Envisioning the Future of Communication

Vol 2, No 1 (2025)

Envisioning the Future of Communication - Conference Proceedings vol. 2

Keynote Speech 2

Andreas Veglis

doi: 10.12681/efoc.7894

Keynote Speech 2

The evolution and prospects of data journalism

Andreas Veglis*

Abstract

Over the past 20 years, there has been remarkable growth in the field of data journalism. Data journalism is a specialization that relies on finding news stories within data. It typically involves a visualization that can make the complex results of an article understandable. Today, there is a trend in data journalism to use programming techniques (usually utilizing the Python programming language). This trend mainly stems from large journalistic organizations involved in big data journalism projects that use big data analysis. The chapter examines the evolution of data journalism from its inception to the present day. It also discusses the necessity of using programming for practicing data journalism, the available methods, and the difficulties encountered by journalists in using programming. The chapter concludes with a reference to future developments in the field of data journalism.

Keywords: data journalism, visualization, programming, Python, future developments.

Introduction

The introduction of ICT (Information and Communication Technologies) in the last decades of the 20th century had a significant impact on almost every industry, including the news industry. Journalism was one of the fields where ICT had a substantial impact, transforming the profession through the digitization of the work process and the introduction of new online content transmission services (Siapera & Veglis, 2012; Veglis, 2009). Today, journalists can use numerous tools and services to stay informed about breaking news and current events and to use a variety of tools and applications for the preparation and dissemination of news (Veglis & Brasas, 2017a). Concurrently, the introduction of ICT has significantly altered the way news is consumed by the public. New content-sharing channels have emerged, and media outlets are leveraging these channels to create alternative ways of distributing journalistic content (Siapera & Veglis, 2012). It is worth noting that many new types of journalism have emerged (algorithmic journalism, drone journalism, multimedia journalism, etc.) (Gray, Chambers, & Bounegru, 2012), which require journalists to possess specialized skills.

In recent years, data journalism has attracted significant attention in academic literature as well as in the field of new developments in digital news production (Hermida & Young, 2017;

* Professor, Media Informatics Lab – School of Journalism & MC, Aristotle University of Thessaloniki, veglis@jour.auth.gr

Loosen, Reimer, & De Silva-Schmidt, 2020; Weber, Engebretsen, & Kennedy, 2018). Data journalism is now considered an established form of journalism. It gradually emerged at the start of the new century. Many factors have contributed to the introduction of data journalism, but one of the most important is believed to be the availability of data in digital form. Another contributing factor was the availability of data visualization and management tools (Veglis & Bratsas, 2017a). Data Journalism is a journalistic specialty that reflects the increased role numerical data plays in the production and distribution of information in the digital age. Data can be the source of data journalism and/or it can be the tool through which the story is told (Gray, Chambers, & Bounegru, 2012).

The report on the state of data journalism by the EJC (2024) highlights significant trends and challenges in the field. The majority of data journalists are employed full-time, reflecting a stable job market, although there is a notable shift towards freelance work, indicating a trend towards more flexible working arrangements. Technical skill shortages, particularly in data analysis and visualization, remain a challenge despite increased fundamental journalistic skills. Public government data is used more frequently, emphasizing its crucial role, with usage varying by region due to different legal frameworks. The integration of artificial intelligence is gradually increasing, primarily for content verification and search, but it faces issues related to ethical concerns and a lack of understanding regarding these technologies.

The introduction of data journalism has significantly increased journalists' needs in data management. On this issue, there are currently two approaches. The first suggests using independent applications that support different stages of data journalism (data finding, cleaning, verification, combination, etc.), and the second involves using programming to perform the necessary processes for practicing data journalism (Porlezza, 2024).

This chapter attempts to describe the current state of data journalism. Section 2 discusses the historical evolution of data journalism. The next section provides definitions of data journalism and outlines the stages that comprise it. Section 4 addresses the issue of the two approaches to practicing data journalism. The available solutions for adopting programming in the practice of data journalism are presented in Section 5. The chapter concludes with Section 6, where future developments in the field of data journalism are discussed.

