ARTICLE

COMPETENCES AND SKILLS IN DATA JOURNALISM:
perceptions on the profile of Brazilian professionals

ANA PAULA OLIVEIRA
Universidade Municipal de São Caetano do Sul – São Paulo – Brazil
ORCID: 000-0002-4093-0590

ALAN CÉSAR BELO ANGELUCI
Universidade Municipal de São Caetano do Sul – São Paulo – Brazil
ORCID: 0000-0001-6596-2831

ABSTRACT – This study is an investigation of the main competences and skills necessary to become a data journalism professional in Brazil. To accomplish this, a theoretical-conceptual review was performed and data were collected from 14 semi-structured interviews conducted with two groups: the first with professionals who work with data journalism; and the second with researchers at universities across the country. The results from the analysis, supported by the ATLAS.ti software and by Bardin's studies, were organized into six categories that indicate logical and analytical reasoning as competences and skills necessary for using programming languages and for the capability of telling stories through interactive visuals. The results were made available in the beta version of the 2030 Data Journalism – an interactive cartography tool that uses geolocation and multimedia information to monitor the trends of this profession over the next years.
Key words: Data journalism. Digital journalism. Data imaging. Interactive mapping. University.

COMPETÊNCIAS E HABILIDADES NO JORNALISMO DE DADOS:
percepções sobre o perfil do profissional brasileiro

RESUMO – Neste estudo, buscou-se investigar quais seriam as principais competências e habilidades necessárias para atuar como profissional de jornalismo de dados no Brasil. Para isso, uma revisão teórico-conceitual foi realizada e dados foram coletados em 14 entrevistas semiestruturadas com dois grupos: o primeiro, com profissionais que atuam

com jornalismo de dados no mercado; e o segundo, com pesquisadores em universidades do país. Na análise dos resultados, feita com apoio do software ATLAS.ti e respaldada nos estudos de Bardin, foram organizadas seis categorias em uma matriz que, entre outros aspectos, indicam como competências e habilidades necessárias o raciocínio lógico e analítico na apropriação de linguagens de programação e a capacidade de contar histórias por visualizações interativas. Os resultados foram disponibilizados na versão beta do 2030 Data Journalism – cartografia interativa que visa acompanhar, por meio de geolocalização e informações multimídias, as tendências dessa especialidade nos próximos anos.

Palavras-chave: Jornalismo de dados. Jornalismo digital. Visualização de dados. Cartografia interativa. Universidade.

COMPETENCIAS Y HABILIDADES EN EL PERIODISMO DE DATOS: percepciones sobre el perfil del profesional brasileño

RESUMEN – En este estudio, se buscó investigar cuáles serían las principales competencias y habilidades necesarias para actuar como profesional de periodismo de datos en Brasil. Para ello, una revisión teórico-conceptual fue realizada y datos fueron recolectados en 14 entrevistas semiestructuradas con dos grupos: el primero, con profesionales que actúan con periodismo de datos en el mercado; y el segundo, con investigadores en universidades del país. En el análisis de los resultados, realizado con apoyo del software ATLAS.ti y respaldado en los estudios de Bardin, se organizaron seis categorías en una matriz que, entre otros aspectos, indican como competencias y habilidades necesarias el razonamiento lógico y analítico en la apropiación de lenguajes de programación y la capacidad de contar historias por vistas interactivas. Los resultados se publicaron en la versión beta del 2030 Data Journalism – cartografía interactiva que pretende acompañar, por medio de geolocalización e informaciones multimedia, las tendencias de esta especialidad en los próximos años.

Palabras clave: Periodismo de datos. Periodismo digital. Visualización de datos. Mapeo interactivo. Universidad.

1 Introduction

One symptom of today's hyper-connected society has brought some challenges to the *praxis* and epistemology of journalism: billions of digital data and information generated by private and public individuals and entities; *big data* made available on digital platforms accessible through the Internet. Journalism, in terms of its ideal function of defending the public interest, has been practiced across a multiplicity of platforms, from traditional or digital media, including contemporary challenges of accelerationism and the (dis) information generated on social networks, algorithms and filters, to ethical questions that surround accessing databases on the *web*.

The data journalism practiced in this current challenging time focuses on a scenario in which discovering, collecting, classifying and using information for verifying, elaborating and visualizing news are reshaped in an intermittent digital environment. Within this

context emerges the need to develop media and information skills and abilities. Not only must the reading public, as producers and consumers of media, be prepared to produce and consume digital content in an ethical manner, but journalism professionals, from the most experienced to those still in training, must also be trained for handling and processing digital data in a more interdisciplinary, fluid, multiplatform and computational journalism.

