El Español and El Confidencial, models of data sports journalism in Spanish native digital news media

El Español y El Confidencial, exponentes del periodismo deportivo de datos en los medios nativos digitales españoles

José Luis Rojas Torrijos. Universidad de Sevilla (jlrojas@us.es)

Agustín Rivera Hernández. Universidad de Málaga (agustinrivera@uma.es)

Abstract:

This research aims to evaluate the degree of development and the potential of data-driven journalism in sports reporting through the study of two of the most outstanding native digital news media in Spain, which are El Español and El Confidencial. To that end, we have analysed the content of all the reports published by both of these media outlets from February 2015 to July 2016, and have also carried out interviews with their respective sports and data editors in order to see if, in fact, advanced statistics and data visualization techniques are being used in sports journalism today, something that is already occurring in other areas of information reporting and in other countries.

Keywords:

Resumen:

La presente investigación tiene como objetivo evaluar el grado de desarrollo y la potencialidad del periodismo de datos en la información deportiva en España a través del estudio de dos de los medios nativos digitales de información general más destacados, El Español y El Confidencial. Para ello, se han realizado un análisis de contenido de las piezas publicadas por estos medios en el periodo comprendido entre febrero de 2015 y julio de 2016 y diferentes entrevistas a sus respectivos responsables de Deportes y Análisis de Datos; todo ello con el fin de comprobar si en este campo se han empezado a desarrollar técnicas de análisis de estadística avanzada y de representación visual de datos que ya se están empleando en otras áreas informativas y en otros países.

Palabras clave:

1. Introduction

At a time of rapid transition toward digitalisation of information content, sports journalism has become an innovative and highly dynamic field characterized by its continuous exploration of new ways of reporting news by utilizing innovative narrative techniques that have arisen in the web environment. Along this path, more and more media have begun to exploit the true potential of the new techniques for visual representation of information by making use of the vast amount of data being generated and accumulated in the world of sports. There is possibly no other area of specialized journalism apart from sports media where statistics have had such a strong impact on reporting (Rojas, 2015, pp. 79), to the point where in some categories it is very difficult to write a good review without relying on such data (Marrero Rivera, 2010). Moreover, these data also help to contextualize the importance of what has been achieved from a historical point of view.

Given the nature of this specialized area that deals largely with results and data accumulated on athletes and teams throughout a season, sports information is close to what is now called "data-based journalism", more commonly known as "data journalism". According to the *Oxford Dictionary of Journalism* (2014), this discipline relates to "the exploitation of databases for the writing of potential stories." According to the coordinator of this work, Tony Harcup, data journalism entails access to official information that has been released by public administrations (open data) or that has been obtained through informants (more closely related to investigative journalism), but it can also be defined as consulting or creating one's own proprietary database by gathering information using computer software (*computer-assisted reporting*).

In this regard, flowers and Salinas (2014, pp. 244) describe data journalism as a new tool aimed at improving the work in media newsrooms through the use of investigative journalism, knowledge, and the use of new IT tools and a new "good law" related to Access to Information. At the same time, they agree with the five phases that comprise data journalism, according to Giannina Segnini (2010), which are as follows: data collection, data cleansing (filtering or *scraping*), analysis (database processing), verification of sources, and display (through design tools).

From another point of view, Simon Rogers, former head of the Journalistic Data Unit of *The Guardian*, points out that work with large volumes of information "sometimes becomes a cure", as the journalist needs to sift through or narrow down these databases, analyze them, provide key data for every article, and look for the information that is best suited to each report so it is understandable to the reader (2013, pp. 16). At the end of the day, the true purpose is to tell the news story in the best way possible. For that reason, and in order to move toward a more visual and explanatory informative report that is increasingly demanded by new users (Newman, 2014), data journalism often benefits from the possibilities offered by handling large amounts of numbers for the purpose of presenting news stories with graphics or interactive displays (Crucianelli, 2013). These new visual tools have contributed to the rapid growth of data journalism in the past few years (Radcliffe, 2015: pp. 273).

However, rather than just gather and display data, this type of data journalism is actually a discipline of advanced statistics applied to information where the key is to analyze relationships between different variables by using scientific methodology in order to get a certain amount of data that is not usually shown through the use of conventional statistics.

As described by Steve Roig (2012), when we talk about data journalism we talk about social science with a time limit (*"Social Science on deadline"*). According to Howard (2014), such "application of *data science* to journalism" combines, nowadays, three elements: the processing of data that must be collected and validated as a complement to other sources of information (experts, testimonies from witnesses or official sources), the application of statistical science to ask questions and draw conclusions, and the presentation of results using graphic displays.

Nevertheless, the computerized analysis of statistical data to justify journalistic stories based more heavily on numbers as well as for the purpose of making predictions is not new. Sundar (2014) observed that some journalistic endeavors using these techniques come from work done by the Nieman Foundation at Harvard University in the decade of the 1960s of the last century, whereas accessibility to databases through the development of internet favored the emergence of this journalistic discipline in the early years of the millennium and is used in all newsrooms today.

Data journalism techniques have been applied to various specialized news domains, though not all of these areas have the same opportunities given the volume of uneven data that is handled by some of them. Sports journalism is one of the fields that is proving to have the greatest potential in this regard, due to the fact that within this area "these techniques are applied and many innovations are accomplished due to the large amount of data involved together with the help of many people with profound knowledge and interest in this area "(Flores and Salinas: 2014, 245 pp.).

2. Sports journalism based on data, a discipline with a long term future

Data journalism spread to the field of sports mainly due to early experiences in the sport of baseball in the United States of America, where complex scientific methods began to be utilized in order to analyze the statistics of the game and then move on to other sports such as basketball.

This trend has come into use in the last three years with the creation of data journalism units in a number of major media, especially those that are English-speaking. In this regard, 2014 was a watershed year in which *The New York Times*, *The Washington Post* and *The Guardian* took advantage of major international sporting events like the Winter Olympics in Sochi and the World Cup in Brazil to strengthen their information coverage by using new digital tools and by experimenting with narration and presentation of content.

This decision to pursue analytical and precision journalism through the use of data was also reflected in the launch in that same year of the general news *startup Vox.com*, created by former editor of *The Washington Post*, Ezra Klein, and the same

thing occurred within large media organizations such as *The New York Times* with the opening of their news section known as *The Upshot*, which also deals with the sports.

Most importantly, database journalism has demonstrated that sports is one of the fields with the highest potential for development in the coming years, especially after the experiences of specialized sites promoted by journalists such as Nate Silver (*FiveThirtyEight.com*) or Bill Simmons (*Grantland.com*), both affiliated with the digital conglomerate US broadcaster ESPN [1].

