# COMMUNICATION & SOCIETY

**Teresa Sandoval-Martín**
https://orcid.org/0000-0001-5106-7887
msandova@hum.uc3m.es
University Carlos III de Madrid

**Leonardo La-Rosa**
https://orcid.org/0000-0002-9522-4490
leonardolarosa2013@gmail.com
University Carlos III de Madrid

**How to cite this article:**
Sandoval-Martín, T. &. La-Rosa, L. (2018). Big Data as a differentiating sociocultural element of data journalism: the perception of data journalists and experts. *Communication & Society, 31*(4), 193-209.

# Big Data as a differentiating sociocultural element of data journalism: the perception of data journalists and experts

## Abstract

**The use of methods of the social sciences and computational tools to analyze databases in journalism has had several definitions since Philip Meyer called it precision journalism (PJ). In the last decade, this specialty has had an important development under the term data journalism (DJ), in a differentiating technological and sociocultural environment: Big Data. This research aims to differentiate DJ from PJ and computer assisted reporting (CAR) with a perspective taken from the science and technology studies, focusing the news as a boundary object between programmers, designers, journalists and other actors that now are part of the news production process. For this purpose, 14 in-depth interviews have been made from 2015 to 2017 to data journalists from Spain (8), EEUU (1) and Finland (1); PP, PD and transparency academic experts from Spain (1) and Finland (2); and one expert in transparency acts y access to public information in Spain, Europe and Latin American. As a result, it can be affirmed that big data is differentiating element of DJ because it is a sociocultural context where the open data philosophy, free software, collaborative and team work are part of its identity.**

## Keywords
**Data journalism, big data, transparency, open data, visualization, precision journalism, computer assisted reporting.**

## 1. Introduction

The convergence in the newsrooms and the new journalistic culture as a consequence of the use and treatment of large quantities of data in the network (Barbosa & Torres, 2013) have led to the study of a journalistic phenomenon known as data journalism (DJ) in the last few years (Barbosa & Torres, 2013; La-Rosa & Sandoval-Martín, 2016). Despite the boom of this emerging discipline, its study in the academic field has been divided till recently among different approaches that positioned it, on the one hand, as a discipline with several decades of existence (Tejedor & Dader, 2011; Egawhary & O'Murchu, 2013) and, on the other hand, as something new (Ferreras Rodríguez, 2013; Chaparro-Domínguez, 2013; Antón-Bravo 2013; López & Martínez, 2014). This confusion is produced –in our opinion–, because the term data journalism (DJ) already existed several decades ago along with data driven journalism (DDJ)

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:**
**the perception of data journalists and experts**

inside of what is known as precision journalism (PJ) and computer assisted reporting (CAR). It was considered as the quantitative turn of journalism that erupted in the United States in the seventies –PJ– that had its origins in the use of computers by Philip Meyer (1973) on his researches as a Miami Herald journalist in 1959.

However, the technological and sociocultural context DJ has been developing, as we know it today, is very different from the other models already mentioned, and this is the aspect this research is focusing on, clarifying the definition and characterization of this journalism from similar ones previously born that are still exercising.

Besides, the vertiginous technological advances and the irruption of many disruptive technologies have been in DJ a breaking point derived from the influence of the *Open Data* movement and the massive leaks of Wikileaks in 2010. In fact, even though there were some DJ projects a few years before done by *Propublica* –in 2008– and *The Guardian* –in 2009 with the *Data blog*–, the term data journalism became popular in the media since the huge quantity of compromising files revealed at Julian Assange's web site. That year *The Guardian* multiplied the production of DJ pieces (Rogers, 2011). Besides the leaks, the trend towards greater transparency brought with it the liberation of other databases through government transparency websites, making the term popular in newsrooms and new digital medias, carrying on both its academic study in universities and the creation of independent DJ teams and data desks in the newsrooms. This research considers this environment and pretends to discover if in the professional practices of the data journalists is reflected the open data and transparency culture, as well as getting to know the opinion of experts in this matter. This will confirm for the first time the believe that *big data* is a differentiating element of DJ in front of other kinds of journalism that work with database analysis to narrate journalistic stories.

As the ultimate aim of the research we consider that due to the credibility crisis that the media and the journalistic profession are going through, the presence of journalists –in the newsrooms– with a mentality close to the fundamentals of the open data –hacker's work ethic, free software and collaborative work– could contribute to improve the quality of journalism. This is a way of thinking already originated with precision journalism because of the use of social sciences techniques that now is also supported by the mythologies of big data. Besides, the fact that the experience of working in multidisciplinary teams in DJ is positively valued, is a good indicator that in the future journalists could work more closely with other technical professions in the newsrooms.

The cultural aspects reflected in the professional practices of Spanish data journalists will be analyzed from a global perspective with the point of view of professionals and experts from the USA, United Kingdom, Spain and Finland, countries where the open data movement –through organizations such as Access Info Europe and the Open Knowledge Foundation among others– has participated actively in the disclosure of public data and the training of data journalists.

## 2. Conceptual framework

In an article published in 2016 by the same authors of this research, we gathered the differences existing in other forms of database research that preceded DJ, such as PJ and CAR. We said about the last one –that as it is known, came up in the 80's with the personal computers– that, among other cultural and technical aspects, during that time CAR journalists centered their efforts in searching and gathering information in a world where it was scarce. In its place, in DJ the information on the web is abundant and the effort goes to data processing instead of search and gathering information (La-Rosa & Sandoval, 2016). Besides this variable, in DJ "we can sense in DJ a transparency culture owns by hacktivists that was not evident in CAR" (La-Rosa & Sandoval, 2016).

We agree with the DJ definition made by Rodríguez-Brito and García-Chico (2013, p. 60):

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

The development of data journalism summarizes digitalization processes; transparency, access to public information and open data philosophy; statistics and visualization tools; and the research skills of the journalist of all times to compile, filter, contextualize, contrast, prioritize and count a story in an attractive manner.

Nevertheless, from an exhaustive literature review about the subject, we were able to extract in 2015 our definition of the term, which is confirmed in some aspects with the results of this study:

Data journalism is the journalism proposal that has emerged from the web that shares with precision journalism and CAR a methodological rigor similar to the one owned by the social sciences for data extraction, standardization and procession in order to publish journalistic stories.