This study, developed as part of a Master's research project between 2017 and 2018, studied the profile of professional data journalists in Brazil and identified their main skills and competences. This set of skills and abilities is seen as a key element for the practice of data journalism – and that they may change over time due to new demands and technological innovations. This study also proposes a prototype for a communicational product – the beta version of *2030 Data Journalism*, a cartography tool for data journalism – which uses geolocation and multimedia elements to present the information collected in this and future works, thereby acting as a reference for trends in this field over the next few years.

2 Data journalism: theoretical and conceptual synthesis

From the beginning, news and ultimately journalism have always been based on facts and data collected from interviews and public and private statements from official spokespersons taken from research, reports, books and studies. The concept of data adopted in this study is broader than that which is applied to the public databases made available to citizens and the media through Information Access Law n° 12,527 (Brazil, 2011).

The fast pace of life, the variety of platforms, and the convergence of media have transformed the production and consumption of news, thus having a direct effect on the search for information. Jenkins (2009) argues that new media are not replacing old ones, just transforming them. Old media tries to be faster, more transparent and interactive in order to survive. Faced with this scenario, journalists do not necessarily have to multitask, but they do need to be aware of new practices. Saad Corrêa and Silveira (2017) identify what they call "multifaceted communicators", those with skills and abilities that add to the knowledge and practices of journalism, public relations and publicity:

> In addition to having (often personally and directly) skills and competences from areas "outside the field" that come from digitization, such as design, computer science, information science, and others. It is clear that professional journalism, in spite of all the complexities of business, is part of this profile. (Saad Corrêa & Silveira, 2017, 174).

In addition to the public and private databases and data used as the primary source for reporting, there is also data which can be collected from a variety of social media platforms; data that can define audience, consumption and behavior profiles in order to put out certain content.

The digitalization of journalism meant journalists had to incorporate databases into their work (BARBOSA, 2004). For this reason, the concept of data journalism – or Data Driven Journalism (DDJ) – is emerging more and more in the field of journalism studies:

> Data Driven Journalism is produced with data which may be generated and made available by a wide variety of public and private sources – this includes *mainstream* journalistic organizations – and can be structured in a very raw, common form, in Excel spreadsheets, or even published according to *design* patterns and diverse formats of journalistic narratives that make use of varied resources for the readers/users and the public can better understand the material. (Barbosa & Torres, 2013, 153).

Machado (2006) emphasizes that databases structure and organize all stages of journalistic production – from calculating and elaborating content to editing and circulating it: "In short, the functions of databases are as follows: 1) formats for structuring information; 2) support for multimedia narrative models; and 3) the memory of published content" (Machado, 2006, p.16).

According to Träsel (2014), DDJ is a set of practices derived from the tradition of Computer Assisted Reporting (CAR) and Precision Journalism (PJ), which was recreated in the mid-2000s by developments in computing and telecommunications, and the changes they brought to the industrial media model of the twentieth century.

> It is about applying computational and scientific techniques to calculating, editing, publishing and circulating journalistic products, which can be in the form of texts, audiovisuals, hypertextual narratives, graphic displays, or news applications. These techniques are used as a way of increasing journalism's capacity to investigate social events and problems, fulfilling its role of watching over democratic institutions in the name of public interest. (Träsel, 2014, p.15).

Barbosa and Torres (2013) define seven characteristics of data journalism: dynamics, automation, inter-relationship / "hyperlinking", flexibility, information density, thematic diversity, and visualization. In order to work with all these characteristics, professionals have to develop multiple skills in data basing, data processing programs and graphic display.

Bradshaw (2010) believes that all these skills together have the potential to generate more complete, multifaceted, and in-depth histories and narratives. He divides data-based journalistic production into three stages: (1) finding data: this requires knowledge to execute basic techniques in *Computer Assisted Reporting* (CAR) and specific knowledge in data mining and MySQL or Python languages; (2) interrogate data: this requires the journalist to have vast knowledge of the context in which the data are inserted, and also of statistics to find the correct numbers; (3) visualize and combine data: this operation is usually performed by designers and programmers, but is already being performed by many journalists. Based on these three stages, Bradshaw states what is fundamental for data journalists is accuracy and time, these are needed in order to understand and clean information, not to mention knowledge of programming and the traditional art of telling good stories.

In this scenario full of new challenges for the journalist in training, Saad Correa (2015) reflects on the constitution of new curricula for a digitized world: "We are facing a scenario that requires the introduction of different *praxis* in the curricular content of Communication" (Saad Correa, 2015, p.13). This influence from digital technologies has meant some of the classic journalistic concepts like the inverted pyramid and the *lead* have had to be reviewed. In this structure, the news is constructed in layers, from the smaller units of information (connected by *hyperlinks*) taken from raw, unedited data, to the complete packages of traditional journalism, including analysis, unfolding and contextualization.