Historical Evolution of Data Journalism

First Generation of Data Journalism (Data Journalism 1.0)

The origins of data journalism can be traced back to the work of John Graunt, who in 1662 published "Natural and Political Observations Made on the Bills of Mortality." Graunt utilized rudimentary statistical methods to analyze the list of deaths in London, thus laying the groundwork for statistical and demographic analysis (Friendly, 2007). Worth mentioning is also the inaugural issue of the Manchester Guardian in May 1821 (see Figure 1). This edition contained a data table - an innovation for that time - listing schools in Manchester and Salford,

with details on the number of students and average annual expenses, providing significant insights into access to education. These data showed that 25,000 children, as opposed to the official estimate of 8,000, received free education, thus questioning the accuracy of the data collected by the four clergymen responsible for these official statistics (The Guardian, 2011).

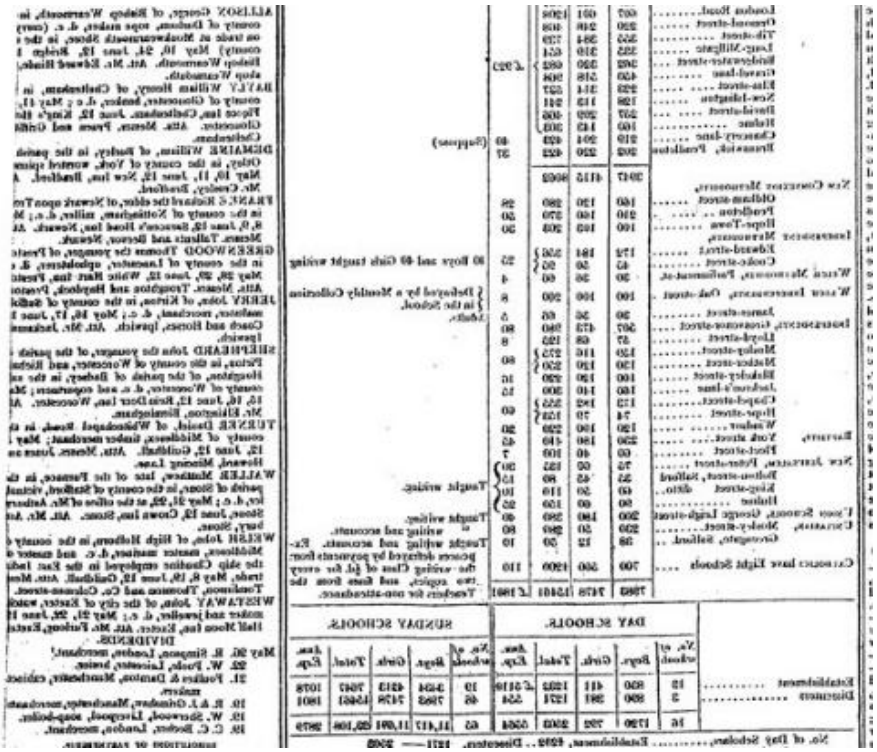

Figure 1: Article in the Manchester Guardian in 1821
(https://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian)

The revelation of data in 1821 was not merely informative but served a crucial socio-political function. It exposed shortcomings in the educational system and highlighted the disparity between official figures and reality. This act of publishing the dataset was a significant move towards transparency, offering a more accurate representation of educational access during that period.

The 19th century saw the expansion of statistical graphics and the first uses of data visualizations in newspapers. William Playfair, a Scottish engineer, was pivotal during this time for inventing several types of graphs, such as the bar chart, pie chart, and line graph. His graphs were later used in journalism to present economic and social data to the public, making complex information more accessible and understandable (Spence, 2005). This era is characterized by the work of Charles Joseph Minard, renowned for his 1869 flow map of Napoleon's Russian campaign in 1812. His innovations demonstrated the power of visualizing temporal and spatial data (Tufte, 2001).

Articles belonging to the first generation of data journalism (Data Journalism 1.0) continue to appear to this day. A characteristic example is in Thailand, where the Isranews agency, in an article covering a company, cites financial data without engaging in any statistical analysis (Thienthaworn, 2021).

