisso porque não é dizer que um curso oferece várias cadeiras, várias disciplinas maravilhosas, tecnológicas absurdamente, e ele não sabe o que isso significa na sua profissão. Então é importante essa formação crítica, conhecer, acima de tudo ler bem, ter acesso aos distintos meios pra reconhecer e saber fazer. Não se deixar levar por aquilo que é o fácil, né? De repente, como aquilo que a gente sabe, copiar e colar o conteúdo. Ver que alguém fez algo, e a gente quer reproduzir? Só reproduzir?

Data literacy – journalist

[literacia dos jornalistas acompanha evolução digital] Olha, do que a gente acompanha, dos relatórios que são divulgados em vários países, com distantes realidades obviamente, ainda deixa muito a desejar, no sentido de que ainda falta esse domínio das técnicas. Mas eu também não creio que isso seja algo que numa redação, todos dominem, não vai ser isso. Nós tivemos um momento de muitas redações com suas equipes de dados, equipes específicas trabalhando conteúdos de dados, editorias de dados, hoje isso já em muitos casos já não tem um editoria conformada dessa forma, mas esses jornalistas estão na redação. Alguns que sabem, que aprenderam, e muitas vezes aprenderam por si, correram atrás, porque nem sempre essas organizações, algumas não dão essa formação. E a gente está num momento muito difícil de redações muito enxutas, no mundo inteiro, isso tem uma redução forte. E essas equipes de dados dentro das redações, elas também como editorias, na maior dos casos já se passou dessa fase porque o jornalismo foi confrontado com um novo ciclo de crise e agora se acirrou diante de uma situação de pandemia. Claro que o jornalismo sempre convive com crises, mas se sabe, aquilo que se avançou em determinados usos e

técnicas e tecnologias e etc. não há como seguir usando. Então, é isso que eu digo, muitas equipes não existem mais como editorias específicas, mas quem aprendeu dissemina isso pela redação de algum modo e um conteúdo ou outro vai ser mais trabalhado tirando partido disso que advém dessa forma, dessas técnicas de uso de dados no jornalismo. Mas eu acho não é assim tão homogêneo. O ano passado eu estive numa banca de defesa de tese sobre o jornalismo de dados na Universidade Carlos III em Madri e o trabalho do Leonardo La Rosa foi justamente em cima disso, das equipes de jornalismo de dados estavam trabalhando, como é que isso no seu dia a dia na produção de conteúdo como é que estava ocorrendo. Nessa pesquisa dele ele demonstra justamente que em muitos casos a Espanha, que foi um dos países que desenvolveu muito bem, muitos jornalismo de dados, já não, as equipes já estavam sendo desfeitas. Então, e muitos casos isso: ficam dois, três, quatro jornalistas, e isso dissemina, tem alguém que em algum momento, a depender do dia a dia da redação, vai compor, vai gerar conteúdos trabalhando a partir desses dados, com o uso dessas técnicas. É parte desse planejamento. Isso no Brasil também a gente teve a experiência. O estadão dados que foi uma equipe que trabalhou durante um tempo numa editoria conformada dessa maneira, mas também já não existe assim, e em outros meios também não. Então é isso. Eu acho que também tem esse item de realidade, na verdade, a realidade impõe muito nesse sentido: quando você está em momentos melhores, que permitem maiores investimentos, você consegue naquelas redações em grandes meios, meios de referência como a gente chama, desenvolvem pra aquilo que normalmente assim: olha isso aqui é pra onde está se caminhando. Precisamos disso, não dá pra continuar sem ter isso. Então, quer dizer, a gente vê várias fazes onde sempre teve algo, algum elemento, algumas dessas inovações entre aspas que foram incorporadas porque já não havia mais como não os ter. O caso das redes sociais, houve muita resistência, depois entrou. Então, a própria possibilidade da participação do usuário, do leitor, da audiência, como se queira chamar, com esses aspectos da interatividade, da interação, enfim. A outra fase da convergência, da integração de redação. E como foi sendo desenvolvidas algumas habilidades que o jornalista até então poderia não trabalhar efetivamente.

Credibility and interest

[relevância do jornalismo de dados no meio jornalístico] Nós tivemos eventos jornalísticos que trouxeram, mostraram a importância, sim, disso. Tivemos WikiLeaks, Panama Papers, e outros exemplos que foram muito bem trabalhados com equipes conjuntas e aí já é outra área que tem muitos pesquisadores também atentos de como se trabalha em colaboração e hoje é mais forte isso, que meios distintos passem a trabalhar conjuntamente na investigação de determinados casos, volta e meia a gente vê casos que ocorreram. Se essa importância, e acho que a gente teria que ter cuidado em como fala, é dizer dessa importância. É como eu falei, eu acho que não é redutor só pra falar de jornalismo

de dados com uso de aplicação dessas técnicas para lidar com dados. Eu considero que isso é mais amplo. Então, por isso que eu to falando, trabalhar o jornalismo e sua interface com a tecnologia. O jornalismo, na verdade, desde sempre ele lida, ele trabalha, ele existe porque há ali a tecnologia vai dizer pra ele o que ele faz, como é que ele pode. A própria máquina, o hino tipo, a prensa, então isso é evolução. Isso são fases distintas historicamente, períodos históricos que acompanham. No caso da tecnologia digital no jornalismo e se a gente considera a web, a partir dali até hoje o quanto modificou, modificou muitíssimo. O que eu estou dizendo é que eu acho que a gente precisa ter cuidado só pra entender que isso é mais amplo e quando a gente fala de jornalismo de dados, técnicas de jornalismo de dados, é algo específico, isso faz parte de algo mais amplo que quer dizer: nós temos desenvolvimentos que são contínuos nessa área e hoje passa pela não só o uso das narrativas em 360 graus, a IA, aquilo que a gente chama da possibilidade de realidade aumentada, o áudio em 360 graus, enfim. Tem muitos aspectos tecnológicos que estão aí. Aquilo que a tecnologia pode ofertar e como o jornalismo vai trabalhar para ofertar conteúdos de melhor qualificação pra gerar formatos mais inovadores, narrativas construídas de maneira distinta porque eu preciso ter o elemento da novidade alimentado porque eu tenho gerações que estão aí, inclusive jovens que em geral não são pessoas que leem da mesma maneira, ou seja, não acessam os conteúdos num site jornalísticos, as vezes não vão nem no site, consomem por outras plataformas. Então, eu acho que isso é muito mais complexo e o jornalismo de dados é um dos movimentos que a gente precisa compreender pra também ver como agrega no dia a dia da produção de uma organização jornalística, nos seus processos, nos processos naquilo que é o aspecto profissional, e naquilo que gera nos produtos dele como é que ele incorpora pra utilizar. Do mesmo modo... tá e os algoritmos? Isso não é jornalismo de dados, mas é uma realidade também. Por isso que eu to entendendo que isso é um aspecto. Quando eu me referi [39] a como eu construir no caso do que foi aquilo que apresentei como tese, era muito mais amplo, era considerar isso de modo dentro do que eu chamei de paradigma porque eu vejo dessa maneira mais ampla.

Production

O jornalismo de dados não é ele sozinho, ele faz parte, ele tá dentro de um conjunto, de um contexto, de novas técnicas, daquilo que o profissional deve conhecer, buscar agregar, ver de que maneira ele consegue inserir no seu trabalho, na sua atuação profissional, e do mesmo modo o professor, o pesquisador que é o que a gente sempre buscar falar. Porque os alunos eles acham assim, o que leva a eles a lerem pode ser só teoria, eles querem saber da prática. Então é por isso que eu sempre chamo atenção deles: você não vai fazer, sem conhecer, sem saber antes do que é que se está falando e do que se trata. Portanto não se chega na prática, sem a teoria. Porque pra eu ter a boa pratica, eu preciso de alguém pensando. É por isso também que tem o outro lado disso nesse tripé, que a gente fala da pesquisa teórica conceitual, no caso do ensino, e da pesquisa aplicada, porque com a pesquisa aplicada eu vou poder desenvolver

protótipos, ver o que funciona, experimentar, e com isso efetivamente contribuir gerando aquilo que pode melhorar os processos, os produtos jornalísticos.

Data Literacy – UFBA

[projeto pedagógico antes e depois] O projeto pedagógico ele precisa estar acompanhando aquilo que as novas diretrizes colocam para a formação do jornalista. Então tem um conjunto de aspectos que não só nessa área digamos mais entre aspas da tecnologia, dessa interface jornalismo e tecnologia. Então tem outros aspectos como a questão da diversidade, dos conteúdos...de estar nesses conteúdos aqueles aspectos não só da diversidade de gênero e outros. [42'] Os conteúdos, aquilo que a gente chama do ensino de conteúdos afro-brasileiros também inseridos aí também com já pesquisa, são pesquisadores dessa linha, e da acessibilidade dos conte...enfim, tem um conjunto. No caso específico nosso do projeto pedagógico, uma das mudanças, na verdade acho que é o principal eixo que nós estamos na busca da formação num eixo que é de formação mesmo dessa área, o que muda é o aspecto que é integrado [artigo visão da formação; integração]. Estamos fazendo essa modificação, essa atualização do projeto pedagógico dentro do Núcleo docente estruturante de jornalismo (NDE) é o órgão assessor do colegiado do curso de graduação. Eu faço parte e também outros colegas integram esse grupo. Nós escrevemos esse artigo como é que muda a visão da formação e passa a ter essa noção da integração. Hoje a gente tem uma oficina de jornalismo digital que está no sexto período, é lá no sexto semestre do curso que o aluno faz. Então, ele começa com oficina de jornalismo impresso, passa pela oficina de jornalismo audiovisual, rádio, telejornalismo, chega no de jornalismo digital e depois tem oficina de assessoria de comunicação. Hoje a realidade é essa. Com a reformulação, ele vai passar a ter, nós construímos a noção do jornalismo integrado, são disciplinas de jornalismo integrado desde jornalismo integrado 1 até o 6. Vai ter sempre um professor coordenador dentro de um desses jornalismo integrados que ele vai trabalhar sempre formatos e linguagens, uma em destaque em cada um desses jornalismo 1,2,3,4,5, pra chegar nos mais 4,5,6 ele vai trabalhar mais, quer dizer, essa complexidade se amplia pra trabalhar, o que? fazer com que esse estudante passa a ter uma formação que ele já vai pensando de modo integrado, que não é mais essa forma *compartimentalizada*, de ensinar e de aprender. Como ele vai atuar? Ele vai atuar no jornal, pressupõe que só vai ser jornalista de impresso. Não é mais isso. Então eu tenho que saber o texto, tenho que saber lidar com recursos que são audiovisuais, tenho que saber como faz podcast, como faz vídeo, e depois como eu integro isso tudo, como eu vou pensar na circulação desses produtos considerando essas várias plataformas, como eu vou adequar esses conteúdos pra diversas plataformas. Então isso é uma mudança na forma da concepção desse currículo para a formação na graduação em jornalismo. A cabeça é essa noção: trabalhar de maneira integrada. O jornalismo de dados ou tudo isso que significa trabalhar com aquilo que é hoje mais importante. Hoje aluno tem que entender a multimodalidade que é justamente trabalhar as distintas formas, percebendo que eu vou trabalhar não só visando uma única linguagem,

