Profile of the data journalist in Spain: training, sources and tools

El perfil del periodista de datos en España: formación, fuentes y herramientas

Diego Arias. Predoctoral researcher in Journalism. He is currently studying a Master's Degree in Professional Multimedia Journalism at Complutense University of Madrid, after having completed a Bachelor's Degree in Journalism at the University of Valladolid. His lines of research are related to the new narrative and professional profiles of Journalism.
Universidad Complutense de Madrid, España
diearias@ucm.es
ORCID: 0000-0003-2023-2530

Pilar Sánchez-García. Professor at the University of Valladolid and a member of the Recognized Research Group (GIR) *Nuevas Tendencias en Comunicación NUTECO* (New Trends in Communication). Professor Sánchez-García is currently the PI of the project entitled, "*Proyecto de investigación aplicada en narrativa periodística multimedia*" C.P. 00200200176 (Applied Research Project in Multimedia Journalistic Narrative). She participates as a researcher in several national projects and is vice president of the Temporary Work Group "Journalism and Communication Education" of the ECREA Association. Her research focuses on professional profiles, multimedia narrative and the evolution of journalistic teaching. She has worked in the field of Journalism for 15 years in various media.
Universidad de Valladolid, España
pilar.sanchez@uva.es
ORCID: 0000-0002-6223-182X

Marta Redondo. Interim contracted PhD Professor of Journalism at the University of Valladolid. She is a member of the Recognized Research Group (GIR) *Nuevas Tendencias en Comunicación NUTECO* (New Trends in Communication), which is dedicated to the analysis of new phenomena related to information science. Professor Redondo is author and co-author of some thirty book chapters and articles published in journals such as *Comunicar, El Profesional de la Información, Revista Latina de Comunicación Social,* and *Comunicación y Sociedad*. Her research focuses on the study of journalistic ethics and deontology, as well as the analysis of infotainment and sensationalist news. She has worked as a journalist in different media, including *Antena 3, Televisión Española* and *Vía Digital,* among others.
Universidad de Valladolid, España
marta.redondo@hmca.uva.es
ORCID: 0000-0001-8143-465X

Abstract:

Data Journalism, the successor of Investigative and Precision Journalism, acquires a new dimension in the Big Data environment that is prompting the redefinition, currently in progress, of a new professional profile: the data journalist. The opening of large databases, the use of new statistical software, and the constant renewal of visualization tools open the door to a journalistic specialization that requires modification of routines and specialized training. This emerging and expanding profile is undergoing uneven development in the Spanish media. With the objective of becoming familiar with its evolution, this research offers the results of an open questionnaire sent to Spanish data journalists that focused the object of study on three categories: routines and training, the use of sources, and visualization tools. The main results seem to indicate that not all media have a specific section of data, that there are training deficiencies remedied by means of self-learning, that they use mostly public sources, and that visualizations are developed with a high level of creative freedom.

Keywords:

Data Journalism; Big Data; professional profile; visualization; technological tool.

Resumen:

El Periodismo de Datos, heredero del Periodismo de Investigación y Precisión, adquiere una nueva dimensión en el entorno del Big Data que empuja a redefinir un nuevo perfil profesional en construcción, el del periodista de datos. La apertura de amplias bases de datos, el empleo de nuevo software estadístico y la constante renovación de herramientas de visualización abren la puerta a una especialización periodística que requiere una modificación de rutinas y una formación especializada. Un perfil emergente y en expansión que ofrece un desarrollo desigual en los medios de comunicación españoles. Con el objetivo de conocer su evolución, esta investigación ofrece los resultados de un cuestionario abierto remitido a periodistas de datos españoles centrando el objeto de estudio en tres categorías: las rutinas y formación, el uso de las fuentes y las herramientas de visualización. Los principales resultados reflejan que no todos los medios cuentan con una sección específica de datos, que existen carencias formativas suplidas de manera autodidacta, realizan un uso mayoritario de fuentes públicas y que las visualizaciones se desarrollan con un alto nivel de libertad creativa.

Palabras Clave:

Periodismo de Datos; Big Data; perfil profesional; visualización; herramienta tecnológica.

1. Introduction

Data Journalism, which is in conformance with the Big Data technology trend (Mayer-Schönberger and Cukier, 2013; McCarthy, 2016), Open Data initiatives (Baack, 2015), and Open Government Data (Shadbolt et al., 2012; Ubaldi, 2013), is progressing as one of the professional journalistic profiles with a very strong future.

The increase in computer tools along with access to new sources is intensifying the investigative work of the journalist, as well as magnifying the need for transparency initiatives (Peiró, 2012). The data, converted into wide sources of information, uncovers a certain reality. Through algorithms (Renó and Renó, 2017), it brings to light a quantitative interpretation of a qualitative reality based on accessibility and reuse of data in order to prevent corrupt practices and to improve the functioning of democratic systems (Antón, 2013; Ferreras, 2013).

Its potential attracts the attention of new markets, innovative business models, and novel consumer and communication trends, especially through the new narratives and the use of networks that promote constant innovation (Dietrich et al., 2012), while at the same time opening up new approaches -frameworks- around the anonymous decisions of the algorithms (Diakopoulos, 2015). Data Journalism is advancing in this context, but with a certain lack of interest by the media, which

can also be attributed to the weakness of an open data culture in most Spanish-speaking countries, differentiating the fact that "visualizing the data is not the same as accessing it" (Crucianelli, 2013: 110).

The popularization and development of the Internet as a source and channel has created a situation in which the volume of information stored on the Internet grows daily (López-García, Toural-Bran, Pereira-Fariña and Barbosa, 2009). However, as a general rule, this huge amount of data in the cloud needs to be treated, analysed and transformed into simpler and clearer publications in a way that is understandable to the general public. The data is converted to interpretive reports and stories typical of Investigative Journalism that goes back to Precision Journalism (Meyer, 2002; Dader, 1997), which has evolved in a digital context of multimedia narratives in which visualization/comprehension is a key factor in offering a clear explanation to the audience of the data obtained (Crucianelli, 2013, Antón, 2013), through the use of infographics, diagrams and interactive applications accompanied by journalistic texts. That is to say, journalism performs the task of 'data literacy' (Gray, Chambers, and Bounegru, 2012) and focuses on a quantitative perspective of professional reality (Coddington, 2015).

Its establishment in the media of different countries is uneven, to the point where some authors continue to mention the idea that we are still 'waiting for Data Journalism' (De-Maeyer et al., 2015), although there are important references in the USA, Europe and Latin America (Ferreras, 2016). In the case of the Spanish media, development has been slow, though it is still considered to be a scarcely exploited specialization (Chaparro, 2013, Ferreras, 2012, 2013, Peiró and Guallar, 2013). From a professional perspective, this is considered a minority discipline due to the fact that although it has experienced a boom phase, its development has stagnated and provides work for a mere 17% of professionals (APM, 2016). In fact, in Spain there are less than twenty media outlets that have specific departments in this discipline (Ferreras, 2016), and they coexist with the impetus of private organizations, with the most well-known case being that of the *Civio* Foundation, among others. However, media that have been born on Internet are the places where this specialization finds the most support (Chaparro, 2014), and in many cases it is integrated into the innovation laboratories of the media (Salaverría, 2015).