Second generation of data journalism (Data Journalism 2.0)

At the beginning of the 20th century, technological advancements in printing and the rise of mass media facilitated the wider use of graphics. Newspapers began to include more detailed charts and graphs to accompany the news. By 1943, data presentation had become much more creative and targeted the average newspaper reader. During that time, The Manchester Guardian featured an article with an infographic (see Figure 2), created by the United States War Information Office to reassure American citizens. At that time, rumors were circulating that most of the military equipment and food produced in the USA were being sent abroad, so this visualization was created to debunk these rumors. It is easy for the reader to see, at a glance, that the solid icons represent exports, and the hollow icons represent materials remaining in the USA (Thienthaworn, 2021).

Figure 2: Data Visualization in a Newspaper Article in 1943 (https://www.theguardian.com/gnmeducationcentre/gallery/2021/aug/13/the-evolution-of-data-journalism-in-pictures)

In the 1960s, the gradual use of electronic computers for composing journalistic content began. This resulted in journalists using computers to analyze large datasets - a practice implemented by Philip Meyer during the Detroit riots in 1967. Meyer's use of survey data in his analysis of the riots for the Detroit Free Press is often cited as one of the early examples of modern data journalism (Meyer, 1973). The skills required for this type of data journalism include social research methodology and computer usage (Thienthaworn, 2021).

Third generation of data journalism (Data Journalism 3.0)

In the late 20th century, the convergence of information technology and communication had significant implications for data journalism. Specifically, in the 1990s, pioneering efforts were made, such as the establishment of research units specializing in data analysis, such as the BBC's News Lab and The Guardian's Data Blog (Coddington, 2015). The rise of the Internet and the development of advanced data analysis software enabled journalists to handle larger datasets and perform more complex analyses. This period also saw the introduction of interactive visualizations and the beginning of online data journalism. An example project that can be considered the starting point of the third generation of data journalism is The Guardian's "Afghan War Logs," which used data science to assist in collecting, organizing, analyzing, and presenting over ninety thousand classified military documents. The skills required for this type of article include programming and the development of interactive websites to enable audiences to interact with the article's content (for example, through interactive visualizations).

Defining data journalism

Megan Knight (2015) attributes the first use of the term "data journalism" to The Guardian journalist Simon Rogers in a post on the Guardian Insider Blog. This concept involves a workflow starting with data analysis, followed by data filtering and visualization to create a narrative, as Lorenz (2010) states. It incorporates elements such as spreadsheets, graphics, data analysis, and significant news (Rogers, 2008), essentially involving the creation of news graphics, incorporating design elements and interactivity (Bradshaw, 2010; Lorenz, 2010; Rogers, 2008). Megan Knight (2015) defines data journalism as a narrative primarily based on numerical data or including significant data or visualization.

Veglis and Bratsas (2017a) proposed a definition to more effectively include the importance of visualization and interactivity in data journalism. They describe it as a process of extracting valuable information from data, shaping articles based on this information, and integrating visualizations (sometimes interactive) into these articles to help readers understand a topic or focus on information relevant to them. They also proposed categorizing the data journalism process into six distinct stages (see Figure 3), namely data collection, data cleaning, data understanding, data evaluation, data visualization, and article writing (2017b; 2016).

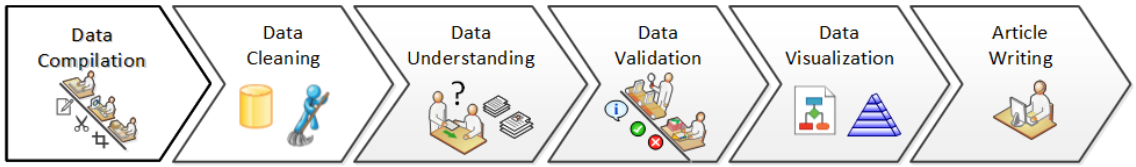

Figure 3: Data journalism stages (Veglis & Bratsas, 2017a).

Data Collection: This initial phase of a data journalism project begins either with a question that requires data or with a dataset in need of exploration. Data collection can be done in various ways, such as obtaining data directly from an organization (often as open data), using advanced

search techniques, web scraping, document conversion for analysis, or collecting data through observation, surveys, electronic forms, or crowdsourcing (Veglis & Bratsas, 2017a).

Data Cleaning: This particular stage involves identifying and correcting incorrect or corrupted records in a dataset (Wu, 2013). It includes removing human errors and standardizing data formats to maintain consistency with other data used by the journalist (Veglis & Bratsas, 2017b).