formato, plataforma, e sempre pensado, até porque a gente sabe que jornalismo é isso. Quem trabalhava em redação de jornal sempre estava com o fotógrafo e com o cara da equipe de editoração que ia diagramar, ou seja, naturalmente somos uma profissão que sempre trabalhamos de modo interdisciplinar. Só que a gente mesmo no interdisciplinar, era uma interdisciplinaridade, digamos, mais básica, e cada um na sua. Mas a gente, e a realidade disso da convergência, da integração, já mudou muito nas redações, pena que aquilo que a gente viu florescer de alguns modelos de integração que também teve suas consequências algumas fatídicas, né, do ponto de vista profissional de aquilo de achar que você quer um grande profissional que faça tudo e receba menos, enfim, essas distorções que sempre ocorrem, nós não estamos livres disso. Mas, não há como retroagir nisso que avançou, então a gente precisa ir olhando mais a frente porque nós acreditamos que trabalhando dessa maneira integrada você tem condição muito mais de ir agregando aquilo que vai ocorrendo porque nós estamos numa área como em qualquer outra. A medicina é assim. Quer dizer, vai pensar no uso da robótica na medicina. Quem entrou lá atrás, não. Óbvio. Então, isso tudo é aquilo que vai nos desenvolvimentos e que de algum modo tem que tá contemplado na formação de alguma maneira.

[como alunos são preparados pra fazer análise de dados] É o que se espera que a gente lá na frente de fato quando isso estiver implantado implementado que possa gerar isso. Nós fizemos um concurso ano passado, justamente, uma vaga de concurso pra professor de jornalismo que foi justamente pensando nessa forma de já a partir das necessidades apontadas por essa construção do projeto pedagógico que esse professor tivesse esse perfil de ser alguém que trabalhe com essa realidade. Nos já fizemos já tem uma pessoa aprovada que tem esse perfil mesmo. Só não tomou posse ainda porque a gente tá nessa situação desde março, justamente quando ela ia tomar posse, entrar em exercício, isso está suspenso até que a gente tenha condição de retomar em modo presencial. Mas nos estamos justamente pensando nisso e, do outro lado, os professores que hoje já são do curso também atentos porque também é uma atualização que cada professor, cada professora, também precisa ter. E isso é do ofício no caso de professor de qualquer área, ele precisa se atualizar. Então nós estamos com um corpo docente que hoje com essa nova professora que foi aprovada no concurso que é da formação do eixo mesmo da formação profissional mais pesada dessas disciplinas que vai trabalhar jornalismo integrado com essa... a expectativa é justamente poder abarcar isso e ofertar essa formação também de compreender, de como gerar visualização, de como fazer, enfim. Mas a gente precisa finalizar pra isso tramitar, porque precisa passar pelas instâncias de aprovação dentro da universidade pra que depois disso possa estar devidamente implementado. A nossa expectativa é de que em 2021 1, ou seja, primeiro semestre de 2021, já pudesse estar com esse novo projeto. Agora não sei se a gente consegue em 2021.1.

[docentes qualificação pra nova necessidade] Na verdade dentro da estrutura da Universidade a gente só pode fazer concurso tendo vaga, sendo autorizado, ou

seja, vagas novas não estão [52'26"] só se o professor se aposenta. E o governo federal também tem colocado algumas limitações pra isso. Então a UFBA conseguiu fazer, teve um edital no ano passado, esse que eu me referi, então a gente teve justamente a vaga de uma aposentadoria de uma professora. Fizemos esse concurso logo. Então nos uma Universidade pública federal dependemos disso, as estaduais também passam por essa via. Nas privadas é uma outra forma, um outro sistema. Então nós temos feito esse esforço. A Facom hoje o corpo docente tem professores que estão na casa há mais tempo, alguns já quase indo pra se aposentar. São, na verdade, professores que estão atentos e que sabem que isso é um desafio que vem pra todos nós. Então eu não estou ministrando aula agora na graduação, disciplina, desde que assumi a direção da faculdade, e eu estou no meu segundo mandato, eu entrei no primeiro mandato em 2013 e 2017 fui reeleita, ou seja, eu vou até novembro do ano que vem, novembro de 2021, na graduação não posso dar aula porque eu tenho muitas tarefas e não posso deixar os estudantes a mercê daquilo que é necessário também fazer como diretora. Então, mas nos tivemos professores substitutos pra essa parte e que estavam cumprindo bem até então. Agora a gente já tá com a professora que é professora efetiva e ela é também alguém dessa área de formação como eu. Então na posso dizer que a gente tem uma realidade que é complicada nesse aspecto, mas que nos exige muito, assim como diariamente. Mas quem é pesquisador também tem esse dado que é um dado de realidade também obviamente que significa dizer que como você é pesquisador você está se atualizando. Não há como...é por isso que nós sempre e eu faço questão sempre de destacar e ressaltar isso: não há como se ensinar na graduação onde for sem se fazer pesquisa porque você cria um hiato muito forte. Então, por essa natureza de que a gente tem professores que são também pesquisadores nos facilita por um lado. Isso é parte. Mas não quer dizer que não tem desafio, obviamente tem desafio. Tem desafio porque tem uma parte que a gente tem que considerar das dificuldades que se interpõem pelas questões de limitação orçamentária, em relação a equipamentos, as vezes a infraestrutura, isso nos lidamos. Na Facom a gente tem tido, pelo nos últimos anos, condição de nos mantermos também com uma atualização no que diz respeito a equipamentos a infraestrutura. Nos somos uma unidade academia com uma boa infraestrutura na Universidade, na significa que a gente não tenha sempre que melhorar, e é isso que é parte também do trabalho de quem tá na direção de uma faculdade. Então isso nos ajuda por um lado. Mas tem muitos entraves. Tem aquilo que a gente quer muito, sabe o que é, mas tem que adequar, tem que adaptar, enfim, mas faz parte. Não existe um mundo perfeito, como não existe pra empresa jornalística, ela não consegue muitas vezes fazer frente a todas essas demandas que são inclusive de tecnologia mesmo, pra poder investir. Então é a realidade e a gente precisa lidar e aprender também.

Big Data

[Big Data no contexto jornalístico] Compreender o que é isso e ter o dado, aquilo que a gente chama de dado de realidade, a concretude de dizer o que é isso e

como isso afeta a as vezes a gente não percebe é que é o ponto. Então, o que eu falei é: como na Facom nós temos grupos de pesquisa, essa disciplina que eu referi que é uma disciplina obrigatória, ela é teórica, é de fundamentos, como a gente chama, que é comunicação tecnologia, que é ministrada pelo professor André Lemos. Então nós temos uma dinâmica interna de pesquisa de pesquisadores dessa área. A Facom na verdade o programa de pós-graduação foi um dos primeiros no Brasil a ter linha de pesquisa em cibercultura. Então, esse contexto na Facom desse estudo da área já vem desde os anos 1990. Então, o aluno já vai sendo de algum modo inserido e vai compreendendo muito porque pessoas que são inclusive referencia na área são professores deles. Então isso faz diferença, mas isso não quer dizer que claramente todos os estudantes dominem isso. Não da pra dizer que num conjunto de 60/65 estudantes as vezes você tem essa compreensão de maneira uniforme. Não é assim porque o processo de aprendizagem não ocorre dessa forma, mas isso é muito, é, nesse sentido, é facilitador pra compreensão desses fenômenos dessa área do que a gente chamava de cibercultura e do que a gente tem hoje como comunicação e culturas digitais. Então o Big Data que é esse fenômeno social que significa assim hoje e o contexto lá nos anos 90 já apontava que os objetos todos vão gerar dados. Aquilo que se fala da internet das coisas, que eu sei que eu tenho celular, uma TV, que passam a ser equipamentos inteligentes, eles geram dados, eles possuem dados. Nós estamos produzindo dados e isso vai criando essa massa imensa de dados que a gente chega no Big Data, ou seja, nas cidades. Tudo que tá nas cidades inteligentes como chama, né? Então esse fenômeno ele está no nosso dia a dia só é a gente compreender melhor aquilo que tá fazendo. Como a gente diz as vezes: tenha cuidado com as informações que você está deixando, quais são suas pegadas que você deixa digitalmente. Por que tanta gente se surpreende “entrei no site tal e to navegando e aparece no meu email publicidade daquilo”. Por que? Porque tem algoritmos e eles vão registrar a nossa navegação os nossos dados. Essas tecnologias como a gente sabe elas geram muita informação sobre nos mesmos, são dados que passam a figurar aí, para aqueles que detém, é o oligopólio da tecnologia, a gente sabe disso né, conhecido pela sigla GARFAM. E nós geramos esses dados e elas é que estão se apoderando desses dados que geramos, é uma parte do Big Data, digamos assim. Mas esse é um fenômeno que está aí, está no dia a dia de todo mundo. Então é essa noção de como lidar com isso, que daí é importante. Por isso que é importante esse contexto de disciplinas teóricas que vão ajudar com esses teóricos da área a compreender melhor e também gerar a criticidade a reflexão que é importante, problematizar, conhecer, saber como atuar, como produzir, e aquilo que a gente tá fazendo. Porque efetivamente tudo o que a gente faz está gerando informação.