Figaro, Nonato and Grohmann (2013) warn that multiplatform journalists have another type of relationship related to the time and space of producing information. They take on an even more important role in ensuring that the criteria for newsworthiness include the values that have made journalistic narratives successful in democratic societies. On the other hand, they are challenged by the "instantaneousness of the new media, which focuses on the norms and requirements of the activity and demands changes to how one

manages oneself at work – to operate with new notions of time and space" (Figaro et al., 2013, p. 128).

3 Methodology

Data journalism as an object of study in Brazil has been studied in recent years by researchers and Master's and doctoral studies, and therefore, there are a few theoretical approaches, parameters and suppositions used as a starting point for investigating this form of journalism; prominent examples can be found in the recent works of Bertocchi (2014), Estevanin (2016) and Gerhke (2018). For this study, we chose to collect data from inductive reasoning and non-participant observation, and then subsequently articulated and related that data into analytical categories.

A methodological framework was proposed in the light of the theoretical-conceptual review and the proposed objectives. This framework was divided into two stages. The first stage involves collecting data from open and semi-structured interviews conducted with three different subject groups: (1) market specialists, (2) national media professionals, and (3) researchers from Brazilian universities. The second stage involves discussing categories defined by Bardin's (2004) Content Analysis (CA), supported by ATLAS.ti® (version 8) software.

The first round of open interviews with two experts in the area, held in August 2017, helped us identify hot topics and subject profiles related to the topic. Two groups of interviewees were then selected for a second round of semi-structured interviews. The first is with a group of professionals who work in data journalism in Brazilian media and entities (we call them "the market group"), and the second is with university researchers and professors who develop studies on the subject in Brazil (we call them "the academic group"). The subjects were chosen was based on recognition among their peers and their respective productions (See Appendix).

It is worth mentioning that using semi-structured interviews as a tool for collecting data, in either audio or video format, has proven effective not only for tabulating information transcribed for CA, but also for generating multimedia material for digital cartography prototypes. For this reason, it was necessary to reveal the identity of interviewees who had given prior consent to divulge their names and/or images for use in this study.

When executing the *Word Cruncher* algorithm of ATLAS.ti®, we obtained two lists of the most frequent key words taken from the set of transcribed interviewee reports – one list per group – which made it possible to generate encodings for the interviewees' quotes listed in a preliminary reading. As a result, 48 codes were assigned to the "market group" answers and 49 codes to the "academy group" answers. The correlations between categories and key words of the groups were then evaluated in line with the CA presuppositions.

Six analysis categories were then defined from the preliminary reading of interview data on the theoretical-conceptual review. The categories were examined by two external and approved teachers. After that, we applied the algorithm from semantic networks in the ATLAS.ti® which allowed us to observe the correlations between the codes and each category for the two groups. A matrix of skills and abilities was generated from the analyses and interpretations. It served as a reference for the interactive cartography beta version with information on trends in data journalism in Brazil, in geolocalized format and multimedia.

4 Study subjects and data collection

In the first stage, held in August 2017, two experts in data journalism were interviewed and answered a number of open questions based on the theoretical-conceptual review. The goal here was to raise and identify perceptions on competencies and skills. The first expert was Simon Rogers, current editor of Google News Lab in San Francisco, USA, and creator of the first blog to be published in *The Guardian,* where he worked between 2010 and 2016. The second expert was João Canavilhas, a PhD in Communication from the University of Salamanca (Spain), recognized internationally for his research in the field of transmedia journalism.

The data collected from this first stage was used to develop a new set of semi-structured questions based on the theoretical-conceptual review and also on the information collected from the aforementioned specialists. We then used these semi-structured questions to interview professionals from five national media outlets and agencies, recognized by their peers as pioneers in data journalism in Brazil. These interviews were held between November 2017 and March 2018. The first professional who was interviewed was journalist

Daniel Bramatti, who works for the *mainstream* Estadão Dados[1], a nucleus of the *O Estado de S. Paulo* newspaper specializing in reports on statistics and developing special projects on data display.

The second professional interviewed was journalist Guilherme Duarte, who works at a smaller and more segmented media outlet, *JOTA*[2], a portal which mainly covers legal and political news. His column, "Jota Dados"[3], is described as covering large political and legal decisions, from plenary sessions to behind-the-scenes events, from court victories to competition among offices. The third professional interviewed was journalist and graphic designer Rodolfo Almeida, from the media outlet, *Nexo Jornal*[4]; a digital newspaper created at the end of 2015 which covers main events in Brazil and throughout the world. The newspaper is a reference in data journalism, specializing in infographics, analyses, and columns. Because it was originally created for the Internet, *Nexo* presents relevant issues with innovative language and visualization.