Data Understanding: In this stage, journalists must decipher various labels in datasets representing categories, classifications, or locations, along with specialized terminology. Often, additional data are required to make sense of existing data. Journalists need to have experience in using data, capable of understanding, formulating, and critically analyzing the data to meet specific needs (Veglis & Bratsas, 2017a).

Data Evaluation: This stage involves cross-referencing initial data and acquiring additional information from sources to enrich the data (Silverman, 2014; Veglis, 2013). It is important to recognize that datasets, like any source, have inherent biases and objectives. Journalists need to investigate the origin, purpose, and methodology of dataset collection (Bradshaw, 2010). This can be done by researching the dataset's creation history, identifying references to the dataset, or using other sources of information related to the research topic (Silverman, 2014; Veglis & Bratsas, 2017a).

Data Visualization: Since statistical information is abstract, transforming it into a physical representation requires an understanding of visual perception and cognition (Cairo, 2013). Effective data visualization should follow design principles based on human perception (Card, Mackinlay, and Shneiderman, 1999; Few, 2013).

The concept of data journalism is closely related to big data (Veglis, et al., 2022). The term "big data" was proposed in the late 20th century. It refers to datasets so large in size that they cannot be collected, curated, managed, and processed by general-purpose software running on typical personal computers (Lewis and Westlund, 2015; Snijders et al., 2012). Kitchin (2014) proposed a more analytical definition: Big data are voluminous (terabytes or petabytes), high-velocity (generated in or near real-time), varied in structure (structured and unstructured in nature), exhaustive in scope (attempting to capture entire populations or systems), high-resolution, uniquely indexed for identification, relational in nature (containing common fields that allow the connection of different datasets), and flexible (expandable).

Big data has influenced the field of media (Veglis and Maniou, 2018), where their application has been facilitated by new technological advancements that have automated and to some extent simplified data analysis (Stone, 2014). Despite the fact that big data are used in many cases in data journalism, it is not always necessary.

The use of programming in data journalism

The use of programming in data journalism is a topic of discussion and debate in the field of media. On one hand, it is argued that programming skills have become essential for journalists, especially data journalists, in today's media landscape. Advocates emphasize that programming allows journalists to collect, analyze, and present data in innovative ways, enabling them to uncover hidden patterns and provide deeper insights into complex issues. They also claim the

utility of programming skills in enhancing storytelling and investigative journalism (Simon, 2021). Porlezza contends that programming skills are often necessary for practicing data journalism (2024). Data journalism involves the collection, analysis, and visualization of data for storytelling and knowledge revelation. To effectively work with data, journalists often need to use programming languages (such as Python or R) to clean and process data, conduct statistical analyses, and create visualizations. Programming skills enable journalists to work with large datasets, automate data processing tasks, and create interactive data visualizations (Porlezza, 2024).

However, there are also concerns regarding the integration of programming into journalistic practice. Specifically, critics argue that journalists should focus on their core skills of storytelling and information gathering, leaving the technical aspects to data specialists and programmers. The concern lies in the fact that specialization in programming may divert journalists' attention from their primary duties and lead to a weakening of journalistic integrity (Cruz, n.d.; Hannaford, 2015). Heravi and Lorenz conclude that the use of programming in data journalism is a matter of debate (2020). While programming skills can be valuable for producing data-driven stories, not all data journalists necessarily need to know how to code. Porlezza (2024) does not directly disagree with this conclusion but believes that possessing programming skills can significantly enhance a journalist's ability to work effectively with data. Programming should not be considered the essence of data journalism but rather a tool that can substantially support the process (Heravi & Lorenz, 2020).

It is important to strike a balance between the positives and negatives of using programming in journalism. While programming skills can empower journalists to uncover significant narratives and engage the audience in new ways, it is crucial to ensure that journalistic standards and ethics are not compromised. Colman et al. (2018) examined the ethical considerations arising when journalists use programming techniques in their work. In conclusion, the use of programming techniques in journalism presents opportunities and challenges. It is essential for journalists to critically evaluate the benefits and risks associated with the use of programming languages, considering the specific context and goals of their reporting. By responsibly leveraging programming skills and adhering to ethical rules, journalists can harness the power of technology to enhance their storytelling and deliver impactful journalistic content to the audience.