Credibility and interest

[transparência no jornalismo de dados e credibilidade] Veja, eu acho que esses limites do tratamento de dados é o que a gente tava falando há pouco, existem as leis. As leis de acesso a informação, as leis em cada país, elas colocam, e nesse aspecto o jornalismo de dados tem relação direta com a entrada em vigência da leis de acesso a informação em países onde elas passaram a vigoraram a pouco

tempo considerando a realidade dos países dos nórdicos, Finlândia e tal, onde tem as leis de transparência mais antigas, o reino unido que também há bastante tempo, mas, por exemplo, na Espanha a lei de acesso a informação de lá ela é dois anos posterior em relação ao Brasil, a do Brasil é de 2011, e passou a vigorar mesmo 2012. Então a nossa lei já vem aí. E nesse momento, com o governo que nós temos, já tivemos algumas situações de tentativas de limitar a lei de acesso a informação e isso é prejuízo principalmente no caso do jornalismo que trabalha, não só o jornalismo, mas principalmente afeta o trabalho do jornalismo de modo geral porque você não tem como recorrer a lei. Mas isso, claro, a gente também recente teve outra decisão do STF de assegurar que a lei seja efetiva como ela esta, sem esses limites. Mas os desafios são desafios que se interpõem de maneira geral ao jornalista e ao jornalismo. Não dá pra dizer que eu vou ter um aspecto que isso vai me favorecer só e por daí eu vou ter credibilidade. O jornalismo pra ele ser jornalismo ele precisa ser credível. E a legitimidade e a autoridade jornalística advêm justamente da maneira, quer dizer, uma organização jornalística que preserva, que atua a partir daquilo que são os códigos *neontológicos*, que assegura isso, trabalha com transparência. Porque o jornalismo, a depender as vezes de alguns meios, buscam, cobram transparência dos outros, mas nem sempre também são transparentes. E o jornalismo de dados tem esse outro aspecto: eu preciso dizer onde eu obtive, como eu obtive, e se a partir daqueles dados que foram coletados se constrói uma base de dados nova, ela tem que estar acessível, o WikiLeaks deixou lá disponível. Então essa é uma das premissas, por exemplo. Se você está trabalhando com essa massa de dados e algum foi compilando e consolidou essas informações, elas precisam, precisa estar dito onde que está aquela base para que as pessoas por aqueles mesmo parâmetros possam ter acesso e quiser fazer suas investigações vão dali buscar aqueles dados que você teve acesso, mas que você compila e resulta numa base de dados que pode estar também disponibilizada. Então isso também é transparência, garantia de transparência do trabalho. É desafio o tempo inteiro, porque acho que o jornalismo desafia concretamente o profissional e os meios diariamente. Então, mas, o que é da legitimidade jornalística, da credibilidade, esses...isso não dá pra brincar. Então não é o jornalismo de dados sozinho que vai garantir. Então acho que isso tá dentro do que é, a...quando a gente pensa percebendo com clareza a função do jornalismo, o que é jornalismo, como instituição social, é, habilitada e a quem se dá... nas sociedades democráticas, quem é que produz o conhecimento da atualidade é o jornalismo. Então esses valores são o básico pro jornalismo. A legitimidade, credibilidade, claro que é afetada e algumas situações como a gente vive aí, né? Tem aqueles que querem o tempo inteiro questionar ou dizer: “que jornalismo? Isso é *fake*”. E a gente tem dois presidentes que vivem a repetir isso, um norte-americano e o nosso aqui. Então isso é muito ruim obviamente, é prejudicial que pessoas públicas que estão em cargos para os quais foram eleitos fiquem o tempo inteiro com declarações que só querem desqualificar e que isso não tenha valor numa sociedade. Num país em que a gente tem tanta desigualdade como Brasil e que o acesso a educação de modo geral é precário, então, a gente, é dilema muito grande, é muito terrível. Mas num momento como esse pandemia que se sabe onde, quais

fontes eu devo procurar pra ter informação de credibilidade e confiável, porque sabem os parâmetros e qual é o processamento dessa informação pra validar. Nenhum meio vai querer arriscar a sua credibilidade de anos, algumas com marca centenárias, ir pro ralo, então isso é o dia a dia do jornalismo.

[Falsa autoridade do dado] É, mas é aquilo que a gente também gosta sempre de frisar, o jornalismo de dados não é só falar de número. A parte de número é importante, mas tem muita informação que não é número. Jornalismo de dados não necessariamente é só eu ver que matérias estão citando, tao colocando estatística, se tem tabelas produzidas, gráficos e etc. Isso é muito complexo. Às vezes se produz gráfico que quem não sabe ler o que está num gráfico, por exemplo, é...por isso que a gente fala, de novo, não é só dizer por dizer, a interdisciplinaridades nas equipes é importante também por isso. Porque esses parâmetros...o jornalista ele tem muito, ele vai lidar com esse grande conjunto de informações de varias ordens e de modo muito complexo, então gerar estatísticas a partir do que ele apura, também a partir do jornalismo de dados com essas técnicas pode gerar o que são tabelas, aquilo, converter muita coisa em números e dar essas informações em números daquilo que a gente tem de maior objetividade nesse caso. Mas jornalismo de dados não é só número. Tem vários conteúdos distintos. O uso dos números é um, quer dizer, eu posso me debruçar numa investigação pra descobrir quantos se chegou de desvio, enfim, em operações sei lá lava jato e outras, vai se saber montantes e montantes que vão gerar a partir da quela confrontação com os números. Mas isso é uma coisa. As outras relações que se vai descobrindo também pra mostrar como isso ...isso é importante, ligar os fatos. É isso que... o jornalista tem que olhar interpretar, saber fazer as conexões e revelar aquilo que está encoberto, não só em números, mas em informações dispersos e claro fazer entrevistas, qualificar essas informações por outro lado. Tem de repercutir, tem que consultar essas fontes de modo diverso. Então não é só o número em si.

Appendix 7 – Federal University of Bahia journalism degree reformulation

Fábio Sadao Nakagawa, Suzana Oliveira Barbosa, Washington José de Souza Filho

O ENSINO DE JORNALISMO E A CONVERGÊNCIA

Integração das redações como proposta pedagógica

Introdução

A Faculdade de Comunicação da Universidade Federal da Bahia comemorou, em 2017, seus 30 anos de atuação como Unidade Universitária autônoma. Mais antiga e mais importante no Estado da Bahia, a Faculdade de Comunicação se desmembrou, em 1987, da Escola de Biblioteconomia e Comunicação, com base em deliberação emitida pela reunião do Conselho Universitário, em 13 de novembro de 1986. Funcionando desde 1999 na atual sede no Campus de Ondina (FIG.1), a FACOM oferece, atualmente, 30 vagas semestrais no curso de graduação em Comunicação com habilitação em Jornalismo; 30 vagas semestrais no curso de Comunicação com habilitação em Produção em Comunicação e Cultura; 60 vagas anuais na área de concentração em Cinema e Audiovisual para os estudantes do Bacharelado Interdisciplinar em Artes e por volta de 60 vagas anuais para o mestrado e o doutorado do Programa de Pós-Graduação em Comunicação e Cultura Contemporâneas (PósCom).

Figura 1. Fachada do Prédio da Faculdade de Comunicação no Campus de Ondina



Fonte: Acervo Labfoto, FACOM | UFBA

O curso de Jornalismo, na UFBA e na Bahia, começou em 1950 (sendo o terceiro no país), “com duas turmas de complementação de currículos para pessoas com graduação em outras áreas. Depois foi interrompido e só voltou a funcionar regularmente em 1962” (MARTINS; GUIMARÃES, 2008, p.15). Em 1969, ele se tornou uma habilitação do curso de Comunicação (BARBOSA; COSTA, 2018, p.2), oferecida pela Escola de Biblioteconomia e Comunicação e, a partir de 1987, passou a ser ofertado como habilitação já na Faculdade de Comunicação, cuja primeira sede foi no Campus do Canela. Atualmente, a habilitação em Jornalismo conta com 311 estudantes ativos no sistema do Colegiado de Graduação¹.

Com índice de 4,8301, o curso em Comunicação com habilitação em Jornalismo da FACOM/ UFBA obteve, em 2015, o maior conceito no Exame Nacional de Desempenho de Estudantes, estando entre os 18 cursos de graduação em Jornalismo que conquistaram o índice de excelência pelo Ministério da Educação, num conjunto de 275 cursos que foram avaliados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, por meio do ENADE.

A grade atual do curso foi implementada em 1999 por meio da última reforma pedagógica, que atingiu tanto a habilitação em Jornalismo quanto a habilitação em Produção em Comunicação e Cultura, que passou a

¹ Dado fornecido pela Secretaria do Colegiado de Graduação em Comunicação da FACOM no dia 27 de março de 2018.

funcionar a partir de 1996. Em 2002, teve início os trabalhos de uma comissão de professores do Departamento de Comunicação para realizar ajustes nas grades curriculares das duas habilitações. No entanto, ao longo dos anos, a proposta foi sendo abandonada. Por conta disso, em 2012, o Colegiado de Graduação, ao invés de retomar a proposta de ajuste curricular, decidiu, em conjunto com o Departamento de Comunicação, instaurar quatro comissões compostas por professores, respectivamente, das áreas das Teorias da Comunicação, do Jornalismo, da Produção Cultural, de Cinema e Audiovisual, para que elas pudessem elaborar os novos projetos pedagógicos dos dois cursos de progressão linear já existentes e, também, o projeto pedagógico do curso em dois ciclos em Cinema e Audiovisual.

No ano seguinte, com a publicação das Diretrizes Curriculares nacionais para o curso de Jornalismo – ocorrida por meio da Resolução Nº 1 de 27 de setembro de 2013 – o Departamento de Comunicação refez as comissões de áreas em três Núcleos Docentes Estruturantes (NDE) para que pudessem repensar e refazer os projetos pedagógicos para os cursos de graduação de Jornalismo, do novo a ser criado de Cinema e Audiovisual e da habilitação Produção em Comunicação e Cultura.