Gustavo Faleiros was the fourth journalist interviewed, and works on the blog *Folha SP Dados*[5], which uses interactive and infographic maps to analyze and visualize information in the articles and reports from both the newspaper and the *website.* The blog works with open data made available by government agencies, universities and independent research institutes. The initiative came about out of a partnership between *Folha* and the *Knight International Journalism Fellowships* program from the International Center for Journalists (ICJ). We interviewed biologist and graphic designer, Daniel Mariani, who publishes data visualizations in the *Folha de S.Paulo* newspaper. We also interviewed journalist Tiago Mali, who specializes in data journalism and is a project and course coordinator for the Brazilian Association of Investigative Journalism (Abraji).

The third stage, held between January and May 2018, used a set of interview questions similar to those applied to the market interviewees. The interviews were conducted with researchers from Brazilian universities, widely cited in literature for their research in data journalism across Brazil. We interviewed three professors from the North and Northeast regions: Professor Walter Teixeira Lima Júnior who, at the time, worked at the Federal University of Amapá (Unifap), coordinating interdisciplinary research in Technology, Communication and Cognitive Sciences; Professor Márcio Carneiro dos Santos, from the Federal University of Maranhão (UFMA), who also works as coordinator of Labcom (Media Convergence Laboratory); and Professor

Suzana Oliveira Barbosa, from the Federal University of Bahia (UFBA), a professor in the Department of Communication (Journalism), the Postgraduate Program in Contemporary Communication and Culture, and Coordinator of the Convergence Journalism Laboratory Project.

We interviewed one professor from the Mid-West, Professor Gerson Martins, from the Federal University of Mato Grosso do Sul (UFMS), a PhD in Journalism (USP) who, at the time, was a postdoctoral researcher in Cyber journalism at the University of Barcelona (Spain). In the south, we interviewed Professor Marcelo Träsel, a researcher and professor of Cyber journalism at the Federal University of Rio Grande do Sul (UFRGS). Lastly, in the Southeast, we interviewed Professor Pollyana Ferrari, a researcher and lecturer at Pontifícia Catholic University of São Paulo (PUC – SP) in Journalism and Multimedia courses at the undergraduate level, and in Digital Intelligence and Design Technologies (TIDD) at the postgraduate level.

5 Categories and matrix of competences and skills

Using codes obtained from the analysis of each interviewee's citations, from the frequency of key words, and from the theoretical-conceptual review, we grouped all the interviewees' information into six categories: (1) differentiated academic curriculum, (2) foundations for programming culture, (3) verification and visualization narratives, (4) projects and partnerships, (5) hybrid teams, and (6) self-directed learning.

Table 01 – Matrix of skills and competences in Brazilian data journalism.

The category...	... shows the need to develop competence in...	... acquired by developing the skill to...
1. differentiated academic curriculum	applying a more logical, analytical and digital-oriented reasoning.	work with Excel spreadsheets, perform models and simulations in Statistics and Mathematics.
2. foundations for programming culture	telling stories and revealing data concealed in *Big Data*. Having independence from programmers.	handle databases and programming languages like Phyton and R.

3. verification and visualization narratives	analyzing complex data extracted from large databases (*Big Data*) and generating compelling digital narratives for online readers.	apply data in graphical languages for the presentation of numbers and results, infographics resources and visualization techniques.
4. projects and partnerships	establishing interdisciplinary partnerships with universities, media outlets and independent associations for the purpose of training data journalists.	understand languages and jargon from different areas, work in different sectors. Seek training as a student. Get trained as a teacher.
5. hybrid teams	being a more complete and experienced journalist with a more analytical view, yet without neglecting the social bias of transparency in news.	work as a programmer, infographic designer and journalist. Interact dynamically with interdisciplinary teams.
6. self-directed learning	being a professional with a broader, interdisciplinary profile and adhering to the demands of a constantly evolving market.	search individually for training and knowledge based on your interests within the digital environment.

Source: the authors.

Differentiated academic curriculum – This category is a consensus among all the interviewees from both the "market group" and the "academic group". There is no classic dichotomy here between theory and practice, which has shown to be the case over the years in journalism practice. Here the interviewee reports point toward a consensus. Both the market and academia recognize that there is a need for new disciplines and curricula which encompass the current scenario of convergence so that journalism students graduate with having had at least minimum training in social and technical aspects of the field, including analytical, logical and technological aspects.