Its fundamental difference with these two branches of journalism is in the addition of arising problems and techniques in the context of big data, which in turn has led to establish the following aspects:

1- its context in the big data phenomenon;

2- the need for data processing and databases in this context;

3- philosophy of open data and transparency from two perspectives:

a- the one needed from public and private institutions for the correct research in this branch of journalism, and

b- another one that the journalist must have as a part of its deontology;

4- data visualization using computer graphics and interactive applications;

5- citizen journalism as a support in the absence of transparency,

6- collective work from two perspectives:

a- that one related to citizen journalism and

b- the other one related to the collective production of information on a data journalism team. (La Rosa, 2015, p. 91).

Currently big data it is not conceived as "merely a technological transition toward data deluge" (Lewis & Westlund, 2015, p. 448), since there's not a unitary big data phenomenon (Boellstorff, 2013). Boyd and Crawford (2012, p. 663) define it as a cultural, technologic and academic phenomenon that rests on the interplay of three factors: (1) the technology –thanks to the power of the algorithms to collect, analyze, link and compare huge quantities of data–; (2) the analysis –that allow to identify patterns with diverse aims–; and (3) the mythology –the large datasets will offer accurate perceptions providing an aura of truth, objectivity and accuracy. This point of view is shared by Lewis and Westlund (2015, p. 447): "Big data is a social, cultural, and technological phenomenon –a complex amalgamation of digital data abundance, emerging analytic techniques, mythology about data-driven insights, and growing critique about the overall consequences of big-data practices for democracy and society."

In the journalistic field, big data needs to be open (Elías, 2015). The term open data is associated to transparency, accountability, accessibility, and free, public and recyclable use. At the core of this philosophy underpin the free software movement and hacktivism. The data publication is closely related to a transparency culture owned by the network society and linked by Himanen (2001) to the hacker work ethics, when he assures that the money is no longer an incentive and the profit is measured in goals as the social value that generates the work undertaken, its transparency and the free access to the work created. As a result of the open data movement in the world we have governmental transparency laws. The publication of data by journalists could be considered as a transparency work as well, but besides these publications , occasionally leaks lead later to "processes of refinement between journalists" (Lesage & Hackett, 2014, p. 41).

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

Despite of the apparent advantages that big data offers to journalism, writing a journalistic story from a huge quantity of data has become a new challenge, not without difficulties and ethical issues to take into consideration. Carlos Elías (2015, p. 94) alerts about the difficulty of "building a theory or narrative when there are only a lot of numerical data without any explanation," and for those cases he considers that grounded theory[1] might be a solution because "it's useful to generate a hypothesis from the data obtained in a research." The author of *Big Data and journalism in the network society* demystifies one of the big data mythologies, the one about the veracity that data offers, because there might be deontological corruptions that willfully distort reality. Sometimes quantity is confused for quality when the so-called *small data* –experts' interviews or opinions– explain better the reality than a dataset. Besides, he highlights the importance of explaining the data within their context and how big data loses its meaning without it, as Philip Meyer (2011) also points out: "today we have a never-ending stream of data, more than we can ever understand, and so the need to explain becomes paramount." Finally, another of the most important difficulties of DJ today is the lack of professionals to cover the needs of the media.

## 3. State of the art

If we observe the results of a simple search of communication articles by terms similar to DJ in the database of Web of Science, we'll have as a result 281 articles since 1992, with a mean of 2,47 articles per year between 1992 and 2007. However, since 2008 the publications take an impulse and it is from 2012 when an exponential grow begins (Figure 1).

**Figure 1**: Articles with terms similar to DJ by years. Source: Self-elaborated.



N º de artículos por año
(Web of Knowledge)
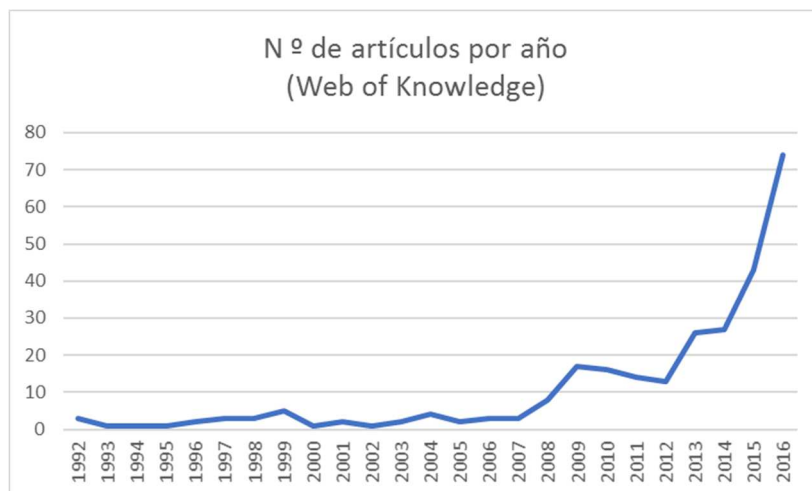
Therefore, even though data journalism as we know it today never existed till 2008 approximately, since 1990 many scholars have studied the use of databases in the newsrooms. Messner and Garrison (2007) gathered the previous literature review in their article "Journalism's 'Dirty Data'. Below Researchers Radar," doing a content analysis of the most

---

[1] Grounded theory (Glasser & Strauss, 1967) was used in the stories of the project "Reading the Riots," made by *The Guardian* and the London School of Economics in 2011, which was inspired by the flagship report of PJ made by Philip Meyer about the riots of Detroit in 1967. The difference between these two lies in the fact that Meyer recollected information by sociological methodologies –surveys and interviews– using computers from that time to code and analyze in order to detect which one of the three possible sociological theories was true, and in "reading the Riots" the social researchers of the London School of Economics that collaborated with journalists from *The Guardian* searched from a huge quantity of data a satisfactory theory that explained the riots, using the grounded theory (Elías, 2014).

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:**
**the perception of data journalists and experts**

prestigious reference publications that have included articles about this phenomenon. Their article was a call for attention to the way databases were being used and to the great mistakes that journalists used to make during their researches due to their lack of technical knowledge. A decade later, these issues still concern the scientists, an example of this is the claim for a radical shift in newsgathering and journalistic interpretation related to stories with statistics (Cushion, Lewis & Callaghan, 2017), as well as the identification of connections across heterogeneous data sources that enable journalists to obtain useful information (Chanial *et al.*, 2018). Other themes of interest are the difficulties to access public data that journalists have (Alppelgreen & Salaverría, 2018; Álamo, Cruz & Elías, 2018), and the training of journalists in new knowledges and skills (Bradshaw, 2018; Gray, Gerlitz & Bounegru, 2018; Splendore *et al.*, 2016).