Passados esses quase cinco anos, os projetos encontram-se em fase adiantada, porém ainda não foram finalizados. Diferentes foram os fatores que dificultaram o cumprimento do prazo estabelecido pela própria Resolução de 2013 para a implementação das novas diretrizes nos cursos de Jornalismo. Dentre eles, destacam-se: a) a difícil negociação interna para a alteração da habilitação do curso de Comunicação para o curso de Jornalismo. Tal alteração não atingiu a Produção em Comunicação e Cultura que continua funcionando como uma habilitação do curso de Comunicação; b) o exaustivo diálogo entre professores e estudantes para elaborar as grades curriculares que ocorreu em diferentes assembleias e em várias reuniões dos NDEs, do Departamento de Comunicação e da Congregação da Unidade; c) a constante renovação dos integrantes dos NDEs por conta dos afastamentos dos docentes da Unidade para realizarem suas pesquisas de doutorado ou de pós-doutorado e, também, por causa da entrada de novos professores no Departamento de Comunicação; assim como d) as diversas greves dos servidores técnicos-administrativos e dos professores ocorridas na UFBA que dificultaram o avanço dos trabalhos e a periodicidade das reuniões. Apesar destes e de outros contratemplos, o Núcleo Docente Estruturante do Curso de Jornalismo conseguiu elaborar a grade curricular e os ementários das disciplinas do novo curso e pretende implementá-lo no ano de 2019.

Neste capítulo, discutiremos primeiramente as características principais do curso de Comunicação com habilitação em Jornalismo que está em funcionamento. Em seguida, vamos expor a ideia e a dinâmica da grade elaborada pelo NDE de Jornalismo e a lógica de funcionamento das disciplinas denominadas Jornalismo Integrado I, II, III, IV, V e VI, por considerarmos que agregam aspecto inovador para a nova grade a ser implementada. Por fim, vamos discorrer sobre o contexto da convergência jornalística e do jornalismo em redes digitais em relação à nova proposta pedagógica do curso de Jornalismo da FACOM/UFBA.

A alteração de habilitação para curso de graduação em Jornalismo

O atual curso de Comunicação com habilitação em Jornalismo da FACOM/UFBA tem a carga horária total de 2.952 horas. São 2176 horas resultantes das 21 disciplinas obrigatórias, 476 horas das 7 disciplinas optativas e 300 horas de atividades complementares obrigatórias. O curso é oferecido somente no período matutino e precisa ser integralizado no tempo mínimo de 4 anos e máximo de 7 anos. As disciplinas obrigatórias são ministradas por 23² docentes do Departamento de Comunicação, de um total de 41 professores, sendo que 12 deles são responsáveis pelas disciplinas específicas da habilitação e 16 atuam também na pós-graduação – a maior parte no Programa de Pós-Graduação em Comunicação e Cultura Contemporâneas (PósCom, da FACOM | UFBA) e os demais no Programa Multidisciplinar de Pós-Graduação em Cultura e Sociedade, do Instituto de Humanidades, Artes e Ciências Professor Milton Santos (IHAC | UFBA).

Além do conjunto de disciplinas da grade curricular atual, o estudante de graduação em Comunicação com habilitação em Jornalismo também encontra outras instâncias da Unidade que colaboram com a sua formação profissional, tais como: o Laboratório de Fotografia (Labfoto); Laboratório de Audiovisual (LabAV); Laboratório de Rádio; o Programa de Educação Tutorial da FACOM (PETCOM); a Agência Experimental em Comunicação e Cultura (AECC); a Agência de Notícias em CT&I - Ciência e Cultura; a Agenda Arte e

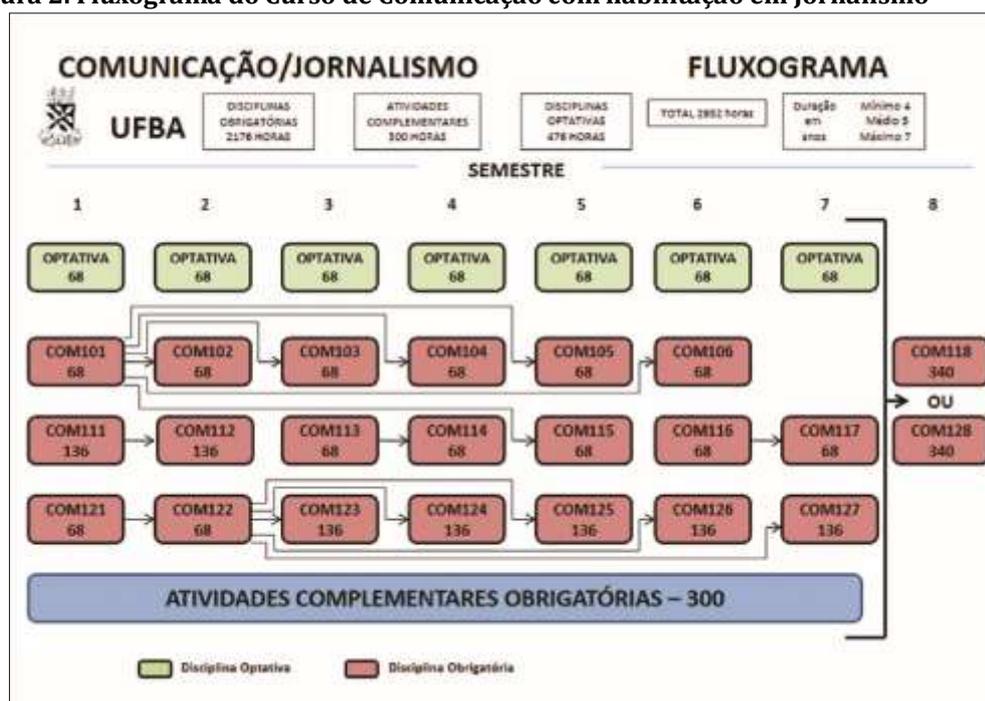
2 Todos os 23 professores possuem o título de doutorado e estão inseridos no regime de 40 horas com dedicação exclusiva.

Cultura; e a Produtora Júnior. Caso o estudante decida direcionar sua formação para a pesquisa, ele encontra também suporte nos diferentes grupos de pesquisa e de estudo existentes na Unidade³. O elenco de disciplinas da atual grade tem como base o seguinte perfil do egresso⁴:

O Jornalista é o profissional que trabalha com a produção de bens simbólicos (notícias; reportagens; etc.) e que, através do domínio de linguagens e técnicas específicas, elabora interpretações da realidade, atuando tanto nos meios de comunicação de massa convencionais (rádio; jornal; e TV) quanto nos mercados emergentes no campo da comunicação institucional.

Inserido no contexto da prática moderna da comunicação massiva e das especificidades da linguagem jornalística através dos meios de comunicações tradicionais (impresso, radiofônico e televisual) e, na época, também pelo emergente meio digital; o perfil do egresso está refletido na grade curricular em funcionamento, principalmente, pelo conjunto de disciplinas específicas da habilitação em Jornalismo. Denominadas como oficinas e voltadas ao ensino teórico-prático em jornalismo impresso, radiojornalismo, telejornalismo, jornalismo digital e assessoria de comunicação; as cinco disciplinas compõem o eixo de formação profissional da área que se inicia com as disciplinas do primeiro ano, “Teorias do Jornalismo” e “Comunicação Jornalística” (FIG.2). Este eixo irá sofrer uma grande modificação no novo projeto pedagógico do curso de Jornalismo, que será discutida no próximo item.

Figura 2. Fluxograma do Curso de Comunicação com habilitação em Jornalismo



Fonte: Acervo Núcleo de Comunicação e Extensão, FACOM/UFBA⁵

3 <https://www.facom.ufba.br/portal2017/pagina/23/pesquisa>

4 Disponível em: <<https://alunoweb.ufba.br/SiacWWW/ListaDisciplinasEmentaPublico.do?cdCurso=307140&nuPerCursoInicial=20091>>. Acesso em: 03 abril 2018, 12:09:00.

5 Listagem das disciplinas que compõem a grade em funcionamento e os seus respectivos códigos: COM101 – Teorias da Comunicação; COM102 – Semiótica; COM103 – Estética da Comunicação; COM104 – Comunicação e Tecnologia; COM105 – Comunicação e Política, COM106 – Comunicação e Cultura Contemporânea; COM111 – Oficina de Comunicação Escrita; COM112 – Oficina de Comunicação

Pensado como uma habilitação do curso de Comunicação, Jornalismo compartilha 14 disciplinas obrigatórias, num total de 21, com a habilitação em Produção em Comunicação e Cultura. Ou seja, 66,66% das disciplinas obrigatórias estão presentes nas grades curriculares das duas habilitações, de modo que apenas 33,34% das disciplinas obrigatórias estão destinadas a oferecer os conteúdos específicos de cada área de atuação profissional.

Na atual discussão para a reformulação dos projetos pedagógicos de ambos os cursos, apenas o eixo de formação teórico fundamental da área da Comunicação será mantido. Nas grades em funcionamento, ele é constituído pelas disciplinas “Teorias da Comunicação”, “Semiótica”, “Estética da Comunicação”, “Comunicação e Tecnologia”, “Comunicação e Política” e “Comunicação e Cultura Contemporânea”. São disciplinas voltadas para uma formação humanística e generalista do futuro jornalista e produtor cultural, e que problematizam e discutem as interfaces entre as diferentes esferas de atuação do sujeito, o fenômeno comunicacional e a área de pesquisa em Comunicação.

O eixo se mantém nas novas grades curriculares com as atualizações dos ementários, objetivos e bibliografias dos seus seis componentes e, também, com a mudança do nome da disciplina “Semiótica” para “Comunicações, Meios e Linguagens”. Tal alteração visa integrá-la ao próprio eixo, constituído pelas fronteiras estabelecidas entre a Comunicação e outros fenômenos socioculturais e não apenas a relação da Comunicação com uma linha de pesquisa específica. Apenas na grade curricular do curso de Jornalismo, manteve-se neste eixo, a disciplina denominada “Comunicação e Ética”, que, na grade em funcionamento, pertence ao segundo eixo de formação tanto generalista quanto profissional.

O eixo composto pelas seis disciplinas de formação teórica na área da Comunicação da grade em funcionamento, com o acréscimo da última disciplina mencionada, irá constituir o primeiro eixo de formação da nova grade curricular. Este conjunto de disciplinas foi mantido em função da sua própria atualidade e, também, devido ao seu importante papel na excelência da formação teórico-conceitual dos estudantes de ambos os cursos. Além disso, o eixo visa a atender tanto o eixo de fundamentação humanística quanto o eixo de formação contextual, ambos mencionados no Art. 6º das Novas Diretrizes dos Cursos de Jornalismo, que versa sobre os seis eixos necessários para a constituição do perfil egresso em Jornalismo e para a promoção das competências deste profissional.