A set of disciplines which are better suited to data journalism is important. Professor Walter Teixeira Lima Júnior gives his thoughts on what an ideal curriculum is:

> In the student's first year, the subject should be Philosophy of information. The next subject is Computational thinking. The third is Theory and interdisciplinary methods, and the fourth is Logic. Philosophy and logic can be included because logic comes from philosophy. When I taught logic and computation, I taught philosophy. The students loved it because they understood it. So, I guess that's it. In the third and fourth year you will say

> things like: Block form, Data Journalism. And some guys will start off really well. It starts at the beginning, the School of Frankfurt. One foot in technology, the other in sociology, and then once he's in his fourth year he will feel isolated, and the guy drops out after four years and is unemployed.

Lima Júnior believes it is important for journalism students to gradually develop skills and logical reasoning; they are being trained as professionals with computational thinking. He also believes that a journalist cannot work in information society without knowing how to do research, conduct surveys, and do basic calculations and visual displays. Speaking on the importance of including more appropriate subjects to a new curriculum, Professor Marcelo Träsel points out:

> As far as I see it, the key is having a course on statistics, or at least incorporating statistics into another course subject. For example, I taught an elective course on statistics in Scientific Journalism. Disciplines like Investigative Journalism, Design, Cyber journalism and Journalistic Production Laboratories would also be natural contexts in which to teach Data Driven Journalism. On the other hand, it is possible to incorporate this type of technique into any practical discipline because the methods for database calculations can be applied not only to digital publications, but also to print journalism, television journalism and radio journalism. However, it is important to keep in mind that not every student has the aptitude or interest to work with data, just as not every student has an interest or aptitude for photography or for public speaking. There should not be too much emphasis given to working with numbers in an undergraduate course.

Professor Träsel argues that the methods for digital verification are also used for print, television and radio journalism, which means that these techniques are not solely taught to professionals who enjoy programming. Not everyone's profile is based around programming, but they do need to understand the basics to work in today's market.

Foundations for programming culture – While the answers given by the "market group" are slightly more in-depth (having a solid basis is important, but it is better to know how to program in both R and Python languages, utilize databases, and have the resources and knowledge for creating displays based on the analyses), the "academic group" deals with the logical, statistical and mathematical basis of reasoning so that students can develop these skills and abilities. This observation is mirrored by Rodolfo Almeida, an infographic designer at Nexo, currently one of the more innovative, completely *web* and technology-based vehicles.

> You have to know how to program. I think it's important for anyone wanting to work with data. You have to know how to use basic statistic software, not because it separates you from the others but because it's essential to be able to carry out research projects or investigations, it has to be something you're interested in, and something you can do on your own, without the help of colleagues or others. I would say this is one of the main competences. It is a kind of newer skill, one that requires journalists to be able to work well with codes because that will give you a little more control over all the stages.

A representative of the Brazilian *mainstream*, infographic designer Daniel Mariani from the *Folha de S.Paulo* supports this idea:

> I believe that knowing how to program is important. You can do good reports using only Excel, I won't say that that is not data journalism, but I have seen better reports on sites that were not programmed. I think programming gives you the freedom to extract data from a huge database, I think you end up being limited if you're not programming. It is important. So, if you know statistics, you should know programming... of course, there are other journalistic parts which are also important.

The interviewees from the "academic group" agree but do point out that some students will always be more interested in this type of narrative than others. Even living in a digital society, not everyone is apt or receptive to learning language programming. For Professor Márcio Carneiro dos Santos, programming is fundamental. He believes it can also serve as a differential when building a narrative and having greater control over how you want your information to be presented to the public.

> So, like I said, even with people from other fields, when producing journalism you should be able to follow and define the processes. Programmers do not make the material, so that's why it is really good having someone who knows these things so they can specify what they want, and sometimes even help out. I study Python programming as a hobby, I don't consider myself a programmer, but doing this gets you out of your comfort zone and forces you to think and structure information in another way, this is good. In addition, this knowledge grants you autonomy in what you're doing, and a kind of exclusivity in your results since few people, or maybe no one, looks at this information in the same way.

Verification and visualization narratives – This third category is based on the importance of skills and improving the quality of content through the use of large amounts of data. Innovative data visualizations, interactive graphics and multimedia features do not make journalists lose their ability to tell stories. In fact, they acquire new tools to make even better and more complete narratives, ones

that interact better with their new readers/users, as pointed out by the interviewees.

For Professor Márcio Carneiro dos Santos, there are three fundamental steps involved in working with narratives in data journalism:

> Being able to find, collect, organize and analyze data sets that are now available to us on the Internet, on transparency portals, on websites and digital archives; then, create stories that this data will support what you are going to write and then do what is essential, listen to people, contextualize everything and finally translate all this into interesting visualizations that help this narrative, because trying to describe numbers using text is a little complicated for readers. So, having knowledge of tools for generating graphics, infographics, etc. helps you complete the process and deliver something interesting.

He also believes that knowing how to use data to create more visual and interactive narratives gives the content value, makes it easier to understand topics that tend to be confusing. Joining infographics, tables, cartography and interactivity together strengthens the testimonies and reports obtained from traditional and official sources.