A particularly illustrative case is that of *FiveThirtyEight* which, before joining ESPN, was linked for three years to *The New York Times*, and at present acts as an innovative site that has been able to use similar scientific models for prediction and measurement of statistics for use in sports and other informative areas as diverse as politics or economics. According to Silver (2015), the analysis of data in data journalism is "a great opportunity", especially if we consider the large volume of statistics generated by competitions and the possibility of measuring these data from a qualitative point of view: "It's not just about *big data*, but something much more -*rich data*- that is to say, data that are accurate, precise and subject to rigorous quality controls."

As Silver explains, data in the world of sports are continuously increasing thanks in part to the existence of companies (a technological industry has already been created around it) that engage in monitoring, storage and the subsequent treatment of data, an example of which is the US company Stats LLC, or the British company Opta Sports, which offer services in real time, not only for media, but also for clubs, professional leagues and sports federations for their application in improving both the performance of athletes and match results as well.

In the case of the United States, specialized media are developing not only in data sports journalism but in specific sports as well, something still far from being fully developed in Spanish journalism. *Nylon Calculis*[1], a native digital media outlet specializing in NBA statistics, had an incredibly strong impact on the finals of the most important basketball competition in the world in late July 2016. In the sport of baseball the company of reference is *Fangraphs*[2], though its displays are quite unattractive, nearly precarious, compared to those of *Nylon Calculis*.

We must not forget the importance of the *Moneyball* book and movie of the same name, starring Brad Pitt, which tells how *big data* was applied for the first time in a professional manner to top-level American sports. Pitt portrayed Billy Beane, *general manager* of the Oakland Athletics, "who became famous for relying on statistics to build a team capable of winning 20 straight games in the 2002 season and finished with a record of 103-59"[3].

---

[1] http://nyloncalculus.com/

[2] http://www.fangraphs.com/

[3] http://www.elconfidencial.com/tecnologia/2016-03-18/el-algoritmo-que-acabara-con-las-discusiones-cunadas-de-futboldel-bar_1169806/

Data journalism and sports information have also found an important common ground in the British mainstream media, especially at *The Guardian* and *The Financial Times*, which in recent years have boosted their commitment to data analysis and display by creating specific work units known as Guardian Visuals and FT Data respectively. These units have interdisciplinary work groups resulting from the need to develop a variety of skills such as debugging data (*data wrangling*), statistics, web design and programming or digital mapping, and they have placed sports in the position of being one of the main fields of narrative experimentation for news storytelling from an angle that is both original and creative, and for the purpose of providing analysis and value to the information repertory offered regularly to the readers.

In this regard, the strategy of *The Financial Times* is particularly significant in understanding the development in the use of advanced statistics in texts and graphics that are designed not only to be shown on the web, but also on small screens of mobile devices and directly on social networks.

As described by one of its data journalists, John Burn-Murdoch (2016), one of the formulas by which an intelligent use of data can add value to the coverage of major sporting events is by telling a story visually through graphics, or by using statistics, which allow the reader to have all the tools necessary to become familiar with the participants in a competition and to follow it properly:

> "The current situation of digital news distribution dictates that the media often have to tell the story without the user even visiting their website. During an event like the Olympics, this need becomes more acute because many people follow them through social network sites to keep track of the dozens of events held there each day. A clear graph is a great opportunity to make the information stand out within such a large amount of data, and it can be fully understood in the same moment".

Judging by what is being done in Anglo media newsrooms, it seems that the revolution in advanced sports statistics is underway. Along with the commitment being made by the US media, and to a lesser extent the United Kingdom, other factors that may favor the popularity of using statistics in sports journalism are, on one hand, the availability of online data (every day more and more clubs and fans make statistics available to the public on websites), and on the other hand, the existence of a very active blogger community (*sportingintelligence.com*, *statsbomb.com*, *crispdata.co.uk*) with specialists in sports data analysis who in many cases have found themselves in demand by the media and ended up working for them.

And while this is happening in the United States and the United Kingdom, in Spain, with the traditional media moving toward digital development, those in charge of these media have started down the road to data journalism in newsrooms (Crucianelli and Zanchelli, 2015) through the creation of specific data analysis units. However, as Mar Cabra noted in the *Annual Report of the Journalistic Profession 2015*, "the fact of being thought of as supporting units hampers the development of genuine data journalism in the sense of being able to work with one's own stories, because often they are employed as auxiliaries to other sections: national, international, society, etc."

In the case of Spain and according to the same report, "despite the creation of specialized departments in some media and media groups, more than half of professionals say it is an activity that is not yet widespread, and these professionals work in media where this activity is not practiced". However, they believe that data journalism will find a way to expand further in so-called native digital media and specialized companies rather than in traditional media.

3. Research Design: case studies in Spain

Among teams that specialize in data analysis that have recently been established in Spanish digital media, those that stand out include the newsrooms of *El Confidencial* through its innovation laboratory, The Confidencial Lab, launched in 2013; and *El Español*, led by Pedro J. Ramirez, former director of *El Mundo*, and launched on October 7, 2015, although it functioned as a blog from early that year until its inception.

Thus, the issue is to learn the degree of relevance that has been acquired so far by journalistic work related to the subject of sports in these digital native news outlets, and also to determine whether or not techniques for advanced statistical analysis and visual representation of data have indeed begun to develop in this field.

3.1. Hypothesis

As such, our first hypothesis is that data journalism is beginning to emerge in sports reporting from Spanish native digital media such as *El Confidencial* and *El Español*, due to the fact that on one hand these two media companies are committed to innovative storytelling and have provided themselves with specific teams for data analysis, and on the other hand, sports journalism is an area of specialization with enormous opportunities for the development of this modality due to the fact that it has to handle a large volume of statistics daily.

3.2. Methodology

To prepare this study, we started with data on innovation and new formats that have emerged in sports journalism and which are regular targets of analysis in the digital media observatory *Periodismodeportivodecalidad* [2]. Based on this information, the methodology used to conduct this research was based on content analysis of reports with data regarding the most outstanding sports that were posted on the websites of the media mentioned for a period of one and a half years between February 2015 and July 2016.

For the purpose of being able to extract variables that are measurable and comparable, we first proceeded to identify the themes according to the type of sport in and out of the competition, journalistic genres (reports, previews, profiles and analysis), and sections where this type of press report (daily information or specials) are inserted. Secondly, the most

innovative contributions of these media in this field are analyzed, both in terms of presentation of content and the use of new digital narratives (long format texts, *scrollytelling*, use of infographics, data displays, interactive and multimedia elements), and especially the use of data (private databases or open source from other people) and the methodology and techniques used for the selection and treatment of these.