Regarding the academic research on Data Journalism, the concept has been approached from different perspectives that could be grouped into three main categories: continuist, rupturist and integrationist. The continuist position considers this not to be a new type of journalism (Rogers, 2013), since journalists have always used data to produce news content, so it can be viewed as information specialization (Chaparro, 2014), seen as an evolution of Precision Journalism. But in the context of media convergence (Jenkins, 2008), driven by the digital and technological environment of Big Data, this is considered a recent discipline that explains why Data Journalism is still a new concept, a 'new type of journalism' (Gertrudis-Casado et. al., 2016), which is in continuous development as a different kind of journalistic discipline because "it works with large volumes of data and uses interactive visualization of such data" (Crucianelli, 2013: 106), and also because the abundance of information differentiates it from Computer-Assisted Reporting (CAR or PAC) (Ferreras, 2013). It is precisely the emergence of Big Data and digital technology that gives it this new dimension (Bradshaw, 2012), which has led to the development of this discipline, understood as a new modality. From this point of view, reference is made to 'data mining' (Riquelme, et al., 2006), which requires different access to the sources and statistical analysis of the data that allows journalists to have immediate access to thousands of files, images and audiovisual material they use to make their information much more elaborate (López et al., 2009).

The third vision, referred to here as the integrationist perspective, considers that Data Journalism is nourished by Investigative Journalism (IJ), by Precision (PJ), and Analytical Journalism (AJ), and is considered to be Deep Journalism (DJ) Assisted by Computers (CAJ). In other words, it constitutes a turning point "between usability (ease of use), design that is oriented toward interaction with the user, infographics, visualization, accessibility, the web, and other technologies" (Antón, 2013: 99). Along the same lines, Gray, Chambers and Bounegru (2012) affirm that this is a type of journalism that allows for the production of news stories that provide added value through techniques of analysis and management of data, as well as others related to presentation and visualization techniques, with the same objective as the old stories, but with sources that allow for the discovery of new ones (Flores, 2012). In other words, Big Data is opening the doors of Data Journalism (López et al., 2016), giving continuity to Precision Journalism (Elías-Pérez 2015), but this has now been transformed into a new open source journalism (SamPedro, 2014).

From this integrationist perspective, the study of the data journalist profile is taken into account, considering that its essential journalistic functions are maintained (Sánchez-García, et al, 2015), yet readjusted to a change of tasks, routines, and training specialization in the use of databases and complex infographics (Cairo, 2012). This is an approach that follows along the lines of precedent studies in Spain, which through various methodologies such as the questionnaire given to experts (Gertrudis-Casado et al., 2016), companies and institutions (Ferreras 2016), confirms the need to define and reinforce skills and specialized training of the data journalist profile that is still in a state of development.

In this context, the main objective of this study is to analyze the professional profile of the data journalist and its implementation in the Spanish media. To this end, three secondary objectives have been established, which in turn have been converted into three main study categories: verifying the professional and educational profile of the data journalist; knowing the predominant and accessible sources and databases; and identifying the most commonly used visualization tools.

With this proposal, the present research ultimately strives to contribute to the enrichment of the object of study in order to update and revise the theory and the results of previous investigations, which allow for the establishment of foundations for the purpose of reflecting upon the current nature of the phenomenon, as well as for measuring its future status.

1.1. Differentiated routines and specialized training of the data journalist

The transformations of journalism in recent years have brought it closer to an interdisciplinary field in which informational and computer skills are required in varying degrees of intensity (Codina, 2016). The data journalist performs an informational news routine that is different from all other journalists in certain tasks that urgently require a search for specialized education, or "reformulated training" (Gertrudis-Casado, et al, 2016). It is essential for this professional profile to be capable of acquiring knowledge in the specialized navigation of large databases, the processing of the information contained therein through software programs, as well as mastery of visualization tools that is transformed into "a functional art" (Cairo , 2011). It is about revitalized skills that converge in a new journalistic profile (Antón, 2013), which requires transversal, general and specific competencies (Gertrudis-Casado, et al., 2016) in the field of analysis, programming and visualization (Ferreras , 2013).

Therefore, it can be argued that the great dilemma of journalism is not so much the incorporation of technology into professional practice as a set of tools, but rather the preparation of professionals with a more technological profile who have the skills and abilities to take advantage of opportunities offered by the computational model (López-García, et al., 2017). This is a specialization that allows for the development of a process of creation based on the compilation, cleaning, contextualization, combination and communication of data (Bradshaw, 2011). Thus, this position requires training in the techniques of search and screening (scraping) of online databases; knowledge of methodology and statistics; and the combining of sources and data from a hybrid web -mashup- (Crucianellli, 2013). Finally, there is the visualization and communication phase that requires the use of maps, charts, infographics and animation that can be combined, or not, with interdisciplinary teams consisting of "developers, designers, statisticians, visualizers and cartographers" (López et al., 2017: 87). This is a profile that performs more efficiently in the hybrid newsroom environment composed of journalists, programmers and designers (Zanchelli and Crucianelli, 2013) with a change of mentality for the purpose of reformulating content and products (Ferreras, 2013). Computer specialists are those who know how to handle data, and journalists are those who know how to transform numbers into fascinating stories that captivate readers (Elías-Pérez, 2015).

In this way, it seems appropriate that the data journalist considers him or herself an expert in searching, ranking and filtering information within the enormous amount of data available on internet, yet at the same time can adapt to the process of continuous training that is a result of the unending technological innovation that affects this profile and should be included in the programs of those universities with Journalistic curricula (Rosique, 2013, Domínguez, 2013). All of this is in response to the need for today's media to innovate quickly without losing the essence of the social and political function of journalism (Orihuela, 2011).

From this theoretical framework, the research presented here begins with two starting hypotheses:

H1   Data Journalism is established in most of the national media analysed through specific departments.

H2. The profile of the data journalist possesses appropriate specialization through training and the use of specific tools and visualizations.

In order to verify or refute the hypotheses, we have used the open questionnaire methodology aimed at data journalism professionals who work in national communication media, and who have become primary sources in this investigation in providing their vision and present experience, which may allow us to indicate the current limitations of this professional profile and future trends as a journalistic discipline.

2. Methodology

This research used a qualitative methodology focused on an organized questionnaire technique by using open questions addressed to professionals with experience in Data Journalism who carry out their work in national communication media of the multimedia format.

Qualitative research is characterized by using multiple "data collection techniques with the main purpose being to obtain information from participants based on perceptions, beliefs, opinions, meanings and attitudes" (Vargas, 2012: 120).