Available solutions for adopting programming

In the field of data journalism, programming languages like Python and R have emerged as pivotal tools, offering powerful capabilities for data processing and visualization. Especially Python, renowned for its simplicity and readability, has become a fundamental tool in data journalism for tasks such as web scraping, data analysis, and interactive visualization. Its extensive libraries, such as Pandas and Matplotlib, enable journalists to efficiently handle large datasets and present complex information in understandable formats. The ability to embed Python code into websites further aids journalists in creating interactive online presentations of their findings. On the other hand, R is highly regarded for its capabilities in statistical analysis and data visualization, making it an ideal choice for in-depth analysis of complex datasets. R allows journalists to uncover hidden patterns and trends, while also providing a platform for

integrating code, visualizations, and narratives, thereby enhancing storytelling in data journalism. The open-source nature of both Python and R fosters a collaborative environment, with a vast community of developers contributing to their development and providing support. This aspect is crucial for journalists who often rely on shared resources and community support to learn and address potential issues. Moreover, the abundance of educational resources and online seminars for Python and R facilitates journalists in seamlessly integrating programming into their skill set (Appelgren & Nygren, 2014; Bradshaw, 2018; Bounegru et al., 2018; Hamilton, 2016; Knight, 2015).

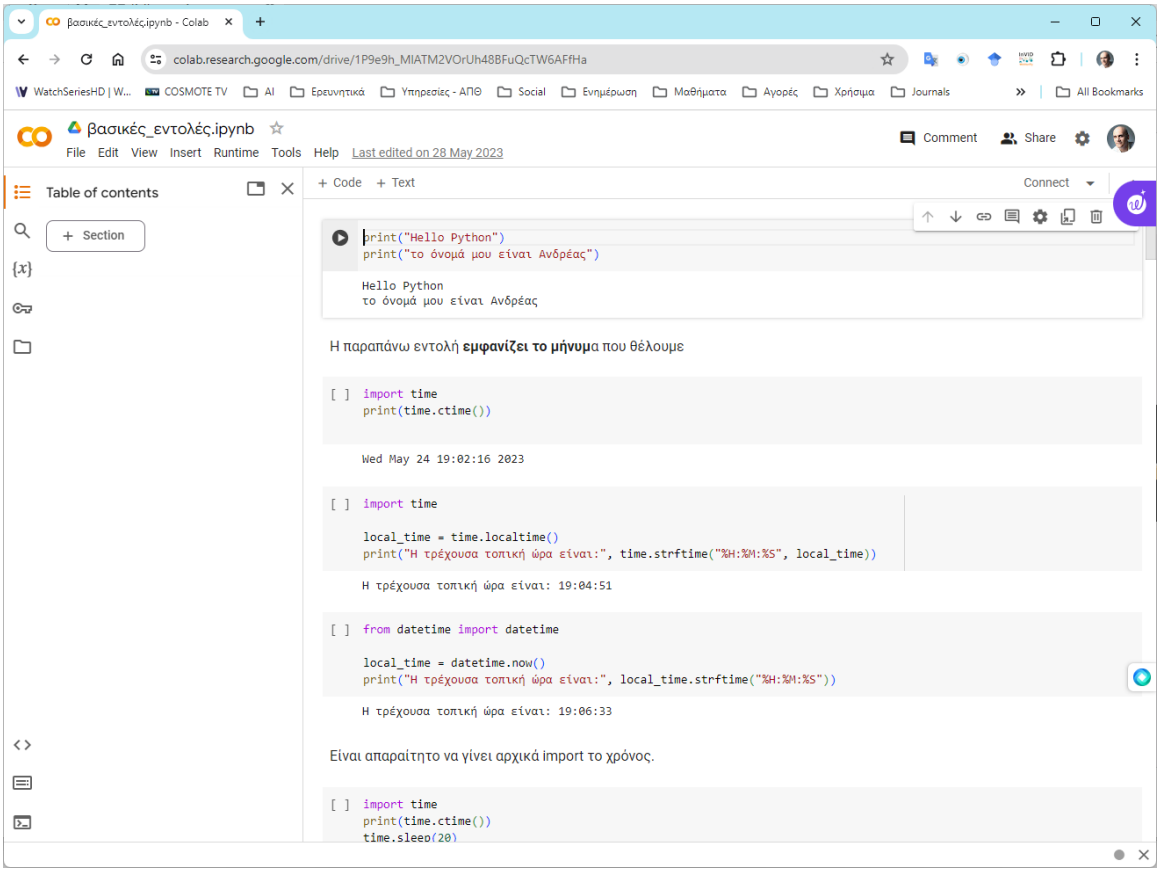

Figure 4: The interface of Google Colab.