To Fink and Anderson (2015, p. 1-2) DJ research might be divided in three great strands:

> The first strand, which for a long time included the majority of computational journalism studies, is geared primarily toward professional journalists and addressed practical concerns (e.g. Cohen *et al.* 2011; Flew, Daniel & Spurgeon, 2010; Nguyen, 2010; Hamilton & Turner, 2009). A second strand maps the infrastructures and arenas that enable the connection between computer scientists and journalists, including research focusing on the relationship between data journalism and the rhetoric and institutional structures of the open-source software movement (Lewis & Usher, 2013); the *New York Times* technology department (Royal, 2010; Weber & Rall, 2013); the organizational roles played by data journalists in Chicago (Parasie & Dagiral, 2012); and Ananny's work on "press–public collaboration as infrastructure" (Ananny, 2013). A third strand historicizes current developments, examining the links between computational journalism and older forms of data-oriented newswork, such as Computer Assisted Reporting (Parasie & Dagiral, 2012; Powers, 2012).

As can be observed, it is since the last years of the first decade of the XXI century that different themes about collaboration spaces between journalism and computer sciences began to be studied and new concepts own by the free software culture are introduced: the transformation that programmers practice in the news (Betancourt, 2009); journalists as programmers (Royal, 2010); the transparency as a new journalistic standard, open data and hacktivism (Karlsson, 2011; Lewis & Usher, 2013 & 2014; Baack, 2015); DJ and open data in the USA, the United Kingdom and Argentina (Aitamurto, Sirkkunen & Lehtonen, 2011); the interaction between information, journalists and experts in technology in EUU and Europe (Flew *et al.*, 2012), as well as that one found between data journalists and civic technologists or hacktivists in the United Kingdom and Germany (Baack, 2018), among others. It is also to be noticed the studies of Parasie and Darigal (2012), that established that the participation in social movements related to free software and initiatives of open government of programmers in the newsrooms in Chicago have influenced in the way that these data tools have to be used in the newsrooms. In a similar way, it is interesting the conclusion reached by Degand (2013) in her research about the role of programmers and data scientists in Belgian newsrooms, whose participation in the news making process has affected the content and the separation between the technician and the journalist is increasingly less. Besides, complex issues related to transparency and open data are investigated, such as the work with data to create stories. That's how Lesage and Hackett (2014) developed three key dimensions in the incorporation of data in DJ practices from the theories of mediation: the problem of the scales, transparent work and access to open data. At the same time, Appelgren and Nygren (2014) assured that the access to open data and the activist's participation are crucial to DJ development in many countries.

Most of this kind of studies have been made from field works in countries like Sweden (Appelgren & Nygren, 2014), Norway (Karlsen & Stavelin, 2014), Belgium (De Maeyer *et al.*, 2015), Canada (Tabary, Provost & Trottier, 2015; Hermida & Young, 2016), the USA (Fink &

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

Anderson, 2015), England (Borges-Rey, 2016) and Great Britain (Borges-Rey, 2017). Others have been made in several countries at the same time such as Finland, United Kingdom and Argentina (Aitamurto, Sirkkunen & Lehtonen, 2011), Finland, the USA and the United Kingdom (Uskali & Kuutti, 2015), Sweden and Spain (Appelgren & Salaverría, 2018) and the USA and Spain (Álamo, Cruz & Elías, 2018). In Spain, this is the first empirical research that indicates that DJ answers to a concrete technological and sociocultural context: big data, with the consequences that this involves to journalistic practices and beliefs in Spanish data journalism as well as in other countries.

To research this kind of connection between different professional cultures and agents involves new theoretical and methodological approaches, more complex than the simple description of the phenomena. In fact, a swift of paradigm is produced, where the way the themes are addressed from a sociological approach are a tendency. This stream, to which this research is added, is not exclusive of DJ and it is observed in approaches and theoretical frameworks in the journalism research in these last years, just as the Nordic scholars Steen Steensen and Laura Ahva (2015) confirmed, when they analyzed the first 14 volumes (2000-2013) of two of the leading journals specialized in journalism (*Journalism - Theory, Practice and Criticism* and *Journalism Studies*). Although journalism (and the implications of digitalization to their practices and institutions) had been investigated so far from others perspectives –normative, empirical and constructivist (Domingo, 2008)–, since recently a forth tendency is observed, when it started to "theorizes the field beyond the traditional institutions and understandings of journalism," and began to study the foundations of digital journalism. "It investigates, for instance, the 'news ecosystem' (Anderson 2010), the 'news landscape' (Peters & Broersma 2013), 'ambient' (Hermida, 2010) and 'networked' (Heinrich, 2011; Russell, 2013) journalism (...)" (Steensen & Ahva, 2015, p. 1). The study concluded that "there has been a broad paradigmatic change in journalism studies since 2000 from perspectives of political science to sociological perspectives," so the "aspects related to philosophical perspectives (e.g. ethics and objectivity) are becoming increasingly important for journalism studies. However, these aspects are not viewed through perspectives of philosophy but are analyzed in a sociologically oriented framework, such as professionalism" (Steensen & Ahva, 2015, p. 12).

It is the case of this research, that it has taken into account the science and technology studies (STS), originated three decades ago from new currents of philosophical and sociological research of science, that have developed concepts to explain how diverse communities gather and interchange knowledge and resources and coordinate productively while maintaining its distinctive social and cultural identities (Galison's, 1997 & 1999; Star & Griesemer, 1989). The scholars from the United States Seth Lewis and Nikki Usher (2016) use this conceptual base to extrapolate to computer journalism what Galinson's (1997) called trading zones and boundary objects in a study of the cooperation between different kind of physicists. According to them:

> Different groups with distinct identities, types of expertise, or occupational backgrounds can come together around particular objects to communicate and collaborate, without losing any allegiance to their own field and its particular frame of reference toward those objects. Boundary objects thus are plastic enough to facilitate a variety of meanings across knowledge boundaries (Lewis & Usher, 2016, p. 5).

Besides, a research made in Sweden about DJ in that country (Appelgren & Nygren, 2014) has inspired us, as well as a study about the future of the journalistic profession due to the great influence of hacktivism in DJ, discourses and practices of different actors –journalists, editors, programmers and activists– in the French speaking part of Belgium (De Maeyer *et al.*, 2015). The approach from a sociological perspective enable us to ask new questions about the transformation of the values and practices of journalism, where the inclusion of these actor –

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

programmers, designers and hacktivists– and the recent identities that arose from their collaborative work pose a new conception about what data journalism is, differentiating it from other similar journalisms.

This new approach not only provides originality and value to the study, but also eliminate a gap in the research about data journalism, because it takes into account the sociocultural environment where it is developing.

## 4. Objectives and hypothesis

The results of this research seek out to clarify a rising field of study, data journalism, which requires more precise definitions.