Specifically, the questionnaire is considered an appropriate tool for gaining knowledge about the first-hand view of the professionals active in newsrooms and thus be able to illustrate accurately the reality of this new professional profile through the use of their contributions as primary sources. The questionnaire technique is useful for an object of study as "a set of carefully prepared questions related to facts and aspects of interest in an investigation" (López and Sandoval, 2006: 5). Among the advantages of conducting an open questionnaire is the greater ease in obtaining a response from a group of experts, as they are usually absorbed in their work and have a lack of time, and gathering them together for a meeting is more difficult as it forces them to make trips and find matching schedules (Ruiz, 2012). Some of these limitations have been solved through the use of an online form with open questions that allows the respondents to answer with the maximum freedom of focus, development, vocabulary and terminology (Ruiz, 2012). In its preparation, validity has been taken into account when establishing "concordance between the research objectives and those of the questionnaire" (López and Sandoval, 2006: 5), so that the data represents what is truly intended to be measured by applying the same questions regarding the same topic in order to obtain the same or similar results. In this way, the questionnaire used in this research is structured and composed of open-ended questions that "do not offer fixed-response alternatives", but rather "a space is left so that the subject can structure his or her response" (Canales, 2006: 79). When all of the answers are received, the researcher is in charge of categorizing and systematizing the data obtained so that all of the perspectives and aspects addressed by the experts are organized and structured (Ruiz, 2012). To avoid any type of bias, the researcher must codify the results by establishing a set of categorized answers (Canales, 2006), which in this case are organized into three main categories that demarcate the object of study in order to guarantee validity: the professional profile and its training, sources of work, and technological tools.

2.1. Delimitation of the sample

This research is based on an initial sample of 10 professionals[1] who work as Data Journalists in national media and/or companies that we previously contacted. Nine responded to the first contact, of which two declined to participate in the project, and one was discarded because he or she did not fulfil the requirement of answering all of the questions raised, which could have resulted in a situation of detriment to the validity of the results. Thus, the valid sample consists of six participants, which also reaffirms the difficulty of the researchers in obtaining the collaboration of the professionals who justified this situation by a lack of time. This limitation provides validity at the same time, precisely due to the complexity of obtaining primary sources from professional newsrooms. On the other hand, it is worth mentioning that we have managed to rely on recognized professionals in Data Journalism and to gather the vision and opinion of these six journalists from

---

1 The most frequently used form of contact to request the collaboration of data journalists was through electronic mail, although it must be emphasized that in some circumstances the social network Twitter and even telephone calls have been used. As for communications, an e-mail was sent first to request participation, and then the topic and questions were explained; those who were contacted by Twitter were sent a tweet or a 'Direct Message' to request their email address in order to repeat the previous step. Once contact was made and possible doubts were clarified, the journalists were given a few days to respond due to the large amount of work they had to carry out on those same dates. In addition, communication was re-established on several occasions with those who did not respond, but without success. The questionnaire was sent and completed between May 21 and June 13, 2018. Some participants responded by email and others by telephone through a recorded message, without altering the order of the questionnaire or the systematic and precise collection of answers.

highly regarded media, including Daniele Grasso (*El Confidencial*), Raúl Sánchez (*eldiario.es*), Hugo Garrido (*El Mundo*), Guillermo Villar (*RTVElab*), Ana Isabel Cordobés (*Cuarto Poder*), and Sandra Sánchez (Weblogs).

2.2. Structured open questionnaire

The questionnaire consisted of fifteen questions grouped into the three categories of study mentioned (Table 1): professional profile and training (C1); sources and databases consulted (C2); statistical and visualization tools (C3).

The first category (C1), which focuses on the professional profile, includes six questions about the explanation of the concept of Data Journalism, how the teams are organized, work routines in the newsroom, and whether they rely on the collaboration of multidisciplinary teams with programmers and web designers, among others. In addition, special attention is paid to issues related to the training and management that this discipline should have in universities.

In the second category (C2), we analyse which sources are used to document the information, as well as its accessibility, comprehension, and reuse of the documents, pages, databases or files consulted. In addition, there is also a small reference to Open Data, related to open access and transparency policies. Finally, in the last category of the qualitative questionnaire (C3), five questions are proposed, which focus on determining the most commonly used computer tools for visualizing stories, how they are selected, and the level of training and creative freedom these journalists have.

Table 1. Questionnaire regarding the profile, sources and tools of a data journalist in Spain.

PROFESSIONAL PROFILE (C1)
1 - The concept of Data Journalism is relatively new, so there is no standard definition to describe this discipline with precision. How would you define it?
2 - Do you consider yourself a specialized data journalist or a multi-tasking journalist?
3 - Does your media company have an exclusive department for Data Journalism? Do you think this is common in Spain?
4 - In your Media company, do you work autonomously, or are you part of a multidisciplinary team with programmers and designers?
5 - How do you assess the training of data journalists in our country?
6 - Do you believe this discipline should have at least one compulsory subject in university curricula?
ACCESS AND USE OF SOURCES (C2)
7 - What type of sources do you commonly use in developing your news information?
8 - How do you gain access to these sources?
9 - On a scale of 1-10, how do you rate the information transparency of the institutions in Spain? Explain why you believe this is the case.
10 - Are the raw data you access understandable and reusable?

STATISTICAL AND VISUALIZATION TOOLS (C3)
11 - What tools do you most often use in your work?
12 - Do you choose them, or is chosen by the company?
13 - Given the constant advancement of these tools, how do you learn to manage them: through training or by self-education?
14 -Do you think a specific profile of an infographic specialist/designer is required for the visualization of data, or is this an activity that a journalist can carry out?
15 - Are there any narrative guidelines set by the company when converting graphics into news?

Source: author-created

3. Results

The results presented below are grouped into the three categories of study and are highlighted by the role of each professional participant in the sample. The answers are completed at the end with a comparison of the results that allow for the examination of similarities and differences of the data journalist profile from the perspective of working professionals. At the end of each section, several representative graphs of all the answers obtained in the corresponding category are shown for the purpose of extracting from the self-administered questionnaires the main ideas according to the objectives set out in the research. Simplification of the answers does not pursue an objective of quantitative representation, but rather an interpretation of compared qualitative answers in the search for concordances or predominant differences by using a table that gathers the perspectives of the participating professionals.

3.1. Professional profile of the data journalist

The first category of analysis focuses on the profile of the data journalist (C1), covering, as already specified, concept, work routine, and education by using six questions (1-6), and the responses are reflected individually (Graphic 1). The first question deals with the relatively recent label of 'Data Journalism', as some authors maintain that there is not yet a standard definition to describe it. In this sense, Cordobés defines the discipline as "a system for generating information, based on data analysis, for the purpose of spreading news and information". Garrido takes the same posture in pointing out that it is "a way to generate information through the analysis of large compilations of data that is presented in the form of articles, visualizations and applications". Other authors, such as Raúl Sánchez, state that it does not differ much from traditional journalism, though "it is more linked to investigative journalism", but rather, "it is about applying the extraction, analysis and visualization of data to traditional journalism to make bigger and better stories". On the other hand, Grasso quotes his French counterpart, Nicolas Kayser-Bril, and affirms that it is "journalism that measures what is not measured". Sandra Sanchez labels it as "a journalistic investigation that tries to find in a set of data a pattern that is susceptible to becoming a newsworthy story", and emphasizes the fundamental need to differentiate data journalism from journalism with data (traditional).