Além da disciplina que estabelece as relações entre a Ética e a Comunicação, será resgatado do segundo eixo de formação da grade em funcionamento o conjunto composto por três disciplinas relacionadas ao trabalho de conclusão de curso que, pelo Art. 11º das atuais Diretrizes, se constitui como componente curricular obrigatório. No entanto, com a revisão da grade curricular pelo novo Projeto Pedagógico do Curso de Jornalismo, a fase de elaboração e aprimoramento do projeto de pesquisa ou de produto deverá ser realizada em um semestre, ao contrário do que ocorre na grade em funcionamento, na qual a feitura e a revisão do projeto são feitas no decorrer de dois semestres, restando apenas um semestre para a realização e defesa do TCC. O Prazo muito exíguo foi revisto pela nova proposta, na qual o desenvolvimento da pesquisa ou a construção do produto jornalístico deverão ser feitos ao longo de dois semestres por meio das disciplinas de TCC I e II.

A alteração de habilitação do curso de Comunicação para um curso de Jornalismo, além das mudanças na grade curricular, no modo de entender e compreender a própria formação do Jornalista, possibilitou ao NDE, em parceria com os estudantes e professores da FACOM, rever e refazer o perfil do profissional que o curso pretende formar. Na nova proposta, pretende-se que o profissional seja aquele com “formação em cultural geral, humanística e teórica do campo” para que possa exercer a profissão de jornalista “com criticidade, reflexão, contextualização e ética para produzir narrativas, textos e discursos de acordo com a diversidade de

Audiovisual; COM113 – Comunicação e Atualidade I; COM114 – Comunicação e Atualidade II; COM115 – Comunicação e Ética; COM116 – Elaboração de Projeto em Comunicação; COM117 – Desenvolvimento Orientado de Projeto; COM118 – Trabalho de Conclusão de Curso em Comunicação; COM121 – Teorias do Jornalismo; COM122 – Comunicação Jornalística; COM123 – Oficina de Jornalismo Impresso; COM124 – Oficina de Radiojornalismo; COM125 – Oficina de Telejornalismo; COM126 – Oficina de Jornalismo Digital; COM127 – Oficina de Assessoria de Comunicação e COM128 – Trabalho de Conclusão de Curso Jornalismo.

linguagens possibilitadas pelas tecnologias da comunicação, pelas mídias e pela multiplicidade de plataformas”⁶.

Integração como estratégia metodológica-pedagógica

A proposta aqui apresentada tem a sua inserção em um entorno acadêmico que, ao menos por duas décadas, como referem Barbosa e Palacios (2018), tem sido ativamente caracterizado, na Faculdade de Comunicação da UFBA, pela busca da associação entre investigação teórica e aplicada, assim como suas ramificações pedagógicas em termos do ensino do Jornalismo frente aos desafios de uma sociedade em rápida mudança tecnológica.

A reestruturação curricular tem como base o ambiente digital. Ela está relacionada a um contexto cujas transformações afetam os modos de produção, as práticas, a distribuição, a circulação, a recirculação dos conteúdos e a própria maneira como o público interage com os conteúdos. Este contexto demarca, para o jornalismo, um dos processos centrais de reestruturação devido à digitalização e às redes telemáticas: o processo de convergência jornalística (SALAVERRÍA, GARCÍA AVILÉS, & MASIP, 2010; SALAVERRÍA; NEGREDO, 2008), o qual é caracterizado pelo *continuum* multimídia de fluxo horizontal e dinâmico (BARBOSA, 2013). Nesse processo, o aspecto da comunicação móvel – *smartphones, tablets, e-readers*, consoles de jogos, *ultrabooks* etc. – tem a sua relevância, pois demarca o ecossistema midiático a partir do desenvolvimento e da apropriação dos dispositivos móveis dentro da noção de distribuição e publicação multiplataformas.

A integração de redações – que materializa a convergência jornalística – serve como modelo para a estratégia metodológico-pedagógica que se definiu para a reestruturação curricular do Curso de Jornalismo da FCOM, caracterizada pela busca da associação entre professores de disciplinas com diferentes linguagens, a partir da forma em que o curso de Jornalismo é ministrado.

A atual forma de reorganizar as redações, que inspira a reestruturação, está baseada em modalidades e fórmulas (SALAVERRÍA; NEGREDO, 2008), a partir de uma referência que tem o jornal impresso como base. A sua representação pode ser estabelecida pela análise dos tipos de coordenação (AVILÉS *et al*, 2009; 2014) e graus de adaptação dos conteúdos, em função das características dos meios (CABRERA, 2010). As modalidades que caracterizam a integração de redações são divididas em duas – uma que é definida pela relação entre os meios; a outra que tem relação com a amplitude da área de cobertura e fórmulas, determinadas por três vertentes (SALAVERRÍA; NEGREDO, 2008, p. 127-133):

1. Convergência a dois, entre o meio impresso e *online*. A modalidade é considerada a mais prática e natural, porque permite racionalizar a colaboração que deve existir na prática. A consequência mais significativa, porém, é a mudança do ciclo de fechamento, além da atuação dos jornalistas, em um ambiente multiplataforma - destinado a mais de um meio e ambiente;
2. Convergência a três, entre os meios impresso e *online* com a televisão. A alternativa representa a maior capacidade de transmissão da internet, pois beneficia a convergência de formatos do jornalismo baseados em texto, áudio e vídeo, além de interativos;
3. Convergência a quatro, entre meio impresso, *online*, televisão e rádio. O modelo corresponde ao mais complexo, já que cada meio depende de uma atuação baseada em especificidades, relacionadas com as características de cada um. As diferenças tornam mais difíceis a adaptação dos jornalistas.

6 O novo perfil do egresso do curso de Jornalismo foi proposto nas reuniões do Núcleo Docente Estruturante do Curso de Jornalismo da FCOM | UFBA e revisto na assembleia com os estudantes e nas reuniões do Departamento de Comunicação. Posteriormente, será publicado no texto do projeto pedagógico que está em fase de finalização. Os professores que compõem atualmente o NDE são: Suzana Barbosa, Washington José de Souza Filho, Fábio Sadao Nakagawa, Carla de Araújo Risso, Lia Fonseca Seixas, Nuno Manna Nunes Cortes Ribeiro e Natália Moura Pacheco Cortez. Em outros momentos, o NDE também contou com a participação dos seguintes professores: Maria Lucineide Andrade Fontes, Leonor Graciela Natansohn, Giovandro Marcus Ferreira, Regina Lúcia Gomes Souza e Silva, Rodrigo Rossoni e Simone Terezinha Bortoliero.

A outra modalidade de integração está vinculada à escala geográfica, que depende da área de atuação dos meios e o tamanho das redações. As características são duas, entre nacionais e locais ou regionais. As fórmulas são determinadas pela associação entre os jornalistas para o trabalho nas redações integradas, nas quais é feita a introdução de uma mentalidade multiplataforma; por ritmo informativo; e autonomia da produção editorial, com a separação do setor comercial.

Do ponto de vista histórico, a experiência que é considerada um marco do processo de integração das redações foi desenvolvida nos Estados Unidos, no *Tampa News Center* (LAWSON-BORDERS, 2006), posto em marcha em 2002, depois de um período de transição digital, iniciado em 1994 (SALAVERRÍA; NEGREDO, 2008, p. 33). A integração representava a convergência entre três meios em uma empresa da Flórida: um jornal impresso, uma televisão e a produção de um jornal *online*.

O processo estava baseado em sete níveis de atuação conjunta, de profissionais com perfis diferentes, dos quais o mais destacado era a atuação de fotógrafos dos jornais e os câmeras da emissora de televisão. Os dois grupos realizavam as coberturas para os dois meios, nas quais utilizavam os equipamentos adequados para cada um. A atuação conjunta de profissionais com perfis distintos, entre eles, impôs a necessidade de adaptar a linguagem que os jornalistas utilizavam em função dos meios em que atuavam. A diferença não era uma consequência da integração, mas de uma assimilação da prática em meios diferentes (SILCOCK; KEITH, 2006).

A integração, definida pela divisão ou concentração de tarefas em uma redação, depende de um padrão, que está relacionado à cultura profissional de cada país (ESSER, 1998). Em um meio como a televisão, da mesma forma que nos outros, a integração pode ser implantada através do computador, como equipamento básico. O equipamento era utilizado para realizar diversas tarefas, mesmo antes do processo de digitalização, na década de 80 (WHITEMORE, s/d; KEIRSTEAD, 2005), em emissoras como a *CNN* (*Cable News Network*) e a *BBC* (*British Broadcasting Company*).

Na primeira fase, os computadores são instalados numa rede, de modo que todos os textos informativos estão dentro do mesmo sistema. De qualquer posto de trabalho se pode ver quais são as notícias do dia previstas para a emissão, quais estão elaboradas e, finalmente, o que está pronto para ir ao ar. [...] o editor tem uma poderosa ferramenta para calcular exatamente como está ocupado o tempo previsto, controlar o número de fitas utilizadas na emissão, as deixas de cada peça ou suas marcações para a troca de câmera, a identificação das pessoas que aparecem nos ‘créditos’ e até mesmo os textos que serão lidos pelo apresentador no estúdio através do teleprompter. (BANDRÉS *et al*, 2002, p. 26-27)⁷

A estratégia de integrar as redações permitiu o surgimento de funções especializadas, profissionais que contribuem para o gerenciamento dos recursos tecnológicos (BANDRÉS *et al*, 2002; JIMÉNEZ, 2014; SOUZA FILHO, 2015). A integração é avaliada de uma forma controversa, pelo o que representa como uma “estratégia empresarial” (SALAVERRÍA; NEGREDO, 2008, p. 73). No Brasil, estudos sobre a integração das redações (BARBOSA, 2009; JORGE; PEREIRA, 2009) relaciona questões influenciadas pela tecnologia, como a mudança da identidade profissional do jornalista.

A integração busca uma forma de reorganizar as redações, por meio da qual as rotinas permitem desenvolver um processo, decorrente da convergência jornalística, que através da adaptação dos antigos meios à realidade do ambiente digital, alcança a distribuição multiplataforma (CABRERA, 2010). A integração tem sido compreendida como uma estratégia que modifica a forma de produzir e distribuir os conteúdos jornalísticos (SALAVERRÍA; NEGREDO, 2008, p. 16). Em uma perspectiva relacionada com a formação de jornalistas adequados ao ambiente digital, existe a necessidade do seu reconhecimento como uma referência, uma alternativa metodológica, que contribui para o ensino de práticas profissionais do Jornalismo, adequadas à realidade atual.