The first objective of this research was to spot –empirically by in-depth interviews to professionals and experts– the existence of the interconnections described in the theoretical framework between DJ as we know it today and the environment where is developing. This environment comprehends the rise of other phenomenon linked to big data and that goes beyond the huge volume of data available with today's technologies and the ease of their management, such as open data (and the transparency culture linked to its philosophy), free software and collaborative work. These technologic, sociological and cultural phenomena own by the big data era, reflected in DJ practices, might be consider as differentiating characteristics regarding other former forms of journalism that used to work with databases –PJ and CAR–, thus contributing to the characterization of DJ.

The second objective is focus on describing the profile of professionals working on DJ and to detect if they work in multidisciplinary teams in the news production process. Built on the premise that the news is a boundary object between programmers, journalists and designers, we want to detect the possible consequences of collaborative work in a profession that till recently was basically a lonely one (Heinrich, 2011), especially in the field of investigative reporting. Among those repercussions we're interested in discovering if the journalists that are part of these teams have incorporated routines own by the culture that covers open data, as well as other ethical issues related to transparency, collaborative work, data visualization and the possibility of reuse of data by other journalists.

Among other issues, we ask ourselves: Is big data from a technological and sociocultural point of view a differentiate element of DJ compared to other former journalisms that used to work with databases –PJ and CAR–?, what technological and sociocultural differences are between DJ, PJ and CAR?, does DJ own elements of the mythologies that goes along with the sociocultural context of big data?, is the new open data and free software philosophy in DJ a consequence of the incorporation of other professions of the digital era such as programmers, developers, computer graphists and designers in the news making process?

## 5. Methodology

In-depth interviews were made within an international approach. In-depth interview was the tool that we considered more suited by the research nature and because of the results obtained in studies made in other countries. Besides, we had the previous experience of its use to elaborate two reports about the condition of journalism in Spain, developed in the core of the research team PASEET (Diezhandino *et al.*, 2008 & 2009). The interview analysis and interpretation have been made by the phenomenological research method of Giorgi (1985), adapted to this study, that through the natural unit of meaning extracts the meaning condensation of the textual quotation, obtaining a compendium of meanings in shorter formulations (Kvale, 1996). We wanted to identify attitudes, values and points of view from different contexts (professional and academic, Spanish and abroad), as well as to understand the interaction processes, along with the interviewees freedom to express their answers.

There were made 14 semi-structured in-depth interviews between 2015 and 2017 to data journalists from Spain (8), EEUU (1), Finland (1) –significant country in open data and

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

transparency–; academic experts from Finland (2) and Spain (1) in PJ, DJ and transparency; as well as one expert in transparency laws and access to public information in Spain, Europe and Latin America. All these interviews were personal.

Specifically, 10 data journalists were interviewed to detect if what we pointed out in the theoretical framework about the influence of the open data culture in DJ was reflected on the beliefs and practices of the actors of the news production process, and to know their opinions about other aspects, amongst which what DJ offers to journalism is very important. We also interviewed experts in data journalism and transparency that we knew could give dissenting opinions on our initial hypothesis: big data is a technological and sociocultural element that characterize DJ and differentiate it from others journalism that share with it some characteristics.

The interviews sample includes prominent Spanish data journalists of reference media: *El Mundo*, *Eldiario.es* y *Elconfidencial.com*. Besides, between 2015 and 2017 –period where all the interviews were made– there was the Spanish Center for Investigative and Data Journalism (*CEPID* because of its Spanish acronym), startup that offered this kind of services. The data journalists interviewed were: Paula Guisado and Hugo Garrido –*El Mundo*–, Belén Picazo and David Ruiz –*Eldiario.es*–, Jesús Escudero and Adrián Blanco –*Elconfidencial.com*–, and Frank Belyeu and Javier Galán –*CEPID*–. The sample also includes experts interviewed in Spain: the journalism scholar and PJ expert, José Luis Dader, and the director of Access Info Europe, Helen Darbishire, prominent specialist in transparency and open government.

Besides, since the Nordic media have a prominent developing in DJ and their researchers are among the theoretical that have published more about this matter, we made an international stance in Finland, country were 3 interviews were made: to a data journalists of the Finnish Broadcasting Company *YLE* –Teemo Tebest– and to the Finnish academic's experts in DJ Turo Uskali y Heikki Kuutti. Also, from the USA was interviewed the professor of journalism and Pulitzer Award winner Steve Doig, pioneer in computer data analysis in investigative reporting.

## 6. Results

### 6.1. *Data journalism in front of other journalisms: elements for its characterization*

First, the data journalists consider that DJ is, before anything, journalism. They linked it to PJ and investigative journalism (IJ). They recognize the existence of differences about former methods of data research. They consider that the technological advances have reinforced the classic values of journalism, such as rigor and veracity.

Frank Belyeu from CEPID calls it in-depth journalism and clears out: "DJ, PJ and IJ are not always the same, because they change in their methodology. DJ uses methods similar to those of PJ, from social sciences, in a more systematic way than IJ." Accordingly, Adrián Blanco from *Elconfidencial.com* considers that DJ has incorporated more accuracy to journalism and now is possible to "make another type of articles and researches that you couldn't prove before." He emphasizes the possibility of fact-checking that DJ enables and remind that many times the veracity of the data is taken for granted because they come from a source, but they're not truthful. In a similar manner, Paula Guisado from *El Mundo*, points out that "quality DJ replies values and good uses of journalism in general" and "adds positive aspects as the requirement of *fact-checking*, because when you work with data the review of the process is essential."

In the same line that other professional colleagues, to David Ruiz "inquiring in the data and the sources inquiring, to analyze," is something that has been always done. He neither considers that "such a radical change in the very concept, in the values" had happened, "but in the way, we can access to data that was impossible before," in a clear mention to transparency, "and how we can work them [the data], process them and take from them much

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

more," in the sense of the technological tools available. In fact, everybody agrees on the fact that technological advances have reinforced classic values of journalism, as well as the rigor and veracity.

All journalists pointed out that the new computer tools –non-existing in the times of PJ and even CAR– had contributed to DJ expansion and its rapprochement to the public, thanks to today's software that works with data and visualize them. According to this, the coordinator of *El Mundo* Data and ex-member of CIVIO Hugo Garrido said: "when come to its popularization what has democratized it a lot was the access to certain technological tools that facilitates the access of a great quantity of data and their visualization." Most of these tools were designed by programmers that defend free software. In fact, regarding the collaborative work and open data philosophy or transparency, Belyeu highlighted that "without the hacker community, the massification of tools for journalists wouldn't be possible" and precisely this availability of software to develop DJ by non-technical users is one of their main strengths.