Finally, Villar maintains that Data Journalism is the discipline responsible for "finding the clue to a concealed story from the analysis and cross-referencing of databases" and tell it in depth through the writing and visualization of data. However, he also explains that referring only to the analysis, extraction and cleaning of data relegates the journalist exclusively to that task and causes him or her not to be taken into account as a key professional throughout the entire process.

Regarding the second question, the focus is on how the professionals would be classified according to the functions they perform in their area. In this case, five of the six participants defined themselves as specialized data journalists, though one of them believes that the figure of a 'multitasking journalist' is a more appropriate term to describe the work. With regard to when they began to work in this discipline, all respondents answered with dates after 2010 (except for one of the journalists who did not specify her beginning). The first to start in this work was Garrido (2012), followed by Grasso and Raúl Sánchez (2013), Villar (2015) and Sandra Sánchez (2016).

Regarding the organization of the newsrooms of each media, they were asked if they have a specific department of Data Journalism, and the answers reflect a diverse panorama, divided at 50%. Thus, *El Confidencial*, *eldiario.es*, and *El Mundo* have exclusive departments to develop this activity, while the other three media do not. In addition, another important element is that most of the participants, five out of six, believe it is not common in this country, compared to one who thinks it is common.

With regard to work routines, it cannot be said that there is a predominant type of organization, since half of them are divided between the ones who work autonomously with regard to the rest of the newsroom team and the other half forms part of a multidisciplinary team with programmers and designers (such as *El Confidencial, El Mundo* and *RTVELab*).

In the last two questions of the professional profile section, they were asked how they valued the education of the data journalist in Bachelor's and Master's degree programs. Only Cordobés defended the training as very good, while Raúl Sánchez and Villar qualified it as good. In contrast, Grasso, Garrido and Sandra Sanchez judged it to be mediocre or improveable. On the question of whether a compulsory subject should be included in the Bachelor's syllabus, four of the six interviewees were inclined toward an affirmative answer, while only two believed it was not necessary.

Graph 1. Grouped answers related to the data journalist profile (C1)
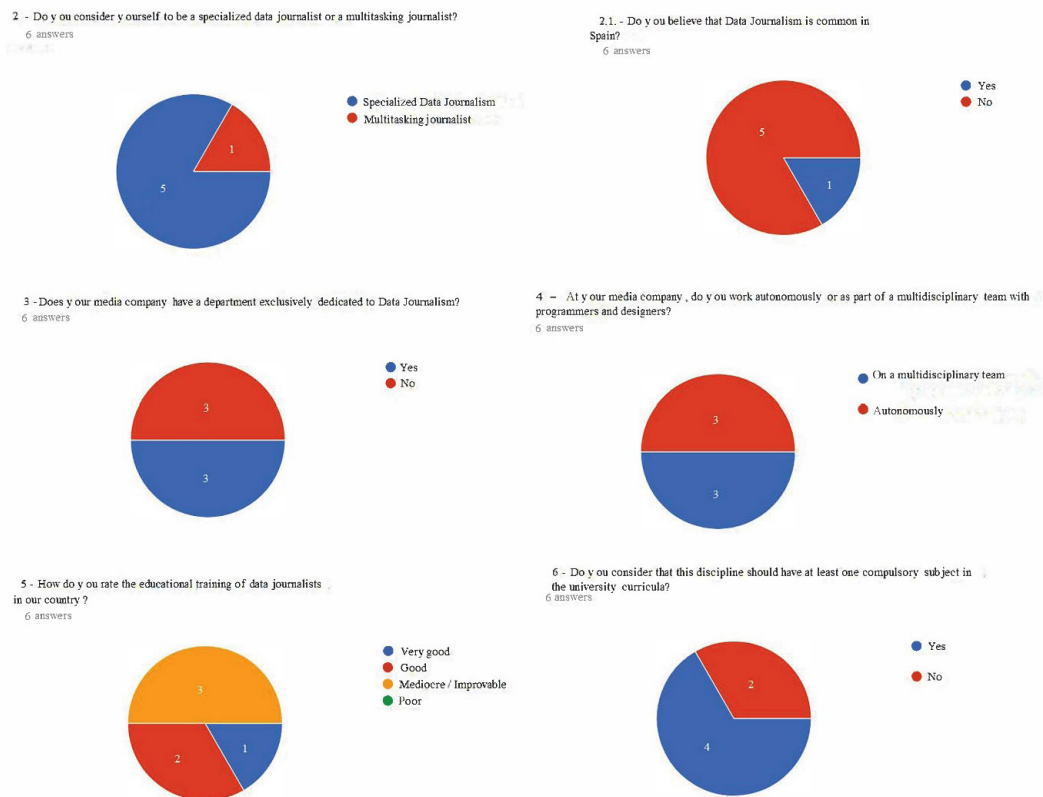
Source: created by the authors

3.2. Access and use of sources

Next, we proceed to a detailed breakdown of the issues related to the second category (C2), which groups together four questions (7-10) concerning the documentary sources and main databases used by the journalists consulted (Graph 2).

In the seventh question, reference was made to the type of sources that are most frequently used by respondents in developing their information. All of them use official public sources to a greater extent, although some diligently carry out requests for information access (2) or use leaks (1). However, none of them mentioned the use of sources from private
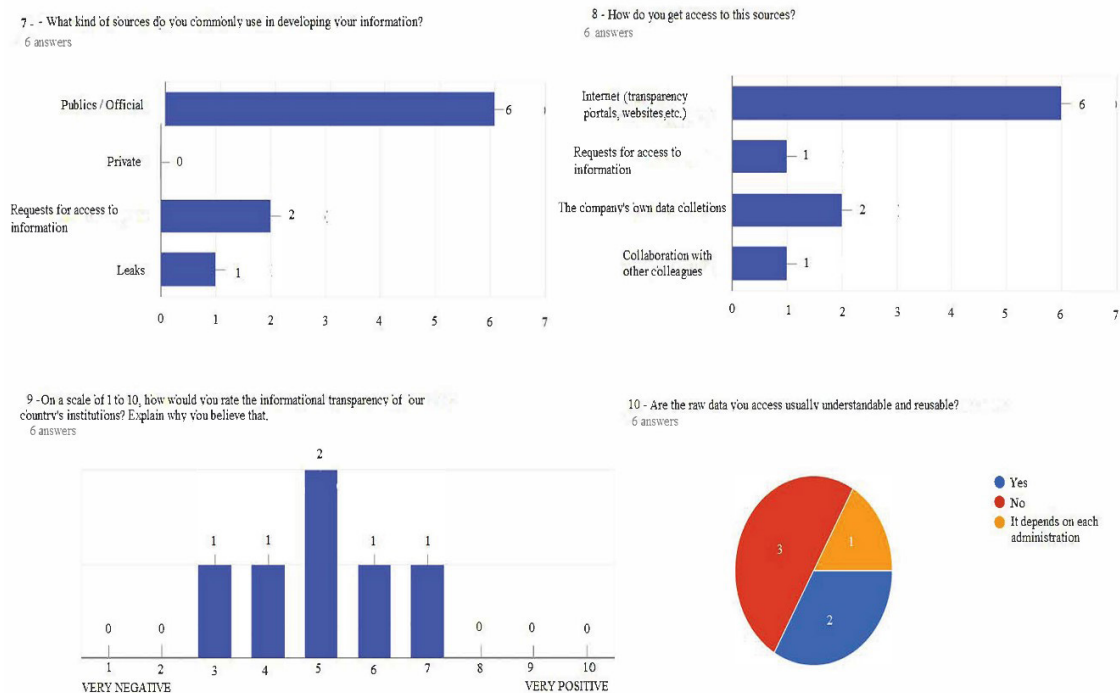
entities. Regarding how to gain access, there was complete unanimity in the use of Internet, or in other words, they focus especially on consulting transparency portals, websites and open databases.