Rodolfo Almeida, infographic designer from the newspaper *Nexo* adds:

> I am responsible for the graphics that I do. I even write, but it is a kind of different writing because writing infographics is more formal, much more exact than regular text. It's also something someone can write without worrying too much about being a journalist. It is more important to write about what is shown in an exact and correct language, there's no need to use flowery language, no need to spice up the report; this is not how it works. So it is a much more clinical form of writing, more surgical.

Projects and partnerships – This fourth category shows the willingness of both the "market group" and the "academic group" to develop initiatives on data journalism that focus on training students and also those professors who will teach analytics courses in journalism. For Professor Gerson Martins, there are a few specific initiatives.

> The projects started with the new journalism guidelines for 2015, the year the curricula were updated. The courses were Cyber journalism and Digital Journalism. The graduation course at UFMS has a Practical Laboratory with analytic databases which generate the electronic newspaper First News. There is also an electronic magazine coordinated by Suzana Barbosa at UFBA which focuses on qualification and graduation. Our goal is for professionals to be competent at whatever type of reporting they may be doing. Professionals trained at UFMS are completely capable of doing any type of reporting in major city centers. At UFSC we oversee cyber media culture with Raquel Longui; it is based on long-form reporting.

Gerson Martins draws attention to partnerships and projects within the academic environment, without market interaction. However, he has a broad view of what is actually developed in the country. Professor Suzana Barbosa is one of the supporters of this market-academic partnership model.

> There are other organizations also working towards better disseminating, and training and establishing a data school. Abraji itself works with this, we will have an event here at Facom in partnership with Abraji to discuss this, as well as the investigative reports that use data and hidden cameras with TV and print media reporters. We already had a course with the Data School in 2014. It was an initiative that took place here and in Rio de Janeiro with the Data School, and here with the research and online journalism group, which I have been a member of since the beginning of the 2000s. Today I am one of the coordinators, along with Professor Marcos Palácios.

The UFBA is the only institution (among those interviewed) which carries out specific actions in partnership with more market-oriented entities such as Abraji and the Data School. Professor Pollyana Ferrari, in São Paulo, believes that the emergence of *fake news* is conducive to these types of partnerships.

> I think when it comes to fake news, to data checking, it's a promising time. A lot of people have started to realize that we need to get data. I mean, from academia to media publishers. You can no longer rely on the source for checking data. But, people in Brazil confuse things; they check facts with the person's last name. No, it's something else. We are close to an important election, so who knows, right? Big vehicles, big media will be paying more attention. I think there might be a promising movement ahead. Media vehicles are keeping an eye on things, and then they ask academia for help. There is a demand. If it was not 2016, from Trump's election to now, we would be stagnant, not discussing anything. So now I think there's a moment of discussion.

Hybrid teams – This fifth category is another consensus among the interviewees. They all agree that a multidisciplinary team – one comprised of journalists and programmers – can create much more interesting and dynamic narratives for readers. Daniel Mariani, an infographic designer from *Folha de S.Paulo*, says:

> It is very difficult to turn a journalist into a programmer who is able to do everything that someone who has studied it since childhood knows how to do. It is an area that is always changing. There are few examples. I worked with a guy who was a programmer; he studied Engineering and then Journalism. It's not easy to find someone like that. But I think the ideal situation is to find people who know how to program, and even if they don't have any journalistic training, they have an interest in numbers, in discovering phenomena".

This example given by Mariani applies directly to him because he is a trained biologist, and had never worked as a journalist when he started to work with infographics and databases for his Master's in Biology. Today, he is the main infographic designer at *Folha de S.Paulo*, after also having worked for the newspaper, Nexo. In both places of employment he worked in multidisciplinary teams comprised of journalists, programmers and infographic designers. Tiago Mali, coordinator of the courses at Abraji, concludes that there is no single perfect model:

> It can be done in many ways. My former boss used to say that it was easier to train a programmer to be a journalist than a journalist to be a programmer. What's better, in my opinion, is having journalists who have had some training in data, who know how to work with bases, and also have developers and programmers in the newsroom and departments. Because developers will think of things that might not be very interesting from an informative viewpoint. So you can see that some tools, some information, are clearly not elaborated with the reader in mind. They were elaborated for technique or data only.

Self-directed learning – This sixth category is also a skill that professionals who want to work in data journalism nowadays need to have. Rodolfo Almeida, journalist and infographic designer at Nexo, believes that self-directed learning is not something one learns at academia:

> You don't get this from college, I had some design classes in my graduation course, but they were more editorial oriented. You learned about recording, you learned how to write a page without complicated questions. I had class with Pollyana too, but it was a study I was doing on my own. It was something I wanted to do for myself. I was right into it and studied online.

