Data Journalism

Andreas A. Veglis
Aristotle University of Thessaloniki, Greece

Charalampos P. Bratsas
Aristotle University of Thessaloniki, Greece

INTRODUCTION

The introduction of ICTs (Information Communication Technologies) had a profound impact on every aspect of human activities. In the case of journalism, the utilization of ICTS has transformed the profession through the digitalization of the work process as well as the introduction of the internet along with its services (Veglis 2009). Today the journalist is expected to have the ability to firstly employ many tools and services in order to be instantly informed about breaking news as well as current events, and secondly, use a variety of tools and applications in order to prepare and disseminate news articles (Veglis & Bratsas, 2017). Many new types of journalism have emerged, among which, data journalism (Gray, Chambers, & Bounegru, 2012), which requires journalists to have special ICT skills.

In the recent years, data journalism has drawn significant attention in the academic literature as well as in the area of new developments in digital news production (Appelgrena & Nygren, 2014; Fink & Anderson, 2015; Mair & Keeble, 2014). Data journalism is considered to be a new form of journalism. It has appeared gradually in the dawn of the new century. Many factors have contributed to the introduction of data journalism, but one of the most prominent is believed to be the availability of data in digital form (Veglis & Bratsas, 2017). Data Journalism is a journalistic specialty reflecting the increased role of the numerical data has in the production and distribution of information in the digital era. Data can be the source of data journalism, and/or it can be the tool with which the story is told (Gray, Chambers, & Bounegru, 2012).

This chapter examines current trends and future perspectives of data journalism. The background section provides historic evolution and definitions of data journalism. Next, the stages of data journalism are presented in detail. Also, the relation between data journalism and open data is discussed due to the importance of the later in the development of data journalism. Finally, recommendations and future research direction are briefly discussed.

BACKGROUND

Evolution and Definition

Although the term data journalism started to attract attention at the end of the previous century, initial examples of data journalism appeared quite early. According to Simon Rogers the first example of data journalism was published at Guardian in 1821. It concerned the number of students who attended school and the costs per school in Manchester (Gray, Chambers, & Bounegru, 2012).

At the end of the 20th century, employing large data to write an article was difficult and required skills that went beyond the capabilities of the average journalist. That resulted in the phenomenon that some news organization in the United States and Great Britain were hiring programmers that worked on novel news products (Parasied & Dagiral, 2013). Traditionally, journalists used to rely on information provided by various sources

DOI: 10.4018/978-1-5225-2255-3.ch103

(governments, officials, research studies, etc.). Of course, there were some cases of investigative journalism where journalists were able to find resources to gather and analyze their own data and publish their results in articles (Veglis & Bratsas, 2017). But as a growing amount of data gradually became available online, and efficient tools with which anyone can analyze, visualize and publish large amounts of data appeared, things changed significantly (Sirkkunen, 2011).

The concept of data journalism in not new. It has been around since the beginning of the digitalization. Digital data has been utilized in news production since the late 60s in US newspapers (Parasied & Dagiral, 2012). Data journalism gradually emerged with the rapid introduction of ICTs and the availability of data in digital form. The term data journalism is synonymous with data-driven journalism while the older term, computer-assisted reporting has vanished since it was introduced at the early stages of computer history (Bradshaw, 2010). It is worth noting that in the case of data journalism there is an increased interaction between journalists and several other fields such as design, computer science and statistics (Thibodeaux, 2011; Veglis & Bratsas, 2017).

The term data journalism is attributed to Simon Rogers that first mentioned it in a post to the Guardian Insider Blog (Knight, 2015). It can be viewed as a process that begins with analyzing, and continues with filtering and visualizing data in a form that links to a narrative (Lorenz, 2010). It combines spreadsheets, graphics data analysis and the biggest news stories (Rogers, 2008). It is fundamentally the production of news graphics and includes elements of design and interactivity (Bradshaw, 2010; Lorenz, 2010; Rogers, 2008). Megan Knight (2015) describes data journalism as "a story whose primary source or "peg" is numeric (rather than anecdotal), or a story which contains a substantial element of data or visualization".

Veglis and Bratsas (2017) proposed a definition in order to better address the power of visualization and interactivity that are significant factors in data journalism. They defined data journalism as the process of extracting useful information from data, writing articles based on the information and embedding visualizations (interacting in some cases) in the articles that help readers understand the significance of the story or allow them to pinpoint data that relate to them.

Journalists' Skills

Today the journalist is expected to possess various ICT skills in order to cope successfully with the challenges in his everyday work (Veglis & Bratsas, 2017). Typical examples are writing news articles, constructing diagrams via spreadsheet applications, communication via e-mail, visualizing data with the help of various applications, publishing material on the WWW (Peebles 2011). Also, they often seek information on the web and by e-mail (Veglis, 2013). Veglis and Pomportsis (2012, 2014) organized the journalists' ICT skills into five categories, namely *Basic skills, Web publishing skills, Web 2.0 skills, Web casting skills and Data Journalism skills*. Veglis & Bratsas (2017) extended this categorization by adding a category that refers to *Web 3.0*. Next, the six categories are briefly presented:

- **Basic Skills:** The journalist has the ability to work efficiently with office automation suites (which include word processing, spreadsheet, presentation, database), and with the basic Internet services (WWW, email). Specifically, the journalist is expected to have basic typing and formatting skills, and he must be able to perform basic functions in a spreadsheet. He must have at least a general understanding of how to use data to support news stories and he must also be able to use relational database programs to cross-check those data files to find various information (Veglis & Pomportis, 2014).
- **Web Publishing Skills***:* In this case, basic knowledge of HTML is considered to be a necessary prerequisite, as well as the abil-