Likewise, in two cases the use of their own data collections was mentioned, one of which was related to requests for access to information, and the other was for databases offered by peers or other sources.

Moreover, in reference to Open Data and Open Government Data, a question regarding how information professionals value information transparency in Spain on a scale of 1 to 10 was included in the investigation. The scores oscillated between 3 and 7, with an average of 5. However, many of the professionals explained their scores, which have been added here because we consider them to be of interest. Cordobés explains that "there is still much to be done", and Grasso clarifies that "it depends a lot on each administration, while stressing that the situation "has progressed a lot thanks to the Transparency Law", because the scores would have gone down to 3 or a 4 if this law had not been passed. Garrido clarifies that "it depends on the culture of that organization in this regard, but "there is still a very patrimonial concept of the information that is generated by these administrations". Villar also stresses that "there is no culture in public administration of giving requested databases directly" and "much less of generating it". On the same issue, Raúl Sánchez confirms the importance of the Transparency Law, but notes that "there is not yet a culture of transparency among officials and senior public sector officials". Sandra Sánchez differentiates between transparency at the municipal level, as they tend to be the ones "who do it best", while at the national level the mark is "unsatisfactory", and "there is still a lot of room for improvement". These reflections confirm the slow development mentioned within the data culture.

Regarding the question of whether it is possible to reuse and comprehend the raw data accessed, different responses appeared: 'no' for three of the respondents, while two said 'yes', and one said 'it depends on each administration'. Raúl Sánchez pointed out that "not even the law obliges administrations to publish data in reusable formats", while Garrido decried that "it is still very common to find administrations that scan documents prior to their release to make something as simple as searching for text difficult". Sandra Sánchez and Garrido clarified that it depends on who generated the data and how it was generated, though in general they judge the data to be understandable. Grasso agrees with the idea, but stresses "it is not always true that they are accessible". Villar states that the data "to a greater or lesser extent needs to be cleaned".

Graph 2. Grouped answers on the use of sources and databases.

7 .- What kind of sources do you commonly use in developing your information?
6 answers

8 - How do you get access to this sources?
6 answers

9 -On a scale of 1 to 10, how would you rate the informational transparency of our country's institutions? Explain why you believe that.
6 answers

10 - Are the raw data you access usually understandable and reusable?
6 answers

Source: created by the authors

3.3. Statistical and visualization tools

The last section of the questionnaire focuses on queries related to the third category of this study (C3) on the use of visualization tools (Figure 3). To start the analysis of this section, it is important to clarify which tools are used most in the media.

Almost all of the journalists consulted agreed that the use of spreadsheets (Excel) is basic and fundamental for managing databases, so this is a tool that must never be missing. In addition, they point to the most popular tools being DataWrapper (5), Tableau (4) and CartoDB (3).

In a smaller number of responses, R, D3 Refine and Flourish (2) also appear. Other tools mentioned by at least one of the professionals are SQL, Minitab, Kimono, Datascript, Tabula, Abbyy, Fine Reader, SPSS, Adobe Illustrator, Photoshop,

Phyton, Infogram and Open Refine, among others. Regarding the choice of tools, five of the six respondents say that they are the ones who select them, although one respondent indicated that he or she works with those that are most commonly used in the company, although it depends on the case, and at no time does the company limit his or her freedom. Therefore, in this regard it must be emphasized that data journalists have the freedom to select the tools they use when developing their activity, and that their choice is influenced by the kind of topic or publication. Aware that digital tools for the treatment of data are constantly evolving as new software and technologies are developed, the question is raised regarding the training they receive in this respect. Moreover, the professionals agreed that they learn in a self-taught way, and in some cases through online courses. The need for specialization in this discipline is driving continuous training of professionals when faced with these new statistical and visualization tools. Furthermore, this need for specialization is also related to the question of whether there should be a specific profile of an infographics specialist or designer to create the data visualizations.

The general view of respondents on this issue is inclined toward a negative response, though with nuances. Three of the six respondents stated it is not necessary, though advisable, while two believed that the journalist can easily take charge of this task. Even Raúl Sánchez categorically denies that the figure of the infographics specialist is necessary, since for him "visualizing data is not designing", but rather "showing information in the most appropriate way so that the human eye understands it better". He believes that everything else "is embellishment".

The last question on the form was about the freedom of data journalists to carry out their work and inquired as to whether there are any narrative guidelines set by the company in terms of converting stories into graphics. All six of the professionals emphasized that there are no narrative guidelines set by the company, and some even pointed out that such graphics are created through their own initiative. Thus, with this data obtained, it can also be confirmed that professionals in this expanding journalistic profession have great freedom in their work.

**Graph 3. Grouped responses on the use of visualization tools (C3)**

11 - Which tools do you most often use in your work?
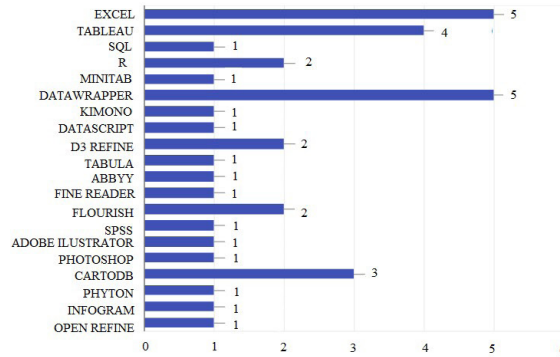
12 -Do you choose the tools yourself, or are they chosen by the company?

- Own choice
- Chosen by the company
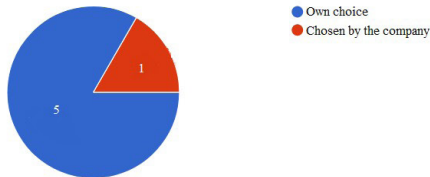
13 – Faced with the continious evolution of tools, how do you learn to manage them: through training or by self-taught learning?

- Self-taught way
- Through training
- Both
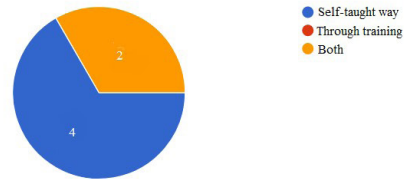
14 - For the visualization of data, do you believe a specific profile of an infographics specialist/designer is necessary, or is this an activity that a journalist can carry out?