Turo Uskali, who has investigated DJ in the USA, the United Kingdom and the Nordic countries, assured that DJ has transformed many journalism's values for the better, but especially when it comes to transparency: "even our first interviews with data journalists [in Finland] and also with internationals, since 2013, they said that is the 'ought to be:' whenever they do a story based on data, they will share this data openly for others reporters to use."

The expert of Access Info Europe, Helen Darbishire, cleared out that "transparency is the result of having the right to access information," and it manifests itself in "the obligation of governments to publish proactively and the right of any citizen to request information and to receive an answer or explanation of why the requested information is not being given to her or him." Many examples of DJ based on open data have already raised the conflict own by IJ – between journalist and the investigated entity– by achieving the publication of data that were not public before proactively through the open data movement, which in turn continues to pressure for the publication of public data in a reactive way through strategic litigation:

> We have had IJ development thanks to the possibility of having access to available information proactively o reactively from governments in other countries since so much time before we had access here in Spain [...] With the open data movement we have today the publication of so much more information than 10 years ago. We have benefited from the mix of new technologies along with claims for transparent governments, and we have advanced very much.

To the expert in PJ José Luis Dader, "DJ and PJ are the same thing, there's not any difference [...] is a problem of trends" and define it like this: "DJ or PJ is any type of journalistic information that use and analyze numeric and alphanumeric data in a systematic way to discover content of news interest." Nevertheless, José Luis Dader distinguishes DJ from journalism in general. He differentiates this discipline –specially, from the point of view of his more sociological perspective own by PJ–, ensuring that traditional journalism is framed in some news conceptions that looks for the exceptional, the rare, however, in PJ the common trend is to look for something frequent: its news are found out when it discovers structures instead of anecdotes. In addition to that, he reminded us that not every DJ or PJ has to be a research or IJ. Even though he says that some data visualizations are very spectacular, he emphasizes that in many cases there's not a truly work of DJ:

> Some dynamic tables that were unable to do in the printing press are being made now, but this work has been simply done with year-books released by ministries, nothing else has been worked: is a technological boast about what to do with some numbers that were already in an Excel spreadsheet that a public office facilitated. There, the truly added value of DJ is missed.

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

In a similar manner that Dader, Steve Doig emphasized that DJ is not a theoretical discovery: "to some this is a new discovery and even though we've pointed out what we were doing since the 80's, there were people like Philip Meyer that was doing this since the 60's." Nevertheless, he does highlights the role of the technological advances of today as a disseminator motor of DJ:

> Even here in the United States we have had fabulous research projects done by small newsrooms in tiny newspapers and online workshops, so realizing that the tools are not so difficult to use once you have trained a little bit to find the data, it will help you to package the story that you want to do.

On his behalf, the Finnish data journalist, Teemo Tebest, assured that DJ "is a tool" and that since long time ago "investigative journalists have been treating with data." In the same line as Dader, he considers that DJ is a trending term and that investigative journalism have been doing the same since decades, even though he recognizes that the technological advances have succeeded in making DJ do things that were impossible to do before.

Following this same line of argumentation, professor Heikki Kuti assured that the treatment of great quantities of data in journalism has always existed, but he highlighted a technological change in the present era under the tag of big data, even though it could owe its existence to a first data digitalization in an age without so much technological advances:

> It's about the data and includes a lot of mathematics and statistics. The data are just a tool to journalism and, in fact, is an old idea [...] when I visited USA 20 years ago there were some tapes in the computers and it wasn't this big data quantity, but if all of this information hasn't been gathered in digital, it would be impossible to discover this, so technologically there's a change.

As for the Finnish academic expert, Turo Uskali, he mentioned the technological differences in three different contexts when he assured that "Meyer was among the firsts in realizing that we could use statistics and computers in the 60's. In the newsrooms computers came in the 80's and now [all of us] have personal computers."

## 6.2. *Multidisciplinary teams and collaborative work in DJ*

It has been noted that the mixed profile of a data journalist requires: IJ skills; data processing and analysis with spreadsheets such as Excel or more complex databases; basic knowledge of statistics, web developing and design; data visualization capacities with online tools and to know some programming language. However, all over the interviews it has been cleared that none of these data journalists cover the whole of the mixed profile that they announce, namely, the ideal profile of a data journalist.

The technical impossibility for now that just one journalist covers completely all the fundamental aspects of the news making process in the Spanish DJ and, therefore, that just one person has all the aspects of the ideal mixed profile, has originated that this discipline organizes itself in Data Desks or DJ Teams. This is a contrasted reality in Spain and EEUU, Finland and the rest of the Nordic countries. As Jesús Escudero says: "In DJ, is very difficult that one person complies with the whole profile: it's usually a multidisciplinary group with several profiles. At the very least you need three people to make a data unit: a journalist, a programmer and a designer."

The DJ news making process linked to the organization of multidisciplinary teams was highlighted by most of the interviewees (journalists and experts). Furthermore, from most of the data journalist's profile described by the professionals, it is extracted that how the work organization in the news making process in DJ has generated new practices. These are directly related to the insertion of new actors in the news making process: the computer specialist and programmer, the designer and computer graphist, and the web developer among others.

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

As a new cultural practice, the collaborative work own by the free software culture is emphasized, as the journalist Teemo Tebest says: "DJ has changed the values of *lonely wolf* journalists. It is becoming well known that the collaborative work is better than keeping the data for yourself." Professor Heikki Kuuti clearly expose the fundamental role of these groups' members: "they usually organize themselves in teams, like *YLE* and *Helsingin Sanomat*. They have different people with different skills. For instance, technical advanced people usually know what you can do or not with data and journalists have the idea about what they want to achieve with data, what kind of answers they want to get from data".