7 Tradução dos autores. No original: “En una primera fase los ordenadores se instalan en rede, de forma que los textos

Estrutura da nova grade curricular

A proposta considerada pela Faculdade de Comunicação da UFBA expressa a compreensão de integração das redações como uma estratégia de inovação do desenvolvimento das práticas do Jornalismo, em torno de um contexto de produção, relacionado às condições tecnológicas e sociais (DOMINGO, 2008), as quais têm interferência sobre os dispositivos e a cultura profissional. A avaliação tem sido realizada pelo Núcleo Docente Estruturante do Curso de Jornalismo, que está dedicado à reestruturação curricular. A proposta, da forma aqui apresentada, é o resultado de um processo contínuo de discussões iniciado a partir de 2015. O conjunto formado pelas disciplinas denominadas Jornalismo Integrado I, II, III, IV, V e VI, oferecidas por 6 (seis semestres), permitirá ministrar os conteúdos dos componentes curriculares a serem substituídos – COM 123 (Oficina de Jornalismo Impresso), COM 124 (Oficina de Radiojornalismo), COM 125 (Oficina de Telejornalismo), COM 126 (Oficina de Jornalismo Digital) e COM 127 (Assessoria de Comunicação). A base da proposta é a aplicação de uma nova metodologia, com foco na convergência jornalística de conteúdos para divulgar a produção por meio das multiplataformas, e considerando o formato multimídia, ou como produções independentes, relacionadas com as linguagens de meios específicos, em um espaço como um portal na *web*.

A integração das práticas, subdivididas nas referências às linguagens desenvolvidas por meios específicos, entre os classificados como “velhos meios”, favorece a formação dos estudantes de Jornalismo da Faculdade de Comunicação da UFBA, especificamente, por estar baseada em atividades mais condizentes com o atual ambiente dos meios de comunicação, marcado pelas mudanças estabelecidas com a tecnologia digital. A partir da associação das práticas profissionais relacionadas com os meios apresentadas na atual estrutura da habilitação em Jornalismo de maneira separada – já que a estrutura curricular está baseada na oferta de disciplinas caracterizadas pela referência a um meio – as atividades das disciplinas na nova conformação permitirá aos alunos uma melhor experimentação da integração permitida pela tecnologia.

A associação das práticas, estabelecida por meio de unidades temáticas relacionadas a cada meio, mantém a noção de diferença entre as linguagens, ao mesmo tempo em que valoriza a noção de multiplataforma. Esta é uma característica visível na atuação de diversas organizações jornalísticas, principalmente naquelas cuja presença digital está atrelada ao modelo de portais *web*. A proposta metodológica-pedagógica permitirá uma formação mais ampla sobre as práticas profissionais do

de todo el informativo se encuentran dentro del mismo sistema. Desde cualquier puesto de trabajo se puede ver cuáles son las noticias del día previstas para su emisión, cuáles están redactadas y, finalmente, las que ya están listas para salir al aire. [...] el editor tiene una herramienta poderoso para calcular exactamente cómo se ocupa el tiempo previsto, para controlar los números de las cintas que intervienen en la emisión, los pies de cada pieza o sus ‘coleos’, los rótulos de las personas que aparecen en los ‘totales’ y hasta los textos que leerá el presentador en el plató para sus entradillas del teleprompter.”

Jornalismo, ainda que possibilite o reconhecimento das especificidades de cada meio. No escopo da proposta também está prevista a associação conjunta de práticas do fotojornalismo e do design editorial, as quais são separadas na estrutura curricular atual do curso.

As disciplinas relacionadas com a proposta – Jornalismo Integrado I, II, III, IV, V e VI – serão ministradas pelos professores que atuam nas atuais Oficinas, subdivididas em componentes curriculares, em um total de 6 (seis), sendo uma por semestre, através de práticas jornalísticas relacionadas pelas linguagens – digital, áudio e audiovisual –, complementadas pelas novas disciplinas – Fotojornalismo e Design Editorial –, estabelecidas com a reestruturação. A indicação é que cada componente possa ter uma carga de 136 horas/aula por semestre.

Os componentes curriculares denominados como Jornalismo Integrado I, II, III, IV, V e VI substituirão as disciplinas oferecidas atualmente com as designações de COM 123 (Oficina de Jornalismo Impresso), COM 124 (Oficina de Radiojornalismo), COM 125 (Oficina de Telejornalismo), COM 126 (Oficina de Jornalismo Digital) e COM 127 (Assessoria de Comunicação). A convergência de conteúdos será realizada através de práticas estabelecidas a partir da linguagem relacionada com o conteúdo de cada componente com a denominação de Redação Integrada de Jornalismo, destinada, na forma de produtos, para uma publicação digital (Tabela 1).

Tabela 1 – Disciplinas Jornalismo Integrado I, II, III, IV, V e VI

Disciplina	Jornalismo Integrado I	Jornalismo Integrado II	Jornalismo Integrado III	Jornalismo Integrado IV	Jornalismo Integrado V	Jornalismo Integrado VI
Finalidade	Apuração, pauta e redação do texto escrita	Texto escrito, captação de áudio e redação, organização e hierarquização da informação	Apurar, produzir pautas, além de redigir, organizar e hierarquizar a informação	Texto escrito, produções editoriais, visual e multimídia	Comunicação estratégica, produção, circulação e recirculação da informação	Jornalismo multimídia
Linguagens	Escrita e visual	Escrita, visual e sonora	Escrita, visual e audiovisual	Escrita, visual e audiovisual	Escrita, visual e audiovisual, em torno do planejamento e cobertura jornalística	Escrita, visual e audiovisual
Formato	Notícia e entrevista	Notícia, entrevista e perfil	Entrevista, perfil e reportagem	Projeto editorial, dossiês especiais e grandes reportagens	Produto editorial	Produtos experimentais jornalísticos

Fonte: Elaboração dos autores

O ensino das práticas, por meio das disciplinas denominadas Jornalismo Integrado I, II, III, IV, V e VI, deve ser realizado sucessivamente e de forma separada por meio dos componentes com a designação de Jornalismo Integrado I, Jornalismo Integrado II, Jornalismo Integrado III e Jornalismo Integrado IV. A disciplina Jornalismo Integrado I deve ter um caráter introdutório sobre as práticas do Jornalismo, com uma ementa que estabeleça que o conteúdo deva incluir as técnicas de seleção (apuração), produção (pauta), realização (entrevista, reportagem e redação) e edição da informação, por meio das diferentes formas para divulgar a notícia. Nas disciplinas Jornalismo Integrado V e Jornalismo Integrado VI, se trabalhará com maior articulação, aprofundamento e maior nível de complexidade, uma vez que elas reunirão conjuntamente o que foi sequencialmente sendo abordado nas anteriores, enfocando ainda mais na produção multimídia, considerando os diversos formatos de conteúdos, os aspectos de gerenciamento e coordenação editorial, e enfrentando o desafio constante da inovação quanto ao uso de tecnologias, softwares, bases de dados, algoritmos, linguagens de programação e de publicação, sistemas de gerenciamento de informações, técnicas de visualização, metadados semânticos, etc.

Os conteúdos das disciplinas Fotojornalismo e Design Editorial – aprovadas em uma proposta anterior de 2007, mas não implantadas – estarão incorporados às disciplinas de Redações Integradas de Jornalismo para permitir o desenvolvimento das atividades voltadas para a produção da forma prevista. As suas ementas, assim como as outras, buscam considerar a possibilidade do desenvolvimento de atividades específicas, mas precisam ser contempladas como parte da proposta de estabelecer uma nova metodologia adequada ao ambiente digital para promover a convergência jornalística de conteúdos.

Considerações finais

Como aqui foi demonstrado, a proposta das disciplinas denominadas Jornalismo Integrado I, II, III, IV, V e VI como estratégia metodológica-pedagógica do processo de reestruturação curricular para o Curso de Jornalismo da Faculdade de Comunicação da Universidade Federal da Bahia está em consonância com a noção de que na contemporaneidade a formação do profissional jornalista deve ocorrer de maneira integrada e articulada.

Com a sua implementação, busca-se conferir maior identidade para o próprio curso, estabelecendo a sua diferenciação em relação à habilitação Produção em Comunicação e Cultura. Isso porque, até então, as habilitações Jornalismo e Produção estão separadas apenas por sete disciplinas específicas de cada uma. A nova estrutura curricular possibilitará também a melhor articulação entre os docentes das disciplinas e, conseqüentemente, a configuração de um conjunto diversificado de produtos laboratoriais experimentais que

possam efetivamente trabalhar a inovação do ponto de vista dos formatos, das técnicas e métodos de apuração, bem como das linguagens para os conteúdos.

Evidentemente, toda mudança traz sempre desafios e também algumas resistências. Mas, levando-se em conta o histórico da Faculdade de Comunicação e o que se conseguiu realizar ao longo dos seus 30 anos no que diz respeito aos desenvolvimentos teóricos e à pesquisa aplicada com as ramificações e desenvolvimentos nas atividades pedagógicas de ensino, acredita-se que se obterá êxito com a proposta que aqui foi apresentada.