- Yes, a specialist is necessary
- No, but a specialist is recommendable
- No, a journalist can do it
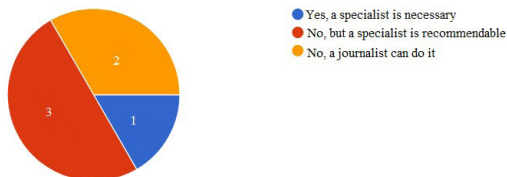
15 - Are the narrative guidelines set by the company when converting graphics into stories?

- Yes
- No
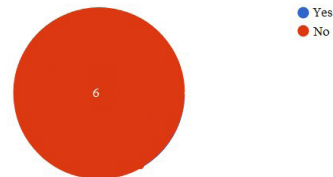
Source: created by the authors

3.4. Comparison of answers among data journalists

Taking into account the qualitative nature of this research, shown below are the individualized results of each participant of the sample, which are grouped around the three categories that focus on the object of study. To do this, we have resorted to simplifying the answers by using key words and ideas that make it possible to establish an overall comparison in the questionnaire. Therefore, in each section a comparative sub-section of all the answers is shown in order to extract the main ideas from the self-administered questionnaires according to the objectives set out in the research. Visualization, in spite of its simplification, is considered useful in obtaining an overview with the main ideas that allow us to compare convergent or divergent points of view of the entire sample.

Table 2 - Comparison of the data journalists' responses grouped by category.

	C1 - ANSWERS RELATED TO THE PROFESSIONAL PROFILE					
JOURNALIST	**Ana Isabel Cordobés**	**Daniele Grasso**	**Raúl Sánchez**	**Hugo Garrido**	**Sandra Sánchez**	**Guillermo Villar**
MEDIA	*Cuarto Poder*	*El Confidencial*	*Eldiario.es*	*El Mundo*	*Weblogs*	*RTVElab**
Definition of Data Journalism	Information-generating system	Journalism that measures what is not measured	Traditional journalism with data analysis	Generates information through Big Data analysis	Research that tries to look for patterns in data	To find a concealed story through analysis and cross-referencing of databases
Specialized or multitasking data journalist	Multitasking data journalist	Specialized data journalist	Specialized data journalist	Specialized data journalist	Specialized data journalist	Specialized data journalist
Start date	Not specified	2013	2013	2012	2016	2015
Does the media have its own department?	No	Yes	Yes	Yes	No	No
Is this common in Spain?	No	No	Yes	No	No	No
Do you work autonomously or on a multidisciplinary team?	Autonomously	Multi- disciplinary team.	Autonomously	Multi-disciplinary team	Autonomously	Multi- disciplinary team
Assessment of education related to the data journalist	Very good	Mediocre/ Improvable	Good	Mediocre/ Improvable	Mediocre/	Good
Is it necessary to have a specific subject in the Degree?	Yes	Yes	No	Yes	No	No

C2 - ANSWERS RELATED TO ACCESS AND USE OF SOURCES						
Most-used sources	Official / Public (transparency portals)	Public, Leaks, Requests for access to information	Public/ Official	Public/ Official, Own data collection	Public/ Official	Public/ Official, Requests for access to information
Access	Internet (transparency portals, websites, etc.)	Internet, Requests, Collaboration of other journalists	Internet (transparency portals, websites, etc.)	Internet, Requests for access to information, own data collection	Internet through scraping: websites, transparency	Internet through scraping: websites, transparency portals
Information transparency rating (1-10)	5	6	7	5	3	4
Can the raw data be reused and understood?	Yes	It depends on each administration, but not always	Non-reusable formats predominate	It depends on each administration, but generally not	Yes, although it depends on the topic	No, the data requires cleaning
C3 - ANSWERS RELATED TO VISUALIZATION TOOLS						
Most commonly-used tools	*Tableau, Infogram, Datawrapper*	*Excel, Tableau, R, SLQ*	*Excel, SPSS, Tableau, Datawrapper, Flourish*	*Excel, CartoDB, Datawrapper, Flourish, Tableau*	*Excel, OpenRefine, Datawrapper, CartoDB*	*Excel, D3 Refine, CartoDB, Datawrapper*
Was this your own choice or was it the company's choice?	Own choice	Own choice	Own choice	Own choice	Own choice	The company's, but with freedom
Self-taught learning or online training?	Self-taught, but sometimes online training	Self-taught learning	Self-taught learning	Self-taught learning	Self-taught learning and online training	Self-taught learning and online training
Is a specific infographics profile advisable, or can a journalist do the job?	A specific profile is advisable	A specific profile is advisable	A journalist can do it	A specific profile is advisable	A specific profile is advisable, but a journalist can do it too	A specific profile is better
Are the narrative guidelines set by the company?	No	No	No	No	No	No

*Villar's answers are related to his previous period at RTVELab.

Source: author-created

4. Discussion and Conclusions

Both the main objective as well as secondary goals of this research have been attained through coming closer to a description and analysis of the profile of data journalist in multimedia communication newsrooms in Spain from several

points of view: the organization of editorial departments, training of data journalists and their access to sources, as well as use of the most effective visualization tools.

Regarding the confirmation or refutation of the initial hypothesis (H1) related to the issue of whether Data Journalism is established in specific departments within national newsrooms, this cannot be confirmed in its entirety since the answers do not demonstrate clearly that such an organization is prevalent in the media. The results show that half of the media consulted have a specific department for Data Journalism, yet at the same time, five of the six journalists interviewed say that this is not common in Spain. In this way, the idea that this is a discipline and profile with uneven development in the media is confirmed, and it is going through a process of adaptation, as the Madrid Press Association (*APM*, 2016) has already pointed out. Its recent implementation can be added to this reality due to the fact that all of the respondents started their activity from 2012 onward.

The second hypothesis (H2) related to the adequate specialization of the data journalist profile through training and the use of tools has been refuted in part because most professionals explain that the training offered is mediocre or improvable, a deficiency that they counteract through self-education. On the other hand, it has been confirmed that this specialization is acquired through knowledge of data processing and visualization tools that ensure correct utilization with a high degree of autonomy in all cases. This idea confirms the training deficiencies detected in previous studies (Gertrudis-Casado et al., 2016), especially in the field of statistical treatment of data and its visualization.

Once the objectives and hypotheses have been verified, a series of specific conclusions are obtained from the three main categories of study: professional profile and training; sources consulted and access to databases; statistical and visualization tools.

Regarding the consideration of their own professional profile (C1), most of the respondents consider themselves to be specialized data journalists rather than multitasking journalists.

The general opinion as to whether the person in charge of visualizations has to be from a specific designer profile was negated by half of the respondents, though they acknowledged that it is advisable, and two respondents explained that a journalist can carry out this responsibility perfectly if visualization is not considered design. Only one professional defends the position that a specific profile in design should be required. In general, they consider that the training of this profile should be introduced in Universities and they affirm that their own training has been self-taught.