The process has been completed by another phase aimed at confirming and broadening the data obtained through various interviews conducted by using questionnaires given to those responsible for the sports and data analysis sections of the media studied: Jesús Escudero and Daniele Grasso for *El Confidencial*; Mario Diaz and Salva Carmona in case of *El Español*. In the specific situation of Salva Carmona, a specialist who after starting his career with *El Español* ended up working for *El Confidencial*, a second semi-structured interview was subsequently performed, this time more in depth.

3.3. Research objectives

Therefore, the research objectives are as follows:

- To learn whether or not journalism based on data is in a phase of expansion in Spain through the data analysis units created in recent years in media outlets such as *El Español* and *El Confidencial*.

- To verify the potential of data journalism in sports reporting and, within this area of reporting, the types of genres, themes and coverage.

- To have a good knowledge of the most innovative contributions of the work done in this field regarding the presentation of graphics and the use of new digital narratives by the media studied.

4. Results

4.1. El Español

In the case of *El Español*, in the year and a half since the objective sample analysis, 19 news articles have been written using data collection techniques and data visualization. This digital news media clearly demonstrates that the development of information data requires teamwork, and as such, approximately 90% of the articles tested were written by two authors: the news writer and analyst, in this case Salvador Carmona[4], and the computer graphics designer in charge of visual design, Luis Sevillano. Both are under the direction of Antonio Delgado, head of the Data and Visualization Department of *El Español*[5].

---

[4]   Carmona left El Español in February 2016
[5]   Delgado left El Español in May 2016

As for the topics addressed in this news outlet when working with data in sports, 17 of the 19 articles (nearly 90%) were related to football, while only two were related to basketball. In this regard, it is worth noting that most of the news containing data was limited purely to competition, that is to say, to recent events, and in ten cases (nine in football and one in basketball), the focus was on the performance analysis of a specific player.

Despite the fact that data journalism relates to recent events, the laborious nature of this type of news report, as compared to day to day news reporting, makes them more specialized new stories that act as preview to an important competition, (on nine occasions - more than 47% of the total) or an explanatory or informative report (53% of cases) with the objective of analyzing the performance of one player or an entire team.

It is also interesting to note the databases with which the people in *El Español* have worked in order to create these articles. Thus, we see how most of the reports have been developed from the combination of their own databases and other sources, such as official data from leagues like the NBA or the CBA, or the company Opta Sports, one of the main entities supplying statistical sports data in the world.

An example of this is found in the report regarding Madrid-Barcelona on 21 November 2015[6] (Figure 1). *El Español* noted that they used "advanced technology" for this data article, and together with their own efforts they used Opta Sports as a source [4] to create circumference graphics in the form of a spider web to analyse the determining factors in the match, such as shots on goal, percentage of goal conversions, ball recovery, passes between lines, assists, key passes, passes, ball steals and shots. The aim was to show the greatest number of possible information at a glance.

---

[6] http://www.elespañol.com/deportes/20151118/80242036_0.html

Figure 1

NEYMAR
F.C Barcelona
2015 - 2016
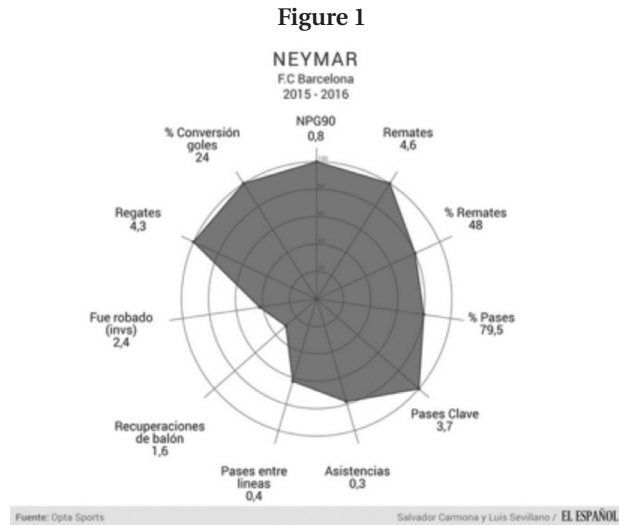
This data collection system was used with star players such as Messi, Neymar and Cristiano Ronaldo. In the case of Bale, a video was used instead of a spider web graph or radar. "The display is made to scale according to values using advanced metrics, giving marks to the player according to the standard deviation of the League for that value. For example, assuming 90% of pass completions for a defender, the display covers the entire framework of the radar, because the player is already a standard deviation above the rest. The display is made so that a perfect player (which rarely occurs), can occupy a complete sphere," explained Carmona at the end of his article regarding the methodology.

The insertion of a tab or methodological note, typical of the most advanced data works, as a manifestation of precision journalism, barely occurs in 40% of the *examples analysed* which indicates a lack of useful information for the reader that prevents him/her from being able to correctly interpret the display of various statistics that were chosen.
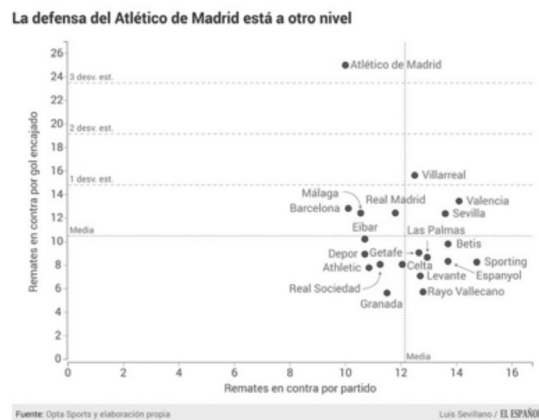
Another crucial aspect of this type of journalism in digital media is the degree of use of new narratives and the innovative nature of the graphics used to display data. In this sense, almost all of the articles analysed are supported by computer graphics (only one is supported by video). Regarding the graphics, it is worthy to point out that not only were conventional graphics used (bars and straight or curved lines to define trajectories but no pie charts or portions), as well as the latest graphics, such as the above-mentioned spider web graph (Figure 1), along with a display adjusted to the player's position in the field (contains the data of a player for 90 minutes, so as more space is covered inside the sphere, this indicates that his performance is better).