It is worth noting that installing Python or R on a computer is not an easy process. However, there are alternative solutions, leveraging cloud computing, such as Google Colab, which supports writing and executing Python code through a web browser. Specifically, Google Colab is a free, cloud-based Jupyter Notebook environment (a specialized version of Python) that allows writing and executing Python code in a web browser, with free access to computing power, making it suitable for machine learning and data science (Burke, 2023). Thus, journalists do not need to install Python on a personal computer to use it (see Figure 4).

Another recent development is the integration of Python into the Microsoft Excel application. In this way, Microsoft combines the advanced computational capabilities of Python with the spreadsheet capabilities of Excel (Kinnestand, 2023). Users can use Python directly in Excel cells, leveraging Python libraries for data manipulation and visualization. This integration into

Excel allows journalists to gradually transition to using Python code while still using the familiar environment of Excel (see Figure 5).

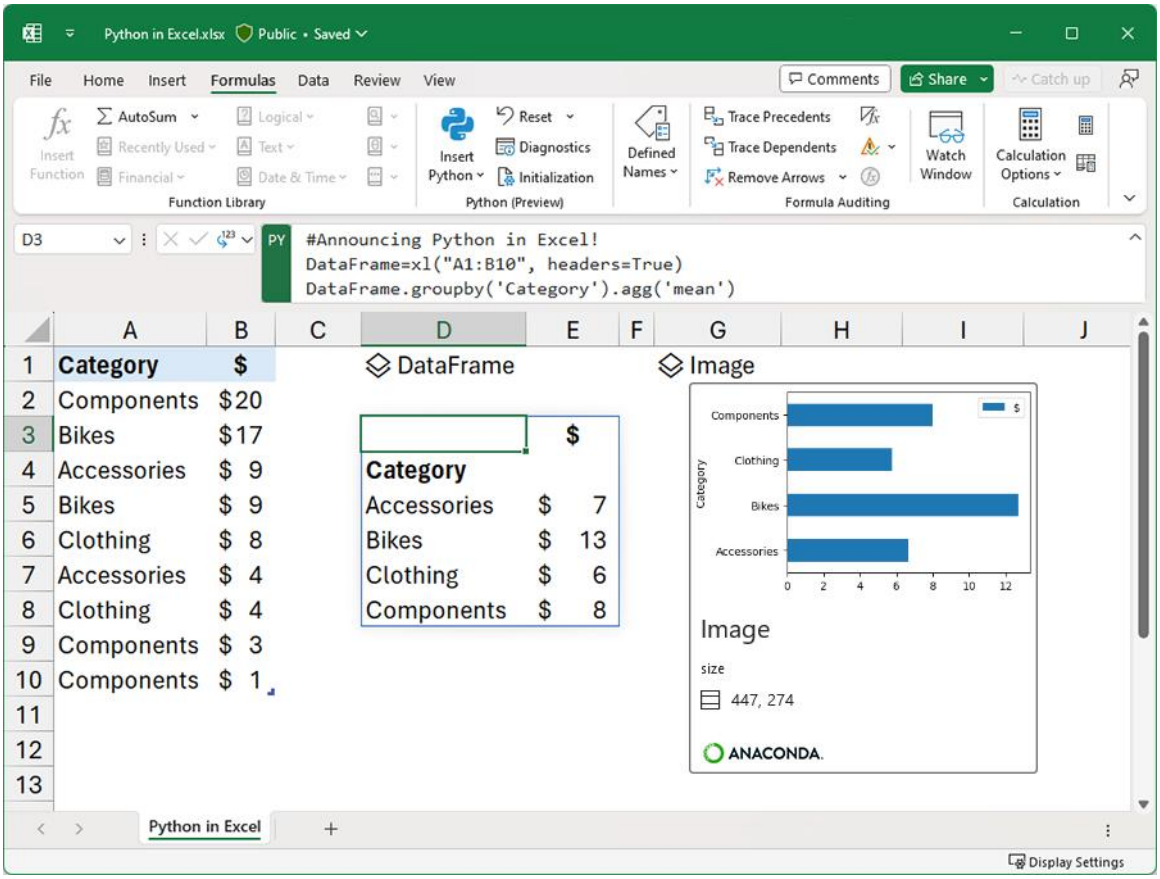

Figure 5: Integrating Python code in Microsoft Excel (https://insider.microsoft365.com/cs-cz/blog/python-in-excel-elevating-your-data-analysis-capabilities)

Another approach available for working with Python involves using artificial intelligence (AI) interfaces like ChatGPT, with the help of specialized add-on programs. ChatGPT can serve as an intuitive interface with natural language processing capabilities for journalists who may not have much programming knowledge. This could include querying datasets, creating statistical summaries, or even generating data visualizations using text commands (Inexturesolutions, 2023). In this way, the utilization of Python for data journalism becomes more accessible, allowing journalists to focus more on storytelling and less on the technical aspects of data handling (Knight, 2015). Using ChatGPT as an interface to Python can also be seen as democratizing data journalism, enabling a wider range of journalists to engage in data-driven storytelling. This approach aligns with the trend towards more accessible and collaborative forms of journalism (Hamilton, 2016). By reducing the barrier to entry into data journalism, AI plays a significant role in the evolution of the field, making it more inclusive and diverse.

Future developments

Tim Berners-Lee asserts that "data-driven journalism is the future" and encourages journalists to seek stories within datasets (Arthur, 2010). The development of the Semantic Web (Web 3.0) and Linked Open Data will play a significant role in the evolution of data journalism (Panagiotidis & Veglis, 2020). In the future, journalists will need to understand and utilize advanced technologies that will involve a more "smart web" (Bradshaw & Rohumaa, 2011). Semantic web technologies, smart devices (such as smartphones, etc.), and AI tools are constantly evolving and upgrading. Data journalism must embrace these technologies as they support the acquisition and verification of data, which is the primary source of data journalism. Significant advancement in the development of the Semantic Web is also expected due to recent developments in the field of Artificial Intelligence, particularly Generative Large Language Models. Understanding the content of the web (which is the goal of the Semantic Web) may become easier with the use of AI tools that can analyze large unstructured datasets (Breit, et al., 2023).

An important issue that needs to be addressed is the lack of online platforms that can support all stages of data journalism. Today, many online tools can be used at different stages of a data journalism project. Each tool has its unique interface, and journalists must spend time learning how to effectively use them. Most of these tools do not offer any integration with other tools. Thus, journalists spend a lot of time transferring data from one platform to another to leverage the strengths of each tool. In other words, the landscape is fragmented, and there is a lack of an integrated platform where journalists can work on all stages of data journalism projects. Data journalists can greatly benefit from the introduction of online platforms that connect various online tools that can be used in data journalism. This is a problem that the media industry must address in collaboration with the field of computer science (Veglis & Bratsas, 2021).

There is no doubt that data journalism today remains more of an exception than the rule, despite the fact that the majority of journalists express interest in it (EJC, 2024). Online journalists seem to be more engaged and have more incentives for data journalism. Given the ongoing shift towards online journalism (Cokley, Edstrom, McBride, & Ranke, 2015), many new developments are expected in the field of data journalism. In the ever-changing media landscape, where traditional media organizations face tough competition from social media services as well as citizen journalists, the level of complexity involved in practicing data journalism may present an opportunity for professional journalists to regain popularity and credibility.

References

Appelgren, E., & Nygren, G. (2014). Data journalism in Sweden: Introducing new methods and genres of journalism into ‘old’ organizations. *Digital Journalism, 2*(3), 394–405.

Arthur, C. (2010, November 22). Analysing data is the future for journalists, says Tim Berners-Lee. The guardian. http://www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee [accessed 29 April 2024].

Bradshaw, P. (2018). *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam University Press.

Bradshaw, P., (2010). How to be a data journalist. The Guardian Data Journalism http://www.Guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide [accessed 29 April 2024].