Daniel Bramatti, from the Estadão Dados, thinks a more analytics-based profile is necessary:

> Their profile is more analytical, more eager, and more open to this type of approach. It is a much more objective approach to reality. Maybe you do a report on hunger, fine. You will need to go somewhere where there are people who are suffering from this, you will tell a story, but this story is anecdotal, casual, if it is not supported by data. So you are showing me a certain element, what does it represent? This phenomenon of hunger, how does it manifest over time? Are we losing? Are we winning this battle? It is super important and you can only draw on the numbers. You could present it formally, without the use of any slang, but a database supports it very well. And this profile, we realize that they are people who are taking a step back, instead of looking at the whole forest they are looking at just a single tree, it is important.

Bramatti believes that professionals who seek self-training have a differentiated profile and are able, through the use of these new skills, to understand, relate and present diverging and innovative angles to stories that would otherwise be simple. Infographic designer Daniel Mariani, from the *Folha de S. Paulo,* adds to this:

> I think you need to understand statistics, and when I say understand statistics, I'm not talking about the theory, dull and formal, like in equations, etc. It is more in the sense of understanding statistical biases, not making large statistical mistakes. For example, you say that more women in daycares are victims of aggression, but the daycare has more female teachers, so this result follows logic. It's little things like that. It's not about knowing one's name and the different distributions; it's about knowing what a name is and what is a distribution and what that represents.

Daniel Mariani is defending logical reasoning. A trained biologist, he was a professional who, because of his knowledge of and work with data, programming and visualization, decided to enter the journalism market to tell stories using infographics.

Lastly, the information and expressions that make up the structure of the matrix of skills and competences served as a source and reference to incorporate the interactive cartographic prototype into a website. We are talking about the *beta-test* for the *2030 Data Journalism* project, a digital archive on data journalism which aims to analyze and monitor the evolution of data journalism over the next few years. The cartography is available at the following link: <https://smartmediausers.com/2030-data-journalism>.

6 Final considerations

This article presents the results of a study developed for a Master's degree with the objective of mapping the skills and competences required to perform data journalism in Brazil. Based on all the researched material and the information which was gathered and analyzed, we found information and expressions that supported the structure for the matrix of competencies and skills presented in this article.

The matrix helped towards developing the beta version of the *2030 Data Journalism* – a cartography program that interacts with the academic research and market initiatives addressed in this study. The cartography is expandable and can be used in

future works. We hope that it serves as an interactive memory and research archive for students, teachers and professionals in the field of data journalism.

Lastly, this study highlighted that at a crucial time when a journalist's profession and reputation is so important, data journalism lends more preciseness and correctness to the classic news story. It complements, validates, qualifies and creates a more consistent, friendly, interactive and digital-oriented narrative.

All the interviewees agreed that a journalist with skills and competences in data journalism is one who is able to rationally combine his or her sociological view of the public interest with a more analytical and precise approach, one that is based on data. They also agreed that this knowledge should be disseminated, not only among students and young professionals, but also among university professors who need to be adequately trained in order that they will be able to train new generations of professionals.

NOTES

1   Estadão Dados (http://blog.estadaodados.com/). The nucleus is part of the *O Estado de S. Paulo* newspaper, in publication since 1875 and one of the leading reference newspapers in circulation in Brazil.

2   *JOTA* (https://www.jota.info/). This is a digital press vehicle specializing in legal and political coverage, one of the most influential in Brazil.

3   Jota Dados (https://www.jota.info/opiniao-e-analise/colunas/coluna-do-jota-dados).

4   *Nexo* Jornal (https://www.nexojornal.com.br/). Established in 2015, this is a digital journalism vehicle nationally recognized for its interactive data and multimedia resources.

5   Folha de S.Paulo Dados (http://folhaspdados.blogfolha.uol.com.br). A nucleus of the *Folha de S.Paulo* newspaper, founded in 1921, recognized as one of the leading reference newspapers in circulation in Brazil.

|REFERENCES

Barbosa, S. (2004). Identificando remediações e rupturas no uso de bancos de dados no jornalismo digital. Proceedings from: *2nd Encontro de Pesquisadores em Jornalismo*. Salvador: Agecom. Retrieved from: https://www.facom.ufba.br/jol/pdf/2004_barbosa_remediacoes_rupturas.pdf

Barbosa, S. A. & Torres, V. (2013). O paradigma 'jornalismo digital em base de dados': modos de narrar, formatos e visualização para conteúdos. *Revista Galáxia* 13(25), p.152-164. http://dx.doi.org/10.1590/S1982-25532013000200013

Bardin, L. (2004). *Análise de conteúdo*. Lisboa: Edições 70.

