

# Data Journalism: Definition, Skills, Difficulties, and Perspectives

**Andreas Veglis**

 <https://orcid.org/0000-0002-0286-2304>

*Aristotle University of Thessaloniki, Greece*

**Charalampos Bratsas**

*Aristotle University of Thessaloniki, Greece*

## INTRODUCTION

The introduction of ICTs (Information Communication Technologies) at the last decade of the 20<sup>th</sup> century had a profound impact on almost every aspect of human activities. Journalism was one of the areas where ICTs had a significant impact, since it transformed the profession through the digitalization of the work process as well as the introduction of internet 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, 2017a). Many new types of journalism have emerged (algorithmic journalism, drone journalism, multimedia journalism, etc. (Diakopoulos, 2019; Ntalakas, Dimoulas, & Veglis & Bratsas, 2017a; 2017b; Tu, 2015)) along with data journalism (Gray, Chambers, & Bounegru, 2012), which require journalists to have special skills.

In recent years, data journalism has drawn significant attention in academic literature as well as in the area of new developments in digital news production (Hermida, & Young, 2017; Loosen, Reimer, & De Silva-Schmidt, 2017; Weber, Engebretsen, & Kennedy, 2018). Data journalism is now considered to be an established 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. Another factor that contributed to its introduction was the availability of visualization and data management tools (Veglis & Bratsas, 2017a). Data Journalism is a journalistic specialty reflecting the increased role that 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 attempts to describe the state of data journalism today. The background section provides the historic evolution and definitions of data journalism. Next, the stages of data journalism are presented in detail, along with the necessary corresponding skills of the data journalist. Also, the various difficulties that the evolution of data journalism is facing are presented. Finally, recommendations and future research directions 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 (currently data journalist and data editor at Google), the first example of data journalism was published at The Guardian in 1821, and 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 20<sup>th</sup> century, employing large datasets to write an article was difficult and required skills that went beyond the capabilities of the average journalist. To address these shortcomings news organizations in the United States and Great Britain began hiring programmers to work on novel news products (Parasie & Dagiral, 2013). Traditionally, journalists have relied on information provided by a variety of sources (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, 2017a). But things changed significantly as data gradually became more available online, along with efficient tools allowing anyone to analyze, visualize and publish large datasets (Gray, Chambers & Bounegru, 2012; Sirkkunen, 2011; Uskali & Kuutti, 2015).

*Big data*, the popular term for large datasets, was introduced into common usage during the last decade of the 20th century. It can be defined as datasets of a size which cannot be captured, curated, managed and processed by commonly used software running on standard personal computers (Snijders, Matzat & Reips, 2012). Is it quite obvious that there is a strong relation between big data and data journalism, since they have developed almost in parallel and data journalism is based/generated by big data (Veglis & Maniou, 2018).

The concept of data journalism is not new. It has been around since the beginning of the digitalization of news. Digital data has been utilized in news production since the late 60s in US newspapers (Parasie & 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 largely 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, 2017a).

The term data journalism is attributed to Simon Rogers who 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 (2017a) proposed a definition in order to better address the power of visualization and interactivity which 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.

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/data-journalism/260256](http://www.igi-global.com/chapter/data-journalism/260256)

## Related Content

---

### Ostra: A Process Framework for the Transition to Service-Oriented Architecture

Fabiano Tiba, Shuying Wang, Sunitha Ramanujam and Miriam A.M. Capretz (2009). *International Journal of Information Technologies and Systems Approach* (pp. 50-65).

[www.irma-international.org/article/ostra-process-framework-transition-service/4026](http://www.irma-international.org/article/ostra-process-framework-transition-service/4026)

### Distributed Autonomous Control Architecture for Intelligent Mobile Robot Systems

Gen'ichi Yasuda (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 6611-6620).

[www.irma-international.org/chapter/distributed-autonomous-control-architecture-for-intelligent-mobile-robot-systems/113122](http://www.irma-international.org/chapter/distributed-autonomous-control-architecture-for-intelligent-mobile-robot-systems/113122)

### WSN Management Self-Silence Design and Data Analysis for Neural Network Based Infrastructure

Nilayam Kumar Kamila and Sunil Dhal (2017). *International Journal of Rough Sets and Data Analysis* (pp. 82-100).

[www.irma-international.org/article/wsn-management-self-silence-design-and-data-analysis-for-neural-network-based-infrastructure/186860](http://www.irma-international.org/article/wsn-management-self-silence-design-and-data-analysis-for-neural-network-based-infrastructure/186860)

### Latent Dirichlet Allocation Approach for Analyzing Text Documents

Parvathi Chundi and Susannah Go (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 1819-1824).

[www.irma-international.org/chapter/latent-dirichlet-allocation-approach-for-analyzing-text-documents/112587](http://www.irma-international.org/chapter/latent-dirichlet-allocation-approach-for-analyzing-text-documents/112587)

### Diagnosing and Redesigning a Health(y) Organisation: An Action Research Study

Christoph Rosenkranz, Marcus Laumann and Roland Holten (2009). *International Journal of Information Technologies and Systems Approach* (pp. 33-47).

[www.irma-international.org/article/diagnosing-redesigning-healthy-organisation/2545](http://www.irma-international.org/article/diagnosing-redesigning-healthy-organisation/2545)