This work organization according to the journalistic production process has been a reality in all the media and collectives interviewed, where the collaborative work is essential. In larger media, like *El Mundo*, this collaboration is more focused between members of the same team that later collaborate with the rest of the media's sections. In *Eldiario.es* case, being a smaller media, the collaboration between those who do DJ is more transversal in relation to the rest of the newsroom. On its behalf, the Data Desk of *Elconfidencial.com* works along with the members of *Elconfidencial.lab* –of which is a small part– and the difference sections of the newsroom. These collaborations not only tend to separate according to the strengths of each data journalist, but to emphasize the computer specialist or web developer labor inside the team, as the statements of the next table shows (Figure 2):

**Figure 2**: Work organization in the DJ teams. Source: Self-elaborated.

| eldiario.es | El Confidencial | EL MUNDO |
|---|---|---|
| "We are two at *Eldiario.es*, though there are more people that collaborate later from different sections. There are people that dedicate themselves visualize, others to data analysis, others to look for the news inside that data, to write." (Belén Picazo). | "We have similar profiles but each of us is specialized in one thing, and later there is also a coder, that is, we're four in the everyday data desk. We have a job well organized with assigned roles." (Adrián Blanco). | "At *El Mundo* we're in luck to count with a data desk made up of three data journalists and a developer […] it's formed by the boss with a mixed profile of journalist and computer specialist, two data journalists and a programmer." (Paula Guisado). |

Since CEPID doesn't belong to any kind of larger and vertical media structure, the different profiles of the team complement in a precise manner each part of the DJ news production process, as there are three data journalists, two developers and one designer. Javier Galán explained that "the journalists have the idea, support themselves on the programmers, and the graphic aspect is incorporated by the designer." At first sight you can see as these entrepreneurs can represent, at least at a structure level, the ideal of a DJ team. Having at least one specialist in each part of the news production process is something that the coordinator of *El Mundo* Data, Hugo Garrido, misses: "We had a complete team till the lasts of June, later a job inter-regulation was applied and one person from the team got fired, precisely the one with the most differentiate profiles."

The most differentiate profile that Garrido talks about is the designer or computer graphist, who is in charge of visualize the data to improve the communication of the information made by the data journalists. This profile, along with the coder-journalist, is one of the key pieces when it comes to boost DJ in Spain.

All of this demonstrates that collaborative work is one of the fundamental aspects and that the news is a boundary object between journalists and programmers. Not for nothing, Jesús Escudero gets to tag as a data journalist a programmer of his data desk and Belén Picazo,

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

a data journalist –who is a perfect example of the intersection of informatics and graphic designs in journalism–, points out: "a data journalist must have basic notions of design and not so basic of coding."

## 7. Conclusions

Regarding the first objective of this research, journalists and experts don't admit explicitly that big data is a sociocultural phenomenon that differentiates DJ from other former journalisms that used huge amounts of data or databases. However, if we take into account the analysis of the natural units of meaning we can observe that they acknowledge elements own by big data in DJ, such as: the massification of tools to work with data –thanks to free software–; the huge importance of the open data existence and the claim for more transparency; the collaborative work with the public disposition of data and results for their reuse –characteristic of the open data movement–; the team work with groups of multidisciplinary professional, and the visualization of data to make more "digestible" the results for any user. All of these characteristics are owned by big data as a phenomenon and, therefore, the main objective of this study is reached, detecting the interconnections described in the theoretical framework between DJ and the environment where is developing, with the rise of some phenomena associated to big data such as open data, free software and collaborative work in the practices of data journalists that are part of the sample. This leads us to confirm the central hypothesis about the existence of differentiating elements of DJ that are based in phenomena –technologic, sociological and cultural– owned by the big data era regarding other forms of journalism that work with datasets.

Nevertheless, even though some experts insist that DJ, PJ and CAR are the same phenomenon, and that this is only a matter of name changing, a fad, all the interviewee agreed on emphasize the technological advance that DJ has benefited. In fact, the digitalization and the possibility of contrasting data are considered by many authors as a differentiate characteristic own by DJ in front of other forms of journalisms. After all, to the data journalists and experts interviewed DJ is a more objective journalism with less bias.

Regarding cultural phenomena, journalists and experts recognize that a greater access to data thanks to the open data philosophy is another characteristic own by DJ, from which another difference has been made regarding former times where transparency laws didn't exist. Inside this open data culture there is also the possibility of results or information reproduction that DJ is looking for, even though we could think that PJ already looked for it with its approach to the scientific method. However, we can see in DJ practices an evolution of the transparency culture in journalistic practices, since the data journalists tend to show the process by which they came with their information and publish the raw data to other researchers' disposal. There's not the *per se* search owned by PJ for social sciences practices.

Specifically, these practices are part of a revalue of the profession through the use of new computer tools didn't exist by the times of PJ and CAR, and at present have extended thanks to the democratization of technology through the activist hacker community. Although these transparency practices made by journalists are similar to those of the scientific method, they are more rooted in a viewpoint similar to the free software community –reuse by other actors– and that Himanen (2001) called the hacker work ethic than in PJ aspirations to approach to social sciences.

Even though there is a consensus (both the experts and the journalists) about the use of big data tools –to process huge amounts of data– as a fundamental aspect of DJ, there hasn't been an agreement on regarding the big data sociocultural context as a differentiate element of DJ in relation to PJ. The reasons that we consider are underlain in these postures, discrepancy relays on considering or not the difference between the transparency culture that comes from the free software philosophy in each of the contexts. Today, more than ever, this culture has been massified to a point where transparency has become a fundamental

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

value not only for journalism, but for many other fields. Is no longer about knowing how to process some data –generated by governments– with techniques own by the social sciences –as in PJ–, but a whole movement that advocates for a more transparent society that exercises in the foreground, making an example, by publishing the data and methodologies of journalistic research.

If the importance of transparency and other elements own by the big data era are acknowledged, like verification or fact-checking, as a differentiate element of DJ, it can be concluded that there's been a substantial change in relation to formers forms -PJ and CAR- that can be summarized in these further points:

1- The use of techniques to generate information and treat huge quantities of data with tools to access and work with (to cross-check, filter, analyze, etc.).

2- The raise of a conflict between the researcher and the entity being investigated through the open data movement and the publication of data actively and reactively by public offices.

3- The reinforcement of the deepening and journalistic research through rigor, veracity and verification of information with fact-checking techniques non-existing when PJ and CAR emerged.

It is confirmed that phenomena and mythologies owned by big data are part of DJ by being directly associated to differentiating technological and sociocultural elements.

Regarding the second objective, due to the impossibility of finding a data journalist that gathers all the proper skills to exercise the ideal profile of DJ, new cultural actors have been incorporated in the news production process. This has led to in the organization of teams where the collaborative work practice by professions of the big data era such as programmers and web developers is essential. This collaborative work implied in the free software culture where the news is a boundary object between journalists, programmers and designers is another fundamental aspect that differentiates DJ from PJ and CAR practices.

In the view of the results we arise the need of developing new research lines about transparency in the newsrooms and news from different perspectives related to data treatment in the big data and post-truth era, specially because of the latest advances on artificial intelligence and language, the algorithms applied to the creation and dissemination of news, fake-news, etc. Our research team is currently finishing a content analysis about news related to the Transparency Law in Spain in the media and another study about transparency in the automated journalism environment.