Bertocchi, D. (2014). *Dos dados aos formatos: um modelo teórico para o design do sistema narrativo no jornalismo digital* (PhD Dissertation). Retrieved from: https://danielabertocchi.com/category/papers-thesis-and-more/tese-de-doutorado/

Bradshaw, P. (2010). How to be a data journalist. *The Guardian*, London. Retrieved from: http://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide

Brasil. (2011). *Lei nº 12.527, de 18 de novembro de 2011*. Regula o acesso a informações previsto no inciso XXXIII do art. 5o, no inciso II do § 3o do art. 37 e no § 2o do art. 216 da Constituição Federal. Retrieved from: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

Estevanin, M. (2016). *Processos no Jornalismo Digital: do Big Data à visualização de dados* (Master Thesis). Retrieved from: www.teses.usp.br/teses/disponiveis/27/27152/tde-06032017-094238

Figaro, R., Nonato, C. & Grohmann, R. (2013). *As mudanças no mundo do trabalho do jornalista*. São Paulo: Editora Atlas.

Gerhke, M. (2018). *O uso de fontes documentais no jornalismo guiado por dados* (Master Thesis). Retrieved from: https://www.lume.ufrgs.br/handle/10183/172614

Jenkins, H. (2009). *Cultura da convergência* (2nd. ed). São Paulo: Editora Aleph.

Machado, E. (2006). *O jornalismo digital em base de dados*. Florianópolis: Editora Calandra.

Saad Correa, E. (2015). Centralidade, transversalidade e resiliência: reflexões sobre as três condições da contemporaneidade digital e a epistemologia da Comunicação. Proceedings from: *XIV Congresso Internacional Ibercom*. São Paulo: USP. Retrieved from http://www3.

eca.usp.br/sites/default/files/form/biblioteca/acervo/producao-academica/002736076.pdf

Saad Correa, E.& Silveira, S. (2017). Proposta teórico-metodológica para a pesquisa de objetos no jornalismo. *Matrizes* 11(2), p. 163-182. http://dx.doi.org/10.11606/issn.1982-8160.v11i2p163-182

Träsel, M. (2014). *Entrevistando planilhas: estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil* (PhD Dissertation). Retrieved from: http://repositorio.pucrs.br/dspace/handle/10923/6841

Appendix - Interviewee List

Interviewee	Interview Period (month and year)	Format	Position
Simon Rogers	August 2017	Face-to-face	Market specialist - The Guardian/ Google
João Canavilhas	August 2017	E-mail	Market specialist – Universidade da Beira Interior - Portugal
Daniel Bramatti	July 2017	Face-to-face	National media professional – Estadão Dados
Guilherme Duarte	July 2017	Face-to-face	National media professional – Jota
Tiago Mali	July 2017	Presencial	National media professional - Abraji
Gustavo Faleiros	September 2017	E-mail	National media professional - Infoamazônia / Knight Center
Rodolfo Almeida	October 2017	Skype	National media professional - Jornal Nexo
Daniel Mariani	November 2017	Skype	National media professional - Jornal Folha de São Paulo
Prof. Dr. Walter Teixeira Lima Junior	February 2018	Skype	Academia – UNIFAP / UNIFESP
Profa. Dra. Pollyana Ferrari	March 2018	Face-to-face	Brazilian researcher – PUC/SP

Profa. Dra. Suzana Barbosa	April 2018	Skype	Brazilian researcher – UFBA
Prof. Dr. Gerson Martins	April 2018	Skype	Brazilian researcher – UFMS
Prof. Dr. Marcio Carneiro dos Santos	February 2018	Skype	Brazilian researcher – UFMA
Prof. Dr. Marcelo Trasel	March 2018	E-mail	Brazilian researcher - UFRGS

ANA PAULA BORGES DE OLIVEIRA. Master in Communication and Innovation at the Municipal University of São Caetano do Sul (USCS), São Caetano do Sul, São Paulo. Professor of undergraduate Journalism course at Municipal University of São Caetano do Sul (USCS), São Caetano do Sul, São Paulo, Brazil. E-mail: ana.ana.oliveira@prof.uscs.edu.br.

ALAN CÉSAR BELO ANGELUCI. Conducted Post-Doctoral studies by the Department of Radio-TV-Film of Texas University at Austin, USA and Department of Culture and Information of the School of Communications and Arts of University of São Paulo (USP). PhD from USP with a sandwich period at University of Brighton, England. Professor of the Graduate Program in Communication (PPGCOM) and Education (PPGE) of the Municipal University of São Caetano do Sul (USCS), São Caetano do Sul, São Paulo, Brasil. E-mail: alan.angeluci@prof.uscs.edu.br.

TRANSLATED BY: Lee Sharp