Referências

- AVILES, J. A. *et al.* Integración de redacciones en Austria, España y Alemania: modelos de convergencia de medios. **Anàlisi**, **38** (p. 173 - 198), 2009.
- AVILÉS, J.A. *et al.* Media convergence revisited - lessons learned on newsroom integration in Austria, Germany and Spain. **Journalism Practice** (p. 1-12), 2014.
- BANDRÉS, E. *et al.* **El periodismo en la televisión digital**. Barcelona: Paidós, 2002.
- BARBOSA, S.; PALACIOS, M. La experiencia con la investigación teórica y aplicada sobre el periodismo convergente y la movilidad en el Proyecto Laboratorio de Periodismo Convergente. **Revista Interin**, v. 23, n. 1, jan./jun. 2018. p. 256-276. Disponível em: <http://seer.utp.br/index.php/i/article/view/604/pdf>
- BARBOSA, S. “Jornalismo convergente e *continuum* multimídia na quinta geração do jornalismo nas redes digitais”. In: CANAVILHAS, J. (Org). **Notícias e Mobilidade. O Jornalismo na Era dos Dispositivos Móveis**. Covilhã, PT: Livros LabCOM, 2013. p. 33-54.
- BARBOSA, Suzana. Convergência jornalística em curso: as iniciativas para integração de redações no Brasil. In Rodrigues, Carla. **Jornalismo online: modos de fazer** (p. 35 - 55). Rio de Janeiro: Ed. PUC-Rio/ Editora Sulina, 2009.
- BARBOSA, Suzana; COSTA, Leonardo. **30 anos da Faculdade de Comunicação. Jornal da FACOM/UFBA**. Salvador, fev. 2018. Editorial, p.2.
- CABRERA, María Angeles. La distribución multiplataforma, objetivo del proceso de convergencia. In García, Xosé; Fariña, Xosé. (Coords.). **Convergencia digital: reconfiguración de los medios de comunicación en España** (p. 149 - 165). Santiago de Compostela. Universidade, Servizo de Publicacións e Intercambio Científico, 2010.
- DOMINGO, David. Inventing *online* journalism: a constructivist approach to the development of *online* news. In: PATERSON, Chris; DOMINGO, David (Eds.). **Making online news** (p. 15 - 28). New York: Peter Lang, 2008.
- ESSER, Frank. Editorial structures and work principles in british and german newsrooms. **European Journal of Communication**, Vol. 13(3), (p. 375 - 405), 1998.
- JIMÉNEZ, José. Montaje y postproducción audiovisuales. In: LEÓN, Bienvenido (Coord.). **Detrás de las cámaras** (p. 187 - 207). Salamanca: Comunicación Social, Ediciones y Publicaciones, 2014.
- JORGE, Thaís; PEREIRA, Fábio. Jornalismo on-line no Brasil: reflexões sobre perfil do profissional multimídia. **Revista Famecos**, Porto Alegre, 40 (p. 57 - 62), Dezembro, 2009.
- KEIRSTAND, Philip. **Computers in broadcast and cable newsrooms. Using technology in television news production**. Mahwah, NJ: LEA (Lawrence Erlbaum Associates), 2005.



- KÜNG, L. **Innovators in Digital News**. London, New York: I.B.Tauris & Co. Ltd/Reuters Institute for the Study of Journalism, University of Oxford, 2015.
- LAWSON-BORDERS, Gracie. **Media organizations and convergence. Case Studies of Media Convergence Pioneers**. Mahwah, New Jersey: LEA, 2006.
- MACHADO, E.; & PALACIOS, M, (Orgs). **O Ensino do Jornalismo nas Redes de Alta Velocidade, Metodologias e Softwares**. Salvador: EDUFBA, 2007.
- MARTINS, Patrícia; GUIMARÃES, Marluce. Todo jornal tem e sempre teve função política. In: MATTOS, Sérgio. (Org.). **Memória da imprensa contemporânea da Bahia**. Salvador: Instituto Geográfico e Histórico da Bahia, 2008. p. 15-35.
- SALAVERRÍA, R., GARCÍA AVILÉS, J. A., & MASIP, P. M. Concepto de convergencia periodística. In: López García, X. & Pereira Fariña, X. (Orgs.). **Convergencia digital: reconfiguración de los medios de comunicación en España** (p. 41-64). Santiago de Compostela: USC, 2010.
- SALAVERRÍA, Ramón; NEGREDO, Samuel. **Periodismo integrado – convergencia de medios y reorganización de redacciones**. Barcelona: EditorialSol90, 2008.
- SILCOCK, Willian; KEITH, Susan. Translating the tower of Babel? Issue of definition, language, and culture in converged newsrooms. **Journalism Studies**, 7, no. 4 (p. 610-627), 2006.
- SOUZA FILHO, Washington José de. A polivalência dos jornalistas em emissoras de TV do Brasil e de Portugal na distribuição multiplataforma de conteúdos. **Ancora**. v3 no. 1(p.7-26), 2015.
- WHITEMORE, Hank. **CNN – a história real** (C. Nogueira, Trad.). São Paulo: Editora Best Seller (Obra original publicada em 1990), s/d.

Appendix 8 – Catholic University of Brasilia data journalism course syllabus

		UNIVERSIDADE CATÓLICA DE BRASÍLIA - UCB <small>Q5 07 LOTE, Nº01, AGUAS CLARAS BRASÍLIA - DF - BRASIL CEP 71920-540</small>								
MATRIZ CURRICULAR										
CURSO: LPE10 - JORNALISMO DE DADOS										
CURRÍCULO: LPE10E01 - CURRÍCULO 01										
HABILITAÇÃO: ESPECIALIZAÇÃO			REGIME: SERIADO		TURNO: INTEGRAL		STATUS: ATIVA			
Nº	Período	Cód. Disciplina	Disciplinas Obrigatórias	Grupo Disc.	Pré-Requisito	CH Teórica	CH Prática	CH Total	Créd. Acad.	Créd. Finan.
1	1ª	LVNFG001	ÉTICA, CONHECIMENTO E ATUAÇÃO PROFISSIONAL	Módulo I		45	15	60	4	4,00
2	1ª	LPE10001	NUMERIALIZAÇÃO DO JORNALISMO: DO COMPUTADOR AOS DADOS	Módulo I		45	15	60	4	4,00
TOTAIS						90	30	120	8	8,00
Nº	Período	Cód. Disciplina	Disciplinas Obrigatórias	Grupo Disc.	Pré-Requisito	CH Teórica	CH Prática	CH Total	Créd. Acad.	Créd. Finan.
3	2ª	LVNBE001	GESTÃO CRIATIVA	Módulo II		15	15	60	4	4,00
4	2ª	LPE10002	PROGRAMANDO DADOS	Módulo II		45	15	60	4	4,00
TOTAIS						60	30	120	8	8,00
Nº	Período	Cód. Disciplina	Disciplinas Obrigatórias	Grupo Disc.	Pré-Requisito	CH Teórica	CH Prática	CH Total	Créd. Acad.	Créd. Finan.
5	3ª	LPE10003	JORNALISMO COMPUTACIONAL	Módulo III		0	60	60	4	4,00
6	3ª	LPE10004	VISUALIZAÇÃO E MAPEAMENTO DE DADOS	Módulo III		45	15	60	4	4,00
TOTAIS						45	75	120	8	8,00
Nº	Período	Cód. Disciplina	Disciplinas Obrigatórias	Grupo Disc.	Pré-Requisito	CH Teórica	CH Prática	CH Total	Créd. Acad.	Créd. Finan.
7	4ª	LPNFG002	TRABALHO DE CONCLUSÃO DE CURSO (TCC)	Módulo IV		0	30	30	2	2,00
TOTAIS						0	30	30	2	2,00
CH Básica:	390									
CH Mínima Optativa:	0									
CH Mínima Atividade Complementar:	0									
CH Mínima Atividade Virtual:	0									
CH Total:	390									
Total de Créditos Financeiros:	26,00									

mayara.oliveira
18 de julho de 2019
1/1

Appendix 9 – Catholic University of Brasilia digital journalism course syllabus

Universidade Católica de Brasília		UNIVERSIDADE CATÓLICA DE BRASÍLIA - UCB										
		Q5 07 LOTE, Nº01, AGUAS CLARAS BRASÍLIA - DF - BRASIL, CEP 71920-540										
		Reconhecida pela Port. Min. nº 1827 de 28/12/1994, publicada no D.O.U de 30/12/1994, Recred. pela P.M. nº 624 de 17/05/2012, D.O.U de 18/05/2012. Credenciada pela P.M. Nº 4419, de 30/12/2004.										
CURRÍCULO PLENO DO CURSO												
Curso:	GPE20 - JORNALISMO DIGITAL	Currículo:	GPE20801									
Carga Horária Total:	3060	Créditos Totais:	136									
Carga Horária Disc. Obrigatória:	2800	Créditos Disc. Obrigatória:	128									
Carga Horária Disc. Optativa:	160	Créditos Disc. Optativa:	8									
Carga Horária Disc. Eletiva:	0	Créditos Disc. Eletiva:	0									
Carga Horária Ativ. Complementar:	100	Créditos Ativ. Complementar:	-									
Grau:		Data Início:	01/08/2019									
Habilitação:	BACHARELADO	Data Término:										
Aprovação: Curso de Graduação em Jornalismo Digital, Bacharelado, criado pela Resolução Consep nº 5 de 23/04/2019.												
TURNOS DISPONÍVEIS: <input checked="" type="checkbox"/> Matutino <input type="checkbox"/> Vespertino <input type="checkbox"/> Noturno <input type="checkbox"/> Integral												
Sem.	Seq.	Cód. Disc.	Disciplinas	Pré-Requisito(s) Disciplina[s]	Min.	Cr.	Qtd. Créd.	Carga Horária				
								Teor.	Sup./Ori.	Prát./Lab.	Tot.	
12	1	GVNF0011	INTRODUÇÃO À EDUCAÇÃO SUPERIOR				4	80	20	0	80	
12	2	GPED0025	PRÁT. PROF. I - INOVAÇÃO, JORNALISMO E CONTEÚDOS DIGITAIS				4	30	100	30	160	
12	3	GPED0033	HISTÓRIA DA COMUNICAÇÃO				4	80	20	0	80	
12	4	GPED0031	GÊNEROS E TÉCNICAS DE JORNALISMO I				4	20	20	40	80	
12	5	GPMB0005	TEXTOS E PRÁTICAS DIGITAIS				4	80	20	0	80	
22	6	GVNF0010	INTRODUÇÃO À PESQUISA CIENTÍFICA				4	80	20	0	80	
22	7	GPED0001	DESIGN E PLATAFORMAS DIGITAIS				4	20	20	40	80	
22	8	GPE20002	PRÁTICA PROFISSIONAL II - JORNALISMO DE DADOS	2			4	30	100	30	150	
22	9	GPED0032	GÊNEROS E TÉCNICAS DE JORNALISMO II	4			4	20	20	40	80	
32	10	GVNF0013	HUMANIDADE, SOCIEDADE E ÉTICA				4	80	20	0	80	
32	11	GPE20061	TEORIAS DA COMUNICAÇÃO	3			4	80	20	0	80	
32	12	GPED0040	RADIOJORNALISMO	9			4	20	20	40	80	
32	13	GPED0025	PRÁT. PROF. III - LINGUAGEM AUDIOVISUAL				4	20	40	20	80	
32	14	GPED0003	VISUALIZAÇÃO DE DADOS	7			4	20	20	40	80	
42	15	GVNF0008	EMPREENDEDORISMO				4	80	20	0	80	
42	16	GPE20004	WEBJORNALISMO	9			4	20	20	40	80	
42	17	GPED0045	TEORIAS DO JORNALISMO	11			4	80	20	0	80	
42	18	GPED0043	TELEJORNALISMO	9			4	20	20	40	80	
42	19	GPED0030	FOTOJORNALISMO	13			4	20	20	40	80	
52	20	GPED0038	PRODUÇÃO E EDIÇÃO EM RÁDIO	12			4	20	20	40	80	
52	21	GPED0035	JORNALISMO POLÍTICO E ECONÔMICO	9			4	80	20	0	80	
52	22	GPE20005	ESTÉTICA E CULTURA MIEDÁTICA				4	20	20	40	80	
52	-	-	DISCIPLINA OPTATIVA				4	0	0	0	0	
62	23	GPED0048	MÉTODOS E TÉCNICAS DE PESQUISA EM COMUNICAÇÃO				4	20	20	40	80	
62	24	GPED0006	REDAÇÃO CONVERGENTE	12,18,21			4	20	20	40	80	
62	25	GPED0007	ÉTICA E COMUNICAÇÃO EM REDE	10			4	80	20	0	80	
62	-	-	DISCIPLINA OPTATIVA				4	0	0	0	0	
72	26	GPED0040	COMUNICAÇÃO PÚBLICA E GOVERNAMENTAL				4	20	20	40	80	
72	27	GPED0008	JORNALISMO AUDIOVISUAL MULTIPLATAFORMA	18			4	20	20	40	80	
72	28	GPED0009	REDES E MÍDIAS SOCIAIS				4	20	20	40	80	
72	29	GPED0028	ESTÁGIO SUPERVISIONADO I - JORNALISMO				4	0	80	0	80	
82	30	GPED0046	TCC - JORNALISMO	23			4	0	80	0	80	
82	31	GPED0010	GESTÃO EM COMUNICAÇÃO E ECONOMIA CRIATIVA				4	20	20	40	80	
82	32	GPED0029	ESTÁGIO SUPERVISIONADO II - JORNALISMO	29			4	0	160	0	160	
TOTAL:								136	1000	1080	720	2880