## References

Aitamurto, T., Sirkkunen, E. & Lehtonen, P. (2011). Trends in Data Journalism. *Espoo: VTT*, 0–27. Retrieved from http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/ D3.2.1.2.B_Hyperlocal_Trends_In Data_Journalism.pdf.

Ananny, M. (2013). Press-Public Collaboration as Infrastructure: Tracing News Organizations and Programming Publics in Application Programming Interfaces. *American Behavioral Scientist, 57*(5), 623–642. Retrieved from https://doi.org/10.1177/0002764212469363.

Anderson, C. W. (2010). Journalistic networks and the diffusion of local news: The brief, happy news life of the "Francisville Four." *Political Communication, 27*(3), 289-309. Retrieved from https://doi.org/10.1080/10584609.2010.496710.

Appelgren, E. & Nygren, G. (2014). Data Journalism in Sweden. *Digital Journalism, 2*(3), 394–405. Retrieved from https://doi.org/10.1080/21670811.2014.884344.

Baack, S. (2011). A new style of news reporting: Wikileaks and data-driven journalism. *Cyborg Subjects*, 1–10. Retrieved from https://www.ssoar.info/ssoar/handle/document/40025.

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big data & Society, 2*(2). Retrieved from https://doi.org/10.1177/2053951715594634.

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

Barbosa, S.O. & Torres, V. (2013). O paradigma 'Jornalismo Digital em Base de Dados': modos de narrar, formatos e visualização para conteúdos. *Galaxia, 25*, 152–164.

Betancourt, L. (2009). How Programmer/Journalists Are Changing the News. Retrieved from https://mashable.com/2009/12/11/programmer-journalists/?europe=true#JJqbAyZiBOqI.

Boellstorff, T. (2013). Making *Big data*, in theory. *First Monday, 18*(10). Retrieved from https://doi.org/10.5210%2Ffm.v18i10.4869.

Borges-Rey, E. (2016). Unravelling Data Journalism. A study of data journalism practice in British newsrooms. *Journalism Practice, 10*(7), 833–843. Retrieved from https://doi.org/10.1080/17512786.2016.1159921.

Borges-Rey, E. (2017). Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism*. Retrieved from https://doi.org/10.1177/1464884917693864.

Boyd, D. & Crawford, K. (2012). Critical Questions for Big data. *Information, Communication & Society, 15*(5), 37–41. Retrieved from https://doi.org/10.1080/1369118X.2012.678878.

Bravo, A. A. (2013). El periodismo de datos y la web semántica. *Cuadernos de Información y Comunicación, 18* (November), 99–116. Retrieved from http://www.homeworkmarket.com/sites/default/files/q4/16/06/servant_leasership.docx.

Chaparro-Domínguez, M. Á. (2013). La evolución del periodismo de precisión: el blog de *The Guardian* sobre periodismo de datos. In B. Lloves Sobrado & F. Segado Boj (Eds.), *I Congreso Internacional de Comunicación y Sociedad Digital*. Retrieved from http://llwh256.servidoresdns.net/handle/123456789/1736.

Cohen, S., Li, C., Yang, J. & Yu, C. (2011). Computational Journalism: a call to arms to database researchers. *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research (CIDR 2011).* Asilomar, California, USA, (January), 148–151. Retrieved from http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//pubs/archive/37381.pdf.

Dader, J. L. (1997). *Periodismo de precisión: vía socioinformática para descubrir noticias*. Madrid: Síntesis.

Dagiral, E. & Parasie, S. (2012). Quand le web colle au territoire : L'exploration de l'information hyperlocale à Chicago. *Sciences de La Société, (84−85),* 81-101.

De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F. & Le Cam, F. (2015). Waiting for Data Journalism. *Digital Journalism, 3*(3), 432–446. Retrieved from https://doi.org/10.1080/21670811.2014.976415.

Degand, A. (2013). Le rôle des techniciens et des informaticiens dans la co-production de l' information en ligne. *Global Media Journal - Édition Canadienne, 6*(1), 25–43.

Diezhandino, M. P., Carrera, P., Sandoval, T., Fernández, J., García, F., Yuste, B., Sánchez, J. C. (2008). *Periodismo en la era de Internet. Claves para entender la situación actual de la información periodística en España*. Barcelona: Ariel.

Diezhandino, M. P., Sandoval, T., Yuste, B., Carrera, P., Fernández-Beaumont, P., Aguado, G., Eiroa, M. & García, J. V. (2009). *Periodismo Digital en tiempos de crisis. Un caso de estudio: el tratamiento de la inmigración en los medios*. Barcelona: Ariel.

Domingo, D. (2008). Inventing Online Journalism: A Constructivist Approach to the Development of Online News. In C. A. Paterson & D. Domingo (Eds.), *Making Online News: The Ethnography of New Media Production* (pp. 15–28). New York: Peter Lang. Retrieved from https://doi.org/10.1177/1461444809106229.

Egawhary, E. & O'Murchu, C. (2013). *Data Journalism Book*. London: Centre for Investigative Journalism. Retrieved from http://issuu.com/tcij/docs/data_journalism_book?e=2989993/2615602.

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

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

Ferreras Rodríguez, E. M. (2013). Aproximación teórica al perfil profesional del 'Periodista de datos'. *Revista ICONO14. Revista Científica de Comunicación y Tecnologías Emergentes, 11*(2), 115–140. Retrieved from https://doi.org/10.7195/ri14.v11i2.573.

Fink, K. & Anderson, C. W. (2015). Data Journalism in the United States. Beyond the usual suspects. *Journalism Studies, 16*(4), 467–481. Retrieved from https://doi.org/10.1080/1461670X.2014.939852.

Flew, T., Spurgeon, C., Daniel, A. & Swift, A. (2012). The promise of computational journalism. *Journalism Practice, 6*(2), 157–171. Retrieved from https://doi.org/10.1080/17512786.2011.616655.

Galindo Arranz, F. (2004). Propuesta de periodización histórica y evolución conceptual del Periodismo de precisión. *Estudios sobre el mensaje periodístico, 10*, 97–112.

Galison, P. (1997). *Image and logic: a material culture of microphysics*. Chicago: The University of Chicago Press.

Galison, P. (1999). Reflections on image and logic: A material culture of microphysics. *Perspectives on Science, 7*(2), 255–284. Retrieved from https://doi.org/10.1162/posc.1999.7.2.255.

Giorgi, A. (1985). *Phenomenology and Psychological Research.* Pittsburgh: Duquesne University.