What stands out in the case of *El Español* is the incorporation of new data measurements applied to sports, that is to say, the intersection of a number of less common variables in conventional statistics. Such measurements include the rate of number of goals made per number of attempted shots, the goal scoring contribution (goals + assists) according to minutes played, or a defender's statistics adjusted to his team ball possession for the purpose of obtaining a more accurate measurement of his performance. This is how Salvador Carmona explained the methodology employed and the reason for crossing these variables in the article "Real Madrid, more reliable defensively than Barça", from 15 November, 2015[7]:

"After many years of studying football patterns, the correlation between clears completed/interceptions/tackles and shots fired by the opponent is downright low. To find the true measurement of the influence of these statistics, the figures have been adjusted according to the possession of the team, because it is not the same, for example, to make more interceptions when your team always has control of the ball in a match. This gives us a much more realistic view of the true contribution of a player within certain parameters in a game or a season."

Another interesting combined statistical measurement used to gauge the defensive level of a player or team is produced by comparing the average number of shots received by the defense per game with the number of shots needed to score a goal against that same defense. These are the variables that were applied, for example, in the article on 20 January 2016 entitled "The Atlético fortress, based on data" that contained the following caption: "Atlético de Madrid, who visits Balaídos in the Cup, confronts the second part of the season with the best defense in the league[8] (Figure 2).

Figure 2

La defensa del Atlético de Madrid está a otro nivel

Fuente: Opta Sports y elaboración propia   Luis Sevillano / EL ESPAÑOL

---

7   http://www.elespañol.com/deportes/20151115/79492081_0.html
8   http://www.elespañol.com/deportes/20160119/95740479_0.html

As for information with data that deals with basketball, different types of charts are also combined within the same article, such as heat maps, bars and lines, spider webs or radar and mapping of points (scatter plots). In this case, Win Shares (WS) can be found among the new applied statistical measurements that provide an estimation of victories that a particular player brings to the team together with the sum of points per one hundred possessions.

Another methodological innovation provided by *El Español* for data projects related to sports information compared to the methods that have been used up to the present in Spain is the utilization of the Elo methodology in the analysis of accumulated historical statistics. This method, invented by Hungarian mathematician Arpad Elo to assess the level of chess players in that period, uses an algorithm, which when applied to football compares the potential of teams, giving priority to past results in order to predict the future and the final scores between rivals.

As Salvador Carmona himself explains in the report entitled, "Barcelona arrives to the Classic in a better position, based on data", from 19 November, 2015[9]:

"Elo considers only the end result, not how a team plays or who plays, and rates all league matches in the same way, because in league matches all parties are equal. The scoring system only changes in single-elimination tournaments (Cup and Champions League), where the importance of a semi-final, for example, is greater than a group phase."

This is the methodology that was used in the case of the special report prior to the classic Real Madrid-Barcelona encounter in order to determine which team was in a better position when they arrived to the match (Figure 3):

Figure 3

Una rivalidad histórica
Puntuación Elo del Real Madrid y FC Barcelona a través de los años

■ Real Madrid CF
■ FC Barcelona

Fuente: Club Elo   Luis Sevillano / EL ESPAÑOL

Puntuación Elo histórica de Real Madrid y FC Barcelona / Elaboración propia

---

9   http://www.elespañol.com/deportes/20151119/80492024_0.html

El Español admits that this is the line of work with data that they want to continue expanding to other sports beyond football and basketball, although they realize that in sports such as handball or volleyball there is still not enough statistical development in their respective leagues that can provide "truly valid data for this type of analysis," says Mario Diaz, head of the Sports section.

"The world of sports has a dose of statistics, a bit of business, and a very high percentage of passion. These three elements offer a cocktail mix that is impossible to analyse 100%, either by analysing each element separately or even by using all three elements together. Luckily for some, but not for others, the X factor in sports, something we might call random, is what makes high level competition so attractive to the masses," says Díaz. *El Español* believes that at the present time there is no need to devote effort to statistical information on a daily basis, "due to the time required for the collection, selection and analysis of data."

Since the departure of Salvador Carmona from sports news writing, *El Español* has slowed considerably in the area of analysis and data visualization in this field. As such, there were only three of these types of reports between February and May 2015. The pieces were written by both Pedro Cifuentes and Patricia Lopez, one of the data unit specialists at this online newspaper.

The first two of these reports consisted of analyses written as a preview to football matches that involved highly intense rivalries (FC Barcelona-Real Madrid and Real Madrid-Atlético de Madrid), and were written by using accumulated statistics offered by Opta Sports and displayed in graphs[10], while the third report entitled "The most practiced sport in Spain is not *running*, but cycling"[11], which used the *Sports Habits Survey 2015* as a source, published by the Spanish Board of Sports to identify those sports in which most Spanish citizens participate.

4.2. El Confidencial

Meanwhile, between January 2015 and the end of the European Championship in France in July of 2016, two periods that are very different, the following can be observed in the data journalism work of *El Confidencial*: the first phase analyzed was operated by the data unit directed by Daniele Grasso and Jesús Escudero and has the mark of the *El Confidencial* data laboratory; a second phase, starting with Salva Carmona's collaboration in *El Confidencial* after his tenure with *El Español*, focuses on author-oriented data journalism in the form of a blog, but always under the coordination of the newspaper data unit .

---

[10] http://www.elespañool.com/deportes/futbol/20160401/113988641_0.html
[11] http://www.elespañol.com/deportes/ciclismo/20160418/118238526_0.html

In the period under review, *El Confidencial* produced 14 articles regarding data sports journalism. The first four articles were developed exclusively by *El Confidencial* Lab and several editors of the media. From March 2016 onward, Salva Carmona has been the only author of these articles commissioned by the head Sports Editor, Kike Marín

On the topic of data journalism applied to the sports field, 100% of the articles published in El Confidential focused on information about football. In addition, most of the published articles were limited to regular competition (League: 2 articles; Champions League: 2 articles; Eurocup: 4 articles). In 7 cases, the purpose was to analyze the performance of an athlete, nearly the same as a personal profile of the player, with the analysis of Cristiano Ronaldo standing out, both in his role as a Real Madrid player in League competition, Champions League and the Portugal National Selection during Eurocup 2016, with up to 3 articles of analysis dedicated to him.

As in the case of *El Español*, the data articles in *El Confidencial* sports section are related to current events, but there is also room for planned special events such as an important match (4 times) or an explanatory release or report from an economic point of view (3 of the 4 articles from the period prior to the incorporation of Salva Carmona), with the latter being one of the newspaper's specializations outside the sports field.