With regard to journalistic sources (C2), there is a predominance of public and official sources such as transparency portals, official websites and public databases. Requests for access to information and leaks are less common. However, although these are less frequent, they have a greater presence in newsrooms than private sources. In this sense, the preferred way of accessing sources is through autonomous searches on Internet, followed by the use of an organization's own data collections, requests for access to official information, and collaboration from sources or other colleagues in the profession. Although it is not the central objective of this investigation, the respondents' negative opinion about the transparency of information in Spain is worth noting

In fact, the average of all the ratings assigned by the journalists consulted was only a passing mark of 5 out of 10. In addition, the majority criticized the fact that in general terms, it is not possible to reuse and understand raw data. This response was

in agreement with the idea that "there is no open data culture in most Spanish-speaking countries" (Crucianelli, 2013: 109) due to official gaps in terms of transparency and reuse of public data. In fact, this limitation of access to data and its non-reusable formats is considered one of the barriers to further development of Data Journalism in Spain (Ferreras, 2013).

Regarding statistical and visualization tools (C3), the respondents agreed that the use of spreadsheets (Excel) is fundamental in managing databases, and therefore is an essential tool.

The most commonly mentioned were DataWrapper, Tableau and CartoDB. They confirm that the journalists themselves select and set up these tools, since they are the ones who have to work with them, and the company trusts their choices without establishing any specific narrative guidelines.

Regarding training in the use of the tools, with the exception of two of the participants, most reiterated that they learned to use the tools through self-education because within the media, specific training courses or workshops is not prioritized, but rather they learn how to use these tools by themselves simply through hands-on use or by collaborating with other colleagues.

In conclusion, despite the constraints of the limited sample, this research confirms the results of previous studies that have indicated uneven development of the data journalist profile in those media that do not have their own departments for this function, which indicates disinterest and difficulties in accessing data. This is a profile that works with a high level of freedom and autonomy and that requires clearer organization regarding hybrid or multidisciplinary work teams. Finally, it has been noted that the teaching of this discipline in Spain still has considerable room for improvement, and consequently it appears necessary to reinforce this profile in undergraduate degrees and in specific master's programmes.

For future research, it would be necessary to study in more depth the evolution of Data Journalism in Spain and its integration into newsrooms, as well as specific educational changes of a journalistic profile that is still being developed and has a solid future.

5. Bibliographic references

Antón, A. (2013): "El periodismo de datos y la web semántica", *Cuadernos de Información y Comunicación (CIC)*, 18, pp. 99-116. Recuperado de: https://goo.gl/M7zwaF [Consultado el 11/04/2018].

Baack, S. (2015): "Datafication and Empowerment: How the Open Data Movement Re-articulates Notions of Democracy, Participation, and Journalism", *Big Data & Society*, 2(2). DOI: http://dx.doi.org/ -10.1177/2053951715594634

Bradshaw, P. (2011): "The inverted pyramid of data journalism", *Online Journalism Blog*. Recuperado de: https://goo.gl/4xWJiS [Consultado el 16/04/2016].

Bradshaw, P. (2012): "What is Data Journalism?", en Gray, Jonathan Bounegru, Liliana and Chambers, Lucy (Eds.): *The Data Journalism Handbook*. Reino Unido, European Journalism Centre and the O'Reilly Media.

Cairo, A. (2011): *El Arte Funcional: infografía y visualización de la información*. Madrid, España: Alamut Ediciones.

Cairo, A. (2012): "Infographics and Visualizations as tools for the mind", en *The Functional Art* (thefunctionalart.com). Recuperado de: https://goo.gl/htpZA8 [Consultado el 16/04/2018].

Canales, M. Coord. (2006): *Metodologías de investigación social: introducción a los oficios*. Santiago de Chile: Editorial LOM. Recuperado de: https://bit.ly/2LWPiYp [Consultado el 15/05/2018].

Coddington, Mark (2015): "Clarifying Journalism's Quantitative Turn. A typology for evaluating data journalism, computational journalism and computer assisted reporting", *Digital Journalism*, 3, 331-348. Recuperado de: https://doi.org/10.1080/21670811.2014.976400 [Consultado el 23/11/ 2018].

Codina, LL. (2016): "Tres dimensiones del periodismo computacional. Intersecciones con las ciencias de la documentación", *Anuario ThinkEPI*, 10, 200-202. Recuperado de: https://goo.gl/U12TwH [Consultado el 18/04/2018].

Crucianelli, S. (2013): "¿Qué es el periodismo de datos?", *Cuadernos de periodistas*, 26. Revista de la Asociación de Prensa de Madrid, pp. 106-124. Madrid, España: Departamento de Comunicación, Publicaciones y Actividades. Recuperado de: https://goo.gl/igiF3R [Consultado el 16/04/2018].

Chaparro, M. A. (2013): "La evolución del periodismo de precisión: el blog de The Guardian sobre periodismo de datos", en F. Segado (dir. Congre), I Congreso Internacional de Comunicación y Sociedad Digital. Logroño: España. Recuperado de: https://goo.gl/KNbhj7 [Consultado el 12/04/2018].

Chaparro, M. A. (2014): "Nuevas formas informativas: el periodismo de datos y su enseñanza en el contexto universitario", *Historia y Comunicación Social*, 19, (nº Esp. Febrero 2014), pp. 43-54. Recuperado de: https://goo.gl/mdW8Nk [Consultado el 14/04/2018].

Dader, J. L. (1997): *Periodismo de Precisión: Vía socioinformática de descubrir noticias*. Madrid: Síntesis.

De-Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F.; y LeCam, F. (2015): "Waiting for Data Journalism: A Qualitative Assessment of the Anecdotal Take-up of Data Journalism in French-speaking, Belgium", *Digital Journalism*, 3(3), 432-446. Doi: http://dx. doi. -org/10.1080/21670811.2014.976415

Diakopoulos, Nicholas (2015): "Algorithmic Accountability, Digital Journalism", 3, 3, 398-415. Recuperado de: DOI: 10.1080/21670811.2014.976411 [Consultado el 26/11/ 2018].

Dietrich, D.; Gray, J.; Mcnamara, T.; Poikola, A.; Pollock, R.; Tait, J.; y Zijlstra, T. (2012): *Open Data Handbook. Open Knowledge Foundation*. Recuperado de: https://goo.gl/qJJDgy [Consultado el 15/04/2018].

Domínguez, E. (2013): "¿Dónde está el periodismo de datos en la formación periodística?", *Revista de los Estudios de Ciencias de la Información y de la Comunicación*, n. 22. Recuperado de: https://goo.gl/QvA5tN [Consultado el 24/05/2018].

Elías-Pérez, C. (2015): *Big data y periodismo en la sociedad red*. Madrid, España: Síntesis.

Ferreras, E. (2012): "Nuevos perfiles profesionales: el Periodista de Datos", en Sociedad Latina de Comunicación Social (SLCS), actas IV Congreso Internacional Latina de Comunicación Social. Universidad de La Laguna. Recuperado de: https://goo.gl/D3BteC [Consultado el 15/04/2018].