CURRÍCULO PLENO DO CURSO

Curso: GPE20 - JORNALISMO DIGITAL

Currículo: GPE20801

DISCIPLINAS OPTATIVAS DO CURRÍCULO

Sem.	Seq.	Cód. Disc.	Disciplina	Pré-Requisito(s)		Qtd.	Carga Horária			
				Disciplina(s)	Min. Cr.		Créd.	Teo.	Sup./Ori.	Prát./Lab.
	33	GPINF012	LÍNGUA BRASILEIRA DE SINAIS - LIBRAS		0	4	60	20	0	80
	34	GVINF014	INGLÊS INSTRUMENTAL		0	4	60	20	0	80
	35	GPE0044	TEORIA, ESTÉTICA E CRÍTICA DO AUDIOVISUAL		0	4	60	20	0	80
	36	GPE0042	EDUCAÇÃO PARA A MÍDIA		0	4	60	20	0	80
	37	GPE20011	PROGRAMAÇÃO EM JORNALISMO		0	4	20	20	40	80
	38	GPE20012	JORNALISMO INTERNACIONAL		0	4	60	20	0	80
	39	GPE0037	COMUNICAÇÃO E GÊNERO		0	4	60	20	0	80
	40	GPE0044	FOTO PUBLICITÁRIA		0	4	20	20	40	80
	41	GPE0036	PRODUÇÃO E EDIÇÃO DE REVISTAS		0	4	60	20	0	80
	42	GPE0034	JORNALISMO ESPECIALIZADO		0	4	20	20	40	80
	43	GPE0046	GESTÃO DE MÍDIAS SOCIAIS		0	4	20	20	40	80
	44	GPE20013	JORNALISMO MÓVEL		0	4	20	20	40	80
TOTALS:						48	520	240	200	960

DISCIPLINAS ELETIVAS DO CURRÍCULO

Sem.	Seq.	Cód. Disc.	Disciplina	Pré-Requisito(s)		Qtd.	Carga Horária			
				Disciplina(s)	Min. Cr.		Créd.	Teo.	Sup./Ori.	Prát./Lab.
Grand Total						0	0	0	0	0
TOTALS:						0	0	0	0	0

Appendix 10 – University of Minho official note

Official note from University of Minho about the suspension of academic activity due to the coronavirus pandemic.



Universidade do Minho
UNIMINHO

Circular
VMT-LL-03/2020

Estratégia face à suspensão das atividades letivas presenciais

Como é do conhecimento de todos, por precaução profilática, o despacho RT-25/2020, publicado ontem, suspendeu as atividades letivas presenciais em toda a Universidade.

Esta suspensão tem implicações, não apenas ao nível das aulas e de outras atividades letivas, mas também ao nível da avaliação das aprendizagens dos estudantes e do registo de faltas e de sumários.

Estas implicações têm suscitado diversas questões por parte dos estudantes dos diversos ciclos de estudos que importa esclarecer.

Assim, no que concerne a aulas e outros tipos de atividades letivas, e até que estejam disponíveis condições e apoios para a eventual realização generalizada de ensino tecnologicamente mediado, devem os conselhos pedagógicos dar orientações internas sobre a estratégia a adotar de modo a garantir que os estudantes usem o tempo que durar a suspensão de atividades letivas presenciais para continuarem a sua formação, o que pode concretizar-se de diversas formas, consoante a natureza do curso e das unidades curriculares (UC) em que os estudantes estiverem inscritos.

No que respeita a atividades de avaliação das aprendizagens de caráter presencial, para além das provas académicas, também as outras atividades de avaliação de natureza presencial, de que são exemplo os testes e as apresentações de trabalhos, estão suspensas, devendo ser oportunamente remarcadas.

Enquanto as atividades letivas presenciais estiverem suspensas e não forem formalmente substituídas, pela Unidade Orgânica, por atividades tecnologicamente mediadas, não serão marcadas faltas aos estudantes.

No que concerne a sumários, enquanto as atividades letivas presenciais estiverem suspensas e não forem formalmente substituídas, pela Unidade Orgânica, por atividades tecnologicamente mediadas, não haverá registo de sumários.

Neste contexto, solicito às unidades orgânicas que comuniquem, preferencialmente até à próxima sexta-feira, qual a estratégia que irão adotar para, durante o período de suspensão das atividades letivas presenciais, continuarem a fomentar a formação dos estudantes dos respetivos cursos.

A Vice-Reitora para a Educação,

Appendix 11 – Email exchange with Público newspaper

Screenshots of the e-mail exchange with David Pontes, one of Público’s directors, about the visit to the newsroom, later cancelled by the data journalism Rui Barros due to the coronavirus pandemic.



Outlook - Alice Melo - Outlook - Google Chrome
outlook.office.com/mail/deeplink?version=2020052401.07&popoutv2=1&leanbootstrap=1

Responder a todos Excluir Lixo Eletrónico Bloquear

Re: Contato - Público

Para: Alice Melo <asam@ua.pt>
Cc: Rui Miguel Barros <Rui.Barros@publico.pt>
Assunto: Re: Contato - Público

Cara Alice,

Nesta altura, apesar de trabalhar connosco há pouco tempo, a pessoa mais indicada é o Rui Barros, que coloco em cópia.

Abç,

David Pontes

From: Alice Melo <asam@ua.pt>
Date: Thursday, 5 March 2020 at 09:57
To: David Pontes <dpontes@publico.pt>
Subject: RE: Contato - Público

Bom dia, David. Tudo bem?

Pedi, na semana passada, um contato de alguém do Público que possa me ajudar com informações sobre o jornalismo de dados do jornal. Meu objetivo é entrevistar os profissionais envolvidos com a produção de notícias baseadas em dados e observar o processo de produção de uma delas. Acha que é possível?

Obrigada,

Alice Assunção

Escreva aqui para procurar

12:08 02/04/2020

Outlook - Alice Melo - Outlook - Google Chrome
outlook.office.com/mail/deeplink?version=2020052401.07&popoutv2=1&leanbootstrap=1

Responder a todos Excluir Lixo Eletrónico Bloquear

Re: Contato - Público

<https://aliceassuncao.journoportfolio.com/>

De: David Pontes <dpontes@publico.pt>
Enviado: terça-feira, 10 de março de 2020 11:57
Para: Alice Melo <asam@ua.pt>
Cc: Rui Miguel Barros <Rui.Barros@publico.pt>
Assunto: Re: Contato - Público

Sim, claro.

From: Alice Melo <asam@ua.pt>
Date: Tuesday, 10 March 2020 at 11:37
To: David Pontes <dpontes@publico.pt>
Cc: Rui Miguel Barros <Rui.Barros@publico.pt>
Subject: RE: Contato - Público

Obrigada, David. Rui e eu estamos conversando para agendar o melhor dia para a entrevista. Seria possível visitar a redação do Público para observar como o trabalho de jornalismo de dados é feito?

Obrigada mais uma vez e um abraço,

Alice Assunção

De: David Pontes <dpontes@publico.pt>
Enviado: segunda-feira, 9 de março de 2020 20:50

Escreva aqui para procurar

12:09 03/06/2020

Outlook - Alice Melo - Outlook - Google Chrome
outlook.office.com/mail/deepink?version=2020052401.07&popout=2=1&cleanbootstrap=1

Responder a todos Excluir Lixo Eletrónico Bloquear

Re: Contato - Público

DP David Pontes <dpontes@publico.pt>
Ter, 10/03/2020 12:12
Para: Alice Melo
Cc: Rui Miguel Barros <Rui.Barros@publico.pt>
Sim, please.

From: Alice Melo <asam@ua.pt>
Date: Tuesday, 10 March 2020 at 12:06
To: David Pontes <dpontes@publico.pt>
Cc: Rui Miguel Barros <Rui.Barros@publico.pt>
Subject: RE: Contato - Público

Ótimo. Obrigada! Penso que também posso organizar essa visita com o Rui, certo?

Abraço,

Alice Assunção

<https://aliceassuno.journoportfolio.com/>

De: David Pontes <dpontes@publico.pt>
Enviado: terça-feira, 10 de março de 2020 11:57
Para: Alice Melo <asam@ua.pt>
Cc: Rui Miguel Barros <Rui.Barros@publico.pt>
Assunto: Re: Contato - Público

Escreva aqui para procurar

12:09 03/16/2020