Jesús Escudero, one of the people responsible for the El Confidential Data Unit, believes that the introduction of data journalism in Spain is due to three factors: imitation of data reports published by US media, British media sources, and "the outstanding work of civic institutions unrelated to the traditional media world but which carry out top quality data journalism, one example being the Civio Foundation". The head of the *El Confidencial* Data Unit appreciates the great potential of data journalism, but warns of the "risk of turning information and media articles into reports plagued with numbers and statistics, distancing the essential information from the reader."

There are no topics or genres more likely to be used by data journalism. Of course, it is not so much the exploitation of existing databases as the development of ad hoc databases that did not previously exist "and with further analysis these will provide insight into situations unknown to citizens".

Figure 4
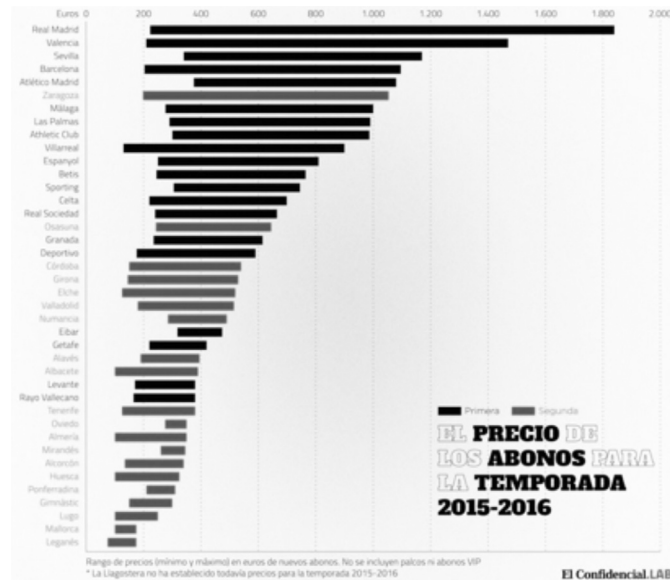
Most of the data articles produced by *El Confidencial* have been differentiated by the development of single databases that are created after collecting scattered data.

There are two examples that have been developed by this newspaper: during the 2015-2016 football pre-season, the annual season tickets from the First and Second Football Divisions were compiled in order to analyze where it is more expensive and more economical to watch professional football in Spain (Figure 4).

Another notable project that was carried out involved searching in libraries to find the sponsors of football shirts since 2000 in order to analyze the economics of marketing connected to the game of football (Figure 5). Precisely along the same lines they have created a database of football players' agents and those they represent in order to know "who belongs to each".

Figure 5
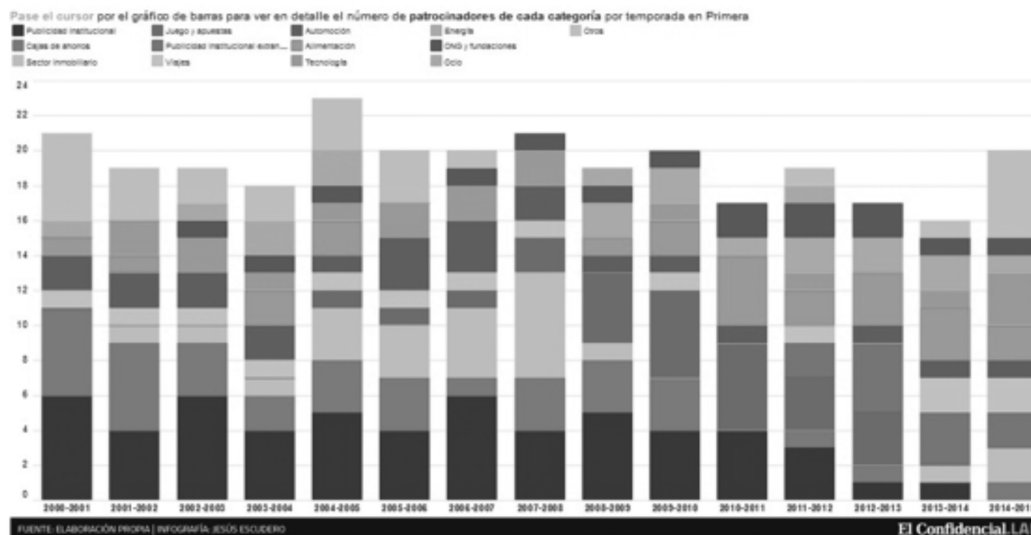
El Confidencial opts for this kind of informative article more than for the analysis of game statistics. This can be seen in topics and titles such as the following: "Cristiano equals Raul as the top all time scorer of Real Madrid"[12].
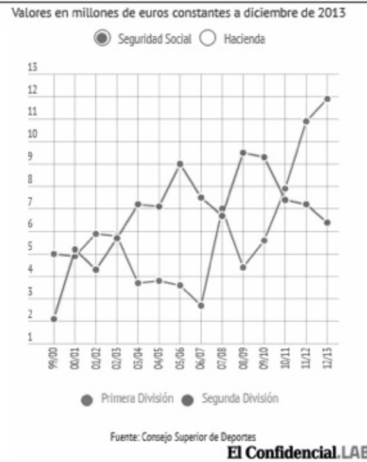
In each complex job completed by the Data Unit, the exact day of the update and the sources used are offered. In this case, there are five as follows: La Liga, UEFA, Cero a Cero, BDFútbol and Wikipedia.

One of the renowned specializations of *El Confidencial* is economic and financial journalism. The Data Unit of the newspaper takes this factor into account as well as the reader profile of the so-called "newspaper of influential lectors," even though the articles are supported but not written by the Unit. This is the case of the report entitled "Football's public debt, 738.5 million Euros that will not be paid until 2021", published on 22 January 2015. This is an article crafted by *Desmanesdelfutbol.com*, which became the winning project of the II Data Journalism Workshop: Following the Public Money Trail, organized by Medialab-Prado and *El Confidencial*[13].