Hamilton, J. T. & Turner, F. (2009). Accountability Through Algorithm: Developing the Field of Computational Journalism. *A Center for Advanced Study in the Behavioral Sciences Summer Workshop Report*, 1–22. Retrieved from http://www.stanford.edu/~fturner/Hamilton Turner Acc by Alg Final.pdf.

Heinrich, A. (2011). *Network journalism: Journalistic practice in interactive spheres*. New York: Routledge. Retrieved from https://doi.org/10.4324/9780203830451.

Hermida, A. (2010). Twittering the news: The emergence of ambinet journalism. *Journalism Practice, 4*(3), 297–308. Retrieved from https://doi.org/10.1080/17512781003640703.

Himanen, P. (2001). *The Hacker Ethic and the Spirit of the Information Age*. New York: Random House. Retrieved from https://archive.org/details/TheHackerEthicAndTheSpiritOfTheInformationAge.

Karlsen, J., Stavelin, E. (2014). Computational Journalism in Norwegian Newsrooms. *Journalism Practice, 8*(1). Retrieved from https://doi.org/10.1080/17512786.2013.813190.

Karlsson, M. (2011). The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority. *Journalism, 12*(3), 279–295. Retrieved from https://doi.org/10.1177/1464884910388223.

Kvale, S. (1996). *Interviews: An Introduction to Qualitative Research Interviewing.* California: SAGE.

La-Rosa, L. (2015). *El periodismo de datos en España: estudio comparativo entre Eldiario.es y el Centro Español de Periodismo de Investigación y Datos* (Tesis de Máster). Universidad Carlos III de Madrid, España.

La-Rosa L. & Sandoval-Martín, T. (2016). La insuficiencia de la Ley de Transparencia para el ejercicio del Periodismo de datos en España. *Revista Latina de Comunicación Social, 71*, 1208-1229. DOI: 10.4185/RLCS-2016-1142.

Lesage, F. & Hackett, R. A. (2014). Between Objectivity and Openness. The Mediality of Data for Journalism. *Media and Communication, 2*(2), 42-54.

Lewis, S. C. & Usher, N. (2013). Open source and journalism: toward new frameworks for imagining news innovation. *Media Culture & Society, 35*(5), 602–619. Retrieved from https://doi.org/10.1177/0163443713485494.

Lewis, S. C. & Usher, N. (2014). Code, Collaboration, And The Future Of Journalism. *Digital Journalism, 2*(3), 383-393. Retrieved from https://doi.org/10.1080/21670811.2014.895504.

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:
the perception of data journalists and experts**

Lewis, S. C. & Usher, N. (2016). Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers. *Convergence: The International Journal of Research Into New Media Technologies, 22*(5), 543–560. Retrieved from https://doi.org/10.1177/1354856515623865.

Lewis, S. C. & Westlund, O. (2015). Big Data and Journalism. *Digital Journalism, 3*(3), 447–466. Retrieved from https://doi.org/10.1080/21670811.2014.927986.

López, P. J. & Martínez, F. (2014). Definición, contexto y herramientas digitales para el desarrollo del periodismo de datos en España. *Revista Internacional de Tecnología, Conocimiento y Sociedad, 3*(1), 45–55.

Messner, M. & Garrison, B. (2007). Journalism's 'Dirty Data' Below Researchers' Radar. *Newspaper Research Journal, 28*(4), 88–100.

Meyer, P. (1973). *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Bloomington: Indiana University Press

Nguyen, D. (2010). Scraping for Journalism: A Guide for Collecting Data. Retrieved from https://www.propublica.org/nerds/doc-dollars-guides-collecting-the-data.

Parasie, S. & Dagiral, E. (2012). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society, 15*(6), 853–871. Retrieved from https://doi.org/10.1177/1461444812463345.

Peters, C. & Broersma, M. (2013). *Rethinking journalism: Trust and participation in a transformed news landscape*. London: Routledge. Retrieved from https://doi.org/10.4324/9780203102688.

Powers, M. (2011). "In Forms That Are Familiar and Yet-to-Be Invented". American Journalism and the Discourse of Technologically Specific Work. *Journal of Communication Inquiry, 36*(1), 24–43. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/0196859911426009?journalCode=jcia.

Powers, M. & Benson, R. (2014). Is the Internet Homogenizing or Diversifying the News? External Pluralism in the U.S., Danish, and French Press. *International Journal of Press-Politics, 19*(2), 246–265. Retrieved from https://doi.org/10.1177/1940161213519680.

Rodríguez Brito, A. & García Chico, L. (2013). Cartografiando el periodismo de datos. *Big data* y periodismo en el continente americano. Cinco casos de estudio. *Telos*, septiembre, 1–10.

Rogers, Simon (Ed.) (2011). *Facts are sacred. The power of data.* London: Guardian Books.

Royal, C. (2010). The Journalist as Programmer: A Case Study of *The New York Times* Interactive News Technology Department. *International Symposium in Online Journalism Journal, 2*(1), 5-24.

Russell, A. (2011). *Networked: A Contemporary History of News in Transition*. Cambridge: Polity Press.

Star, S. L. & Griesemer, J. R. (1989). Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology. *Social Studies of Science, 19*(3), 387-420. Retrieved from http://www.jstor.org/stable/285080.

Steensen, S. & Ahva, L. (2015). Theories of journalism in a digital age: An exploration and introduction. *Digital Journalism, 3*(1), 1–18. Retrieved from https://doi.org/10.1080/21670811.2014.927984.

Tabary, C., Provost, A.-M. & Trottier, A. (2016). Data journalism's actors, practices and skills: A case study from Quebec. *Journalism, 17*(1, SI), 66–84. Retrieved from https://doi.org/10.1177/1464884915593245.

Tejedor, L. & Dader, J. L. (2011). El patrimonio del César ante el Perro Guardián: un análisis de Periodismo de precisión sobre la transparencia económica de los miembros del Gobierno en España. *Estudios sobre el Mensaje Periodístico, 17*(2), 431-455.

Sandoval-Martín, T. &. La-Rosa, L.
**Big Data as a differentiating sociocultural element of data journalism:**
**the perception of data journalists and experts**

Uskali, T. I. & Kuutti, H. (2015). Models and Streams of Data Journalism. *The Journal of Media Innovations, 2*(1), 77–78. Retrieved from https://doi.org/10.5617/jmi.v2i1.882.

Young, M. L. & Hermida, A. (2015). From Mr. and Mrs. Outlier To Central Tendencies: Computational journalism and crime reporting at the *Los Angeles Times*. *Digital Journalism, 3*(3), 381–397. Retrieved from https://doi.org/10.1080/21670811.2014.976409.