Ferreras, E. M. (2013): "Aproximación teórica al perfil profesional del 'Periodista de datos'", *Icono 14*, 11, n. 2, pp. 115-140. DOI: 10.7195/ri14.v11i2.573 [Consultado el 12/04/2018].

Ferreras, E. M. (2016): "El periodismo de datos en España", *Estudios del Mensaje Periodístico*. 22,1, pp. 255-272.

Flores, J. (2012): "Sinergias en la construcción del Nuevo Periodismo derivadas del Data Journalism y el Transmedia Journalism", en Actas III Congreso Internacional Comunicación 3.0, pp. 476-487, Universidad de Salamanca: Salamanca, 10 y 11 de octubre de 2012. Recuperado de: https://goo.gl/9REC4e [Consultado el 11/04/2018].

Gertrudis-Casado, M.C.; Gértrudix, M. y Álvarez, S. (2016): "Competencias informativas profesionales y datos abiertos. Retos para el empoderamiento ciudadano y el cambio social", *Comunicar*, 47, pp. 39-47. Recuperado de: https://doi.org/10.3916/C47-2016-04 [Consultado el 09/09/2018].

Gray, J., Chambers, L., y Bounegru, L. (2012): *The Data Journalism Handbook*. O'Reilly Media. Recuperado de: http://datajournalismhandbook.org/1.0/en/ [Conusltado el 12/10/2018].

Jenkins, H. (2008): *Convergence Culture: la cultura de la convergencia de los medios de comunicación*. Barcelona: Paidós.

López, N. y Sandoval, I. (2006): "Métodos y técnicas de investigación cuantitativa y cualitativa". Documento de trabajo, Sistema de Universidad Virtual, Universidad de Guadalajara. Recuperado de: https://goo.gl/6JpYNt [Consultado el 24/05/2018].

López, X. (2012): "La educación de los periodistas para los entornos digitales actuales", *Revista de Comunicación*, n. 11, pp. 178-195. Recuperado de: https://goo.gl/z4iuk4 [Consultado el 11/04/2018].

López, X.; Toural-Bran, C.; y Rodríguez-Vázquez, A. (2016): "Software, estadística y gestión de bases de datos en el perfil del periodista de datos", *El profesional de la información*, 25, n. 2, pp. 286-294. Recuperado de: https://goo.gl/FRvZFK [Consultado el 12/04/2018].

López-García, X.; Rodríguez-Vázquez, A. I.; y Pereira-Fariña, X. (2017): "Competencias tecnológicas y nuevos perfiles profesionales: desafíos del periodismo actual", *Comunicar*, XXV, n. 53, pp. 81-90. DOI: 10.3916/C53-2017-08 [Consultado el 14/04/2018].

López-García, X.; Toural-Bran, C.; Pereira-Fariña, X.; y Barbosa, S. (2009): "Automatización de las bases de datos: potencialidades de herramientas básicas para otro periodismo posible", *El profesional de la información*, 18, n. 3, pp. 308-315. Recuperado de: https://goo.gl/yV4FJG [Consultado el 14/04/2018].

Mayer-Schönberger, V., y Cukier, K., (2013): *Big Data. La revolución de los datos masivos*, Madrid, España: Editorial Turner Noema.

McCarthy, M. (2016): "The big data divide and its consequences", *Sociology Compass*, 10(12), pp.1131-40. DOI: 10.1111/soc4.12436[Conuslktado el 02/09/2018].

Meyer, P. (2002): *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Maryland, Estados Unidos: Rowman & Littlefield Publishers, Inc. [4ª edición]

Orihuela, J. L. (2011): *80 claves sobre el futuro del periodismo*. Madrid, España: Anaya Multimedia: 800 Books. Introducción recuperada de: https://goo.gl/RPUMZx [Consultado el 21/05/2018].

Peiró, K. (2012): "Data Journalism: un enorme mundo por explorar". En karmapeiro.wordpress.com (blog personal). Recuperado de: https://goo.gl/F88Ne2 [Consultado el 14/04/2018].

Renó, D. y Renó, L. (2017): "Algoritmo y noticia de datos como el futuro del periodismo transmediaimagético", *Revista Latina de Comunicación Social*, 72, pp. 1.468 a 1.482. DOI: 10.4185/RLCS-2017-1229 [Consultado el 03/07/2018].

Riquelme Santos, J.C., Ruíz, R. y Gilbert, K. (2006): "Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial", *Revista Iberoamericana de Inteligencia Artificial*, 10 (29), 11-18.

Rogers Simon (2013): *Facts are Sacred: The power of data. Guardian Shorts*. Ed: Faber & Faber.

Rosique, G. (2013): "Los estudios universitarios de periodismo en España: la adaptación al EEES y la formación de periodistas en competencias digitales", *Riesed, Revista Internacional de Estudios sobre Sistemas Educativos*, 1, n. 1-2, pp. 117-132. Recuperado de: https://goo.gl/qjKyNC [Consultado el 22/05/2018] .

Ruiz, J. I. (2012): *Metodología de la investigación cualitativa*. Bilbao, España: Universidad de Deusto. Recuperado de: https://goo.gl/6ZHrWt [Consultado el 01/06/2018].

Sampedro, V. (2014): *El cuarto poder en Red: por un periodismo (de código libre) libre*. Barcelona: Icaria.

Salaverría, R. (2015): "Los labs como fórmula de innovación en los medios", *El profesional de la información*, 24, julio-agosto, n. 4. Recuperado de: https://goo.gl/UZdQdF [COnusltado el 15/04/2018].

Sánchez García, P.; Campos-Domínguez, E. y Berrocal, S. (2015): "Las funciones inalterables del periodista ante los perfiles multimedia emergentes", *Revista Latina de Comunicación Social*, 70, pp. 187-208. DOI: 10.4185/RLCS-2015-1042. [Consultado el 21/05/2018].

Shadbolt, N.; O'Hara, K.; Berners-Lee, T.; Gibbins, N.; Glaser, H.; Hall, W. y Schraefel, M.C. (2012): "Linked open government data: lessons from Data.gov.uk". *IEEE Intelligent Systems*, 27 (3), Spring Issue, pp. 16-24.

Ubaldi, B. (2013): "Datos de gobierno abierto: hacia un análisis empírico de iniciativas de datos de gobierno abierto", en Documentos de trabajo de la OCDE sobre gobernanza pública , No. 22, Publicaciones de la OCDE, París, https://doi.org/10.1787/5k46bj4f03s7 -en [Consultado el 02/12/2018].

Vargas, I. (2012): "La entrevista en la investigación cualitativa: nuevas tendencias y retos". *Revista Calidad en la Educación Superior* (CAES), 3, n. 1, pp. 119-139. Universidad Estatal a Distancia: Costa Rica. Recuperado de: https://goo.gl/r9sWD2 [Consultado el 24/05/2018].

Zanchelli, M., & Crucianelli, S. (2013): "Integrando el periodismo de datos en las salas de redacción". Recueprado de https://ijnet.org/sites/default/files/integrating%20data%20journalism-spanish.pdf [Consultado el 02/12/2018].