---

[12] http://www.elconfidencial.com/deportes/futbol/2015-09-30/cristiano-ronaldo-real-madrid-raul-goleador-historico-malmoe_1024492/

[13] http://www.elconfidencial.com/deportes/futbol/2015-09-30/cristiano-ronaldo-real-madrid-raul-goleador-historico-malmoe_1024492/

Figure 6

Así ha evolucionado la deuda de los clubes de fútbol

Valores en millones de euros constantes a diciembre de 2013

◉ Seguridad Social ○ Hacienda

● Primera División ● Segunda División

Fuente: Consejo Superior de Deportes

El Confidencial.LAB

The combination of quantitative and qualitative elements "promote more complete information," says Escudero, who highlights as an advantage the fact that the collection and analysis of data in an article "is based on quantifiable information rather than on perceptions, beliefs or statements." The analysis of data provides headlines, but afterwards, it is the journalistic work and reporting competence that adds context and personal stories to the data analysis.
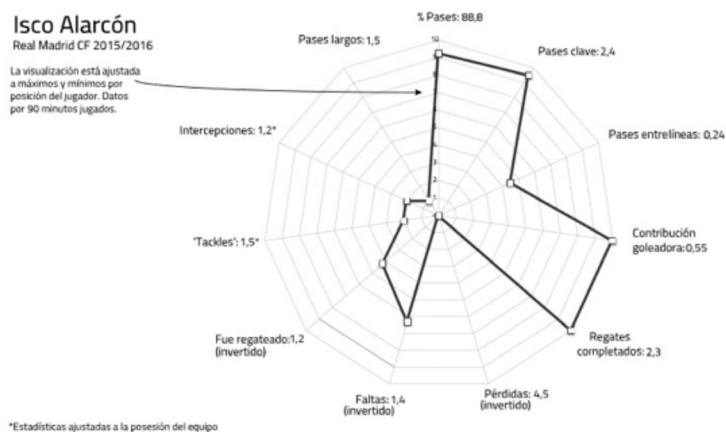
The classic 6Ws of journalism are not all given an answer in this Data Unit. "Answers are given to who, what, where and when, but they will never respond to why and how questions, which are usually the ones most relevant to society and the ones seeking causes and consequences of a given reality. If questions of why and how are not important enough, you can in fact tell a story with just data. However, if why and how are important, a data-only story will fall short".

Salvador Carmona, a highly regarded specialist in sports data journalism, began collaborating with *El Confidencial* on March 16, 2016 through a blog called "El Jurásico". His first analysis was called "Isco, the protected"[14]. In the article he showed how the Real Madrid player from Malaga was a person "who had a high rate of lost ball possessions", averaging 4.5 losses for every 90 minutes of play." Using data from Opta Sports and his own data as well, he analyzed among other parameters the percentage of passes, goal scoring contribution, fouls and the aforementioned losses (Figure 7).

---

[14] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-03-16/isco-real-madrid-bernabeu-florentino-perez-zidane-malaga-cesc-fabregas_1169732/

Figure 7

Isco Alarcón
Real Madrid CF 2015/2016

La visualización está ajustada
a máximos y mínimos por
posición del jugador. Datos
por 90 minutos jugados.

% Pases: 88,8
Pases largos: 1,5
Pases clave: 2,4
Pases entrelíneas: 0,24
Contribución goleadora: 0,55
Regates completados: 2,3
Pérdidas: 4,5 (invertido)
Faltas: 1,4 (invertido)
Fue regateado: 1,2 (invertido)
'Tackles': 1,5
Intercepciones: 1,2*

*Estadísticas ajustadas a la posesión del equipo

"Isco tends to play with a lower level of performance than his peers when his team needs him the most in giving a high performance." The fact is that he was below the level of all his rivals who play in the same position in key matches: against Özil (Arsenal), Pjanic (AS Roma), Arturo Vidal (Bayern Munich) or Rakitic (FC Barcelona). This article had a huge impact on *El Confidencial* readers. In fact, it was the most-read article in the newspaper that day, which explains reader interest in well-crafted and 'slow-cooked' sports reports.

"Klopp, the Simeone of Liverpool"[15] was the title of the game analysis of the legendary English team led by the German coach since the eighth day of the Premier League during the 2015-2016 season. Carmona dissects the differential between shots and goals made by Liverpool and how these data were rewarded: The team plays two finals in just a couple of months.

Salvador Carmona believes that data journalism is associated with a "halo of quality" from which it benefits. "I won't say that isn't true, because top management sees it that way, but it is a huge amount of work, and nowadays there are very few people who can do this on a daily basis." In his view, the problem lies in finding reporters who specialize in data and are at the same time within the information area of sports.

The development of data in sports journalism is still in its embryonic stage. Carmona himself has promoted a tool based on an algorithm that is able to quantify "the fairest result in a match depending on the shots on goal of each team" as Jaume Esteve said in he report entitled "The algorithm that will end the meaningless disagreements in football[16]"..

---

[15] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-05-18/klopp-simeone-liverpool-liga-europa-sevilla-league_1201608/

[16] http://www.elconfidencial.com/tecnologia/2016-03-18/el-algoritmo-que-acabara-con-las-discusiones-cunadas-de-futboldel-bar_1169806/

Carmona's code gathers all the shots that occur during the 90 minutes of play and are given a certain value depending on the proximity to the football goal post, the situation of other players, the position of the goalkeeper or whether or not the ball has gone out of play or been cleared. "I don't know if this could be of value in such a general-interest media outlet such as *El Confidencial* or *La Vanguardia*" [where he has also collaborated since March 2016]. "First you have to write a post to explain the model", he says. In La Vanguardia he prepares not only football articles within the Vang Data department, such as 'Bale and Neymar, the hidden players of the Tridents'[17],, or 'Morata should never have left Real Madrid, as his numbers were undisputed'[18], or the report in basketball entitled, 'Xavi Pascual was reviled despite having a higher percentage of victories than Pablo Laso[19].'

In his article covering the analysis of the preview of the Champions League final between Real Madrid and Atlético de Madrid, the journalist applied the Elo rating system in order to take into account the rivals who competed against each of the finalists, their scoring, and the final results of the games played in the competition. "When two teams compete in a match they exchange points based on the final score and the location of the game – home team, visitors or a neutral field. If the competition is an elimination match, the difficulty of the game is also taken into account in assigning points (a semifinal is worth more than a quarterfinal, which in turn is worth more than the 16 round). This system allows for a very rough estimate in comparing teams from different leagues."

This is the case in the article entitled, "The road to the final for Atlético is the most expensive in history; that of Madrid is the simplest"[20]

Figure 8

El camino a la final de Real Madrid y Atlético de Madrid
Para la elaboración de este *ranking* se ha utilizado el sistema de puntuación ELO

Fuente: Club ELO   Salvador Carmona | ElConfidencial.**LAB**

---

[17] http://www.lavanguardia.com/vangdata/20160401/40801251236/barcelona-real-madrid-bale-neymar-clasic.html

[18] http://www.lavanguardia.com/vangdata/20160623/402705036725/morata-real-madrid-eurocopa-2016.html

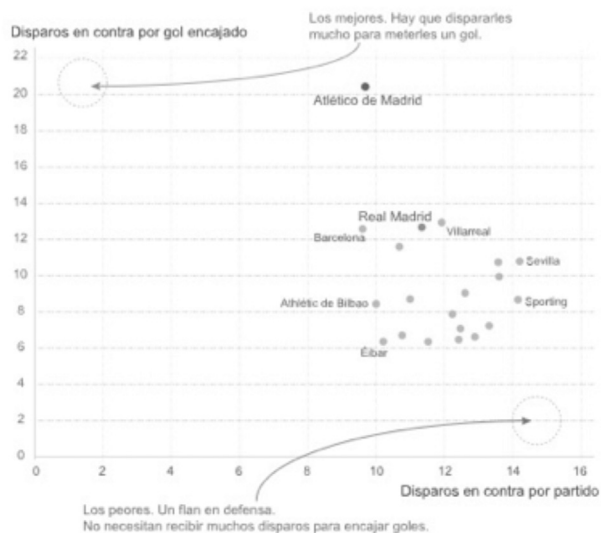[19] http://www.lavanguardia.com/vangdata/20160429/401451882947/pascual-laso-victorias-baloncesto.html

[20] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-05-26/final-champions-real-madrid-atletico-rivales_1206406/

The graph shows the difference between the teams. The only rival of Real Madrid that has exceeded 1,800 points was Manchester City. However, Atlético Madrid had two great rivals, FC Barcelona and Bayern Munich, both of which exceeded 2,000 points.

"In-depth study of a final: the best defense in Europe against the Zidane Effect"[21], was the second analysis by Carmona for the 2016 final of the Champions League (Figure 9). The great defensive ability of Simeone's team is reflected in the graph, which shows that the opposing team had to attempt a high number of shots in order to make a single goal.

Figure 9

Fortaleza en defensa de los dos aspirantes a la Champions

El Confidencial also carried out several data studies for the Eurocup, not only related to the national selections, but regarding the protagonists of the championship as well.

In the first article, entitled "Analysis of a defeat: Italy beat Spain both in theory and in practice"[22], Carmona dissected the failure of the selection of Vicente del Bosque. "The end of an era", wrote the author himself.

---

[21] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-05-27/final-champions-real-madrid-atletico-zidane-simeone-estadisticas-datos_1206989/

[22] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-06-28/eurocopa-espana-italia-octavos-conte-del-bosque_1225011/

As is common in his works, he explained the methodology and tools used. This article was based on Expected Goals (xG or ExpG), an advanced metrics system that dissects the quality of opportunities that take place in a match, as explained in detail in his own personal blog (Carmona, 2016). Each shot has a value based on the probability that the shot will end in a goal. The value is assigned based on the location of the shot, the distance to the goal, the angle, the body part (head or foot) that hits the ball, and the speed of the move or of the preceding pass.

Figure 10
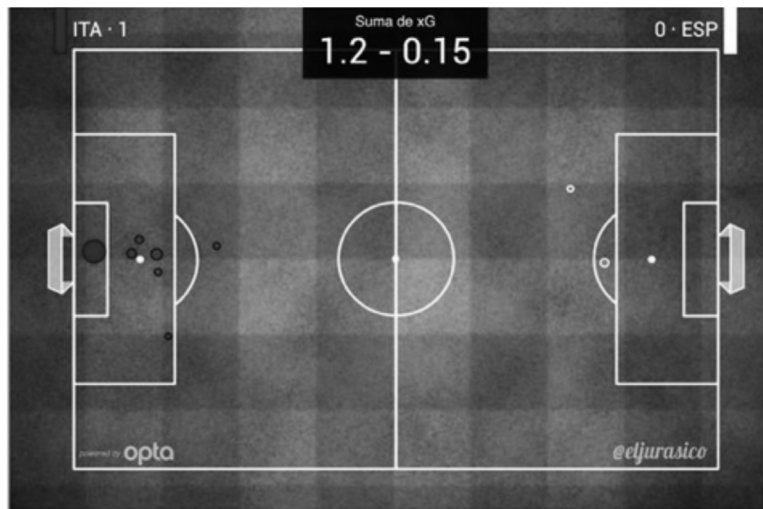
Ocasiones en el Italia-España al descanso. El tamaño indica la calidad de la ocasión.

Salva Carmona has studied the cases of players like Kanté, a former Leicester player and number one in Europe in ball recovery. In his analysis[23], he applies the Elo system and graphics on Twitter by David Sumpter, professor of Applied Mathematics and author of the book *Soccermatics: Mathematical Adventures in the Beautiful Game* (2016).

The case of Cristiano Ronaldo, one of the media's highest profile players who attracts worldwide attention from the football audience, was studied in-depth in two articles that appeared in *El Confidencial* during the Eurocup. In the first one, entitled "Why Cristiano Ronaldo has accelerated towards the Ballon d'Or in this UEFA Euro 2016[24]" the author emphasizes the individual performance of the Real Madrid player in the European competition, where he went from less to more. In the

[23] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-06-25/kante-pogba-francia-deschamps-espana-real-madrid_1222122/

[24] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-07-08/cristiano-ronaldo-gareth-bale-real-madrid-portugal-gales-eurocopa_1229707/

second of the articles regarding the star of the Madrid club, "The hidden power of Cristiano and his youth, keys to an unexpected champion"[25], Carmona once again applied the *expected goals* advanced metrics system.

Carmona, who also works with analytical articles involving data sports journalism at *Jot Down Magazine*, believes this type of journalistic work will grow thanks to the Spanish pool of young, talented computer graphics professionals who are working for leading international media organizations. He cites as examples Alberto Cairo, Knight Chair at the University of Miami, and Chiqui Esteban, assistant director of the Graphics Department at *The Washington Post*, global experts in creative visualization. "My big weakness is visualization. That's where I have to improve" he says, regarding a specialization where diverse profiles come together, such as journalists trained in computer graphics and design or data analysts who end up working in the media.

5. Conclusion

The results of the investigation show that sports reporting is, in fact, becoming a fertile ground for the growth of data journalism in Spain, and digital media outlets such as *El Confidencial* and *El Español* are leaders in this new way of working in newsrooms. Similarly, data sports journalism is in expansion in Spain thanks to the work done by specific data units created in recent years.

The analysis of data articles studied in both of these media outlets also indicates that data journalism is fundamentally a team effort, and as such, its development depends not only on journalists, but also on graphic designers, programmers, statisticians or even mapping experts in order to create visual displays.

Nearly all of the works regarding data sports journalism carried out by *El Español* and *El Confidential* deal with football, especially with regard to competitions like the Spanish National Football League, the Champions League or the European Championship. In this sense, the Spanish media analyzed in this study adopt patterns of Spanish sports journalism in terms of their topic priorities.

To a lesser extent, the data work from both media focused on analyzing the performance of an athlete, which is very similar to a profile story. While articles using data are related to the present, there is also room for special reports that are previously planned regarding an important meeting, an explanatory story, or in case of *El Confidencial*, an article from the economic point of view.

In a high percentage of data journalism reports from the two Spanish digital outlets, the Sports section has chosen to work with their own databases as well as statistics provided by Opta Sports, the latter being one of the most frequently used sources that are employed in creating such reports.

---

[25] http://blogs.elconfidencial.com/deportes/el-jurasico/2016-07-12/portugal-eurocopa-cristiano-pepe-francia_1231456/

Regarding innovations, the figure of analyst Salvador Carmona stands out, due to the fact that he worked for *El Español* early in his career and then later at *El Confidencial*, and he has applied advanced statistical methodologies to the field of sports, which until very recently had not been used in Spain, examples of which are the *Elo* model, or the *Expected Goals* advanced metrics system based on a balance of probabilities.

The main challenge facing Spanish sports journalism is to consolidate this trend, which started a few years ago in other journalistic areas of the world, especially in Anglo-Saxon countries, and to expand technical analysis and data visualization to other sports beyond football. Possible limitations to be overcome are the continuing lack of enough statistical databases available to the media for other sports, the overspecialization of journalists in other sports, or the training and professional competence taught at universities for the new professional profiles required by media companies.

Notes

1. This site, created in 2011, was closed by ESPN on October 31, 2015.

2. This blog, created by Professor and Researcher José Luis Rojas Torrijos in 2010 has included studies on innovation and new trends in journalism in general, and in sports journalism in particular.

 The address is: http://periodismodeportivodecalidad.blogspot.com.es/

3. Publishing products in data journalism offered by Opta Sports are the following: previews, analysis and match reports; predictions; news; sports announcing (minute by minute) in real time, automatic and/or manual; game data and interesting details about any other match in the competition; data and graphic analysis and rankings of teams and players.

Bibliography

APM (2015): *Informe anual de la profesión periodística 2015*. Madrid: APM.

Burn-Murdoch, J. (2016): "There's a golden opportunity for sports journalists to add more quantitative analysis to their repertoire to better serve their audience". En *Periodismo deportivo de calidad*: http://periodismodeportivodecalidad. blogspot.com.es/2016/03/burn-murdoch-theres-golden-opportunity.html

Carmona, S. (2016): "Introducción a Expected Goals (xG)". [Disponible en: http://scarmona.org/introduccion-a-expected-goals-xg/]

Crucianelli, S. (2013): "¿Qué es el periodismo de datos?", *Cuadernos de Periodistas*. Asociación de la Prensa de Madrid. [Disponible en: http://www.cuadernosdeperiodistas.com/que-es-el-periodismo-de-datos/]

Crucianelli, S. y Zanchelli, M. (2015): *Integrando el periodismo de datos en las salas de redacción. Informe del International Center for Journalists*. [Disponible en: http://www.icfj.org/sites/default/files/integrating%20data%20journalism-spanish.pdf]

Doig, S. (2012): "Social science done on deadline", *journalistsresource.org* [Disponible en: http://journalistsresource.org/tip-sheets/research/research-chat-steve-doig-data-journalism-social-science-deadline]

Esteve, J. (2016): "El algoritmo que acabará con las discusiones cuñadas del fútbol", *El Confidencial.com*. [Disponible en: http://www.elconfidencial.com/tecnologia/2016-03-18/el-algoritmo-que-acabara-con-las-discusiones-cunadas-de-futboldel-bar_1169806/]

Flores Vivar, J.; y Salinas, C. (2014): "El periodismo de datos como especialización en los medios generalistas con presencia en internet", en Esteve, F.; y Nieto, J. C. (Coords.): *Nuevos retos del Periodismo Especializado*. Madrid: Schedas, pp. 241-260

Harcup, T. (2014): *Oxford Dictionary of Journalism*. Oxford: Oxford University Press.

Howard, A. B. (2014): *The Art and Science of Data-Driven Journalism*. Nueva York: Tow Center for Digital Journalism. [Disponile en: https://towcenter.gitbooks.io/the-art-and-science-of-data-driven-journalism/content/]

Marrero Rivera, O. (2011): *Fundamentos del periodismo deportivo*. San Juan: Terranova.

Newman, N. (Coord.) (2015): *Reuters Institute Digital News Report 2015*. Oxford: Reuters Institute for the Study of Journalism. [Disponible en: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Reuters%20Institute%20Digital%20News%20Report%202015_Full%20Report.pdf]

Radcliffe, D., Felle, T. y Mair, J. (2015): *Data Journalism: Inside the Global Future*. Bury St. Edmunds: Abramis Academic Publishing.

Rogers, S. (2013): *Facts are sacred. The power of data*. Londres: Guardian Books.

Rivera, A. y Rojas Torrijos, J. L. (2016): "El periodismo deportivo de datos en El Confidencial, El Español y ABC.es", en Blanco, E. y Quesada, M. (Coords.): *Periodismo de datos*. La Laguna: Latina de Comunicación, Cuadernos Artesanos de Comunicación, n. 112, pp. 247-266. DOI: 10.4185/cac112

Rojas Torrijos, J. L. (2015): "Innovación en el periodismo deportivo: medios, formatos y narrativas", en García Avilés, J. A. (Coord.): *Innovar en Periodismo*. Murcia: Diego Marín, pp. 75-86.

Segnini, G. (2010): "El periodismo de datos aumenta la credibilidad y la independencia", *Consejoderedaccion.org*. [Disponible en: http://consejoderedaccion.org/noticias/item/89-giannina-segnini-el-periodismo-de-datos-aumenta-la-credibilidad-y-la-independencia.html].

Silver, N. (2015): "Nate Silver Comes Full Circle With Sports Data Journalism Mecca", *Sporttechie.com* [Disponible en: http://www.sporttechie.com/2015/08/17/nate-silver-comes-full-circle-sports-data-journalism-mecca/].

Sumpter, D. (2016): *Soccermatics: Mathematical Adventures in the Beautiful Game*. Londres: Bloomsbury.

Sundar, V. (2014): "The Power of Data Journalism", *Harvardpolitics.com* [Disponible en: http://harvardpolitics.com/covers/power-data-journalism/].