Breit, A., Waltersdorfer, L., Ekaputra, F. J., Sabou, M., Ekelhart, A., Iana, A., ... & Van Harmelen, F. (2023). Combining machine learning and semantic web: A systematic mapping study. *ACM Computing Surveys, 55*(14s), 1-41.

Bounegru, L., Chambers, L., & Gray, J. (Eds.). (2012). *The data journalism handbook*. European Journalism Centre.

Burke, J. (2023). *Why and how to use Google Colab*. TechTarget Available at https://www.techtarget.com/searchenterpriseai/tutorial/Why-and-how-to-use-Google-Colab [accessed 29 April 2024].

Cairo, A. (2013). *The functional art: an introduction to information graphics and visualization*. New Riders.

Card, S., Mackinlay, J., Shneiderman, B. (1999*). Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

Chen, C. H., Härdle, W., Unwin, A., & Friendly, M. (2008). *A brief history of data visualization*. Handbook of data visualization, 15-56, Springer.

Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital journalism, 3*(3), 331-348.

Cokley, J., Edstrom, M., McBride, J., & Ranke, A. (2011). Moving Away from 'Big Media': Students, Jobs and Long-tail Theory. *Australian Journalism Review, 33*(1), 107-118.

Colman, F., Bühlmann, V., O'Donnell, A. and van der Tuin, I. (2018). Ethics of Coding: A Report on the Algorithmic Condition [EoC]. H2020-EU.2.1.1. – INDUSTRIAL LEADERSHIP – Leadership in enabling and industrial technologies – Information and Communication Technologies. Brussels: European Commission. 732407, 1–54. https://cordis.europa.eu/project/rcn/207025_en.html [accessed 29 April 2024].

Cruz, J. (n.d.). Why journalists need data and coding skills more than ever | International Journalists' Network. International Journalists' Network. https://ijnet.org/en/story/why-journalists-need-data-and-coding-skills-more-ever [accessed 29 April 2024].

EJC, (2024) State of data journalism 2023. EJC.net. https://ejc.net/state-of-data-journalism-survey-2023 [accessed 29 April 2024].

Few, S. (2013). *Data Visualization for Human Perception*. In: Soegaard, Mads and Dam, Rikke Friis (eds.). The Encyclopedia of Human-Computer Interaction, 2nd Ed. Aarhus, Denmark: The Interaction Design Foundation.

Hamilton, J. T. (2016). *Democracy's Detectives: The Economics of Investigative Journalism*. Harvard University Press.

Hannaford, L. (2015). Computational journalism in the UK newsroom. Journalism education, 4, 6-21. https://e-space.mmu.ac.uk/575953/1/Article%20Computational%20Journalism.pdf [accessed 29 April 2024].

Heravi, B. R., & Lorenz, M. (2020). Data journalism practices globally: Skills, education, opportunities, and values. *Journalism and media, 1*(1), 26-40.

Hermida, A., & Young, M. L. (2017). Finding the data unicorn: A hierarchy of hybridity in data and computational journalism. *Digital Journalism, 5*(2), 159-176.

Inexturesolutions, (2023). How To Use ChatGPT API In Python For Your Real-Time Data. Medium.com, https://medium.com/@inexturesolutions/how-to-use-chatgpt-api-in-python-for-your-real-time-data-7838fed712b1 [accessed 29 April 2024].

Kinnestand, S. (2023). Introducing Python in Excel: The Best of Both Worlds for Data Analysis and Visualization, Microsoft 365 Blog. https://techcommunity.microsoft.com/t5/microsoft-365-blog/introducing-python-in-excel-the-best-of-both-worlds-for-data/ba-p/3905482#:~:text=Python%20in%20Excel%20combines%20Python's,to%20further%20refine%20your%20insights [accessed 29 April 2024].

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1-12. DOI: 10.1177/2053951714528481.

Knight, M. (2015). Data Journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice, 16*(1), 55-72.

Lewis, S.C., and Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism, 3*(3), 447-466.

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r) evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism, 21*(9), 1246-1263.

Lorenz, M. (2010). Data driven journalism: What is there to learn? Presented at IJ-7 Innovation Journalism Conference, 7–9 June, Stanford, CA.

Meyer, P. (2002). *Precision journalism: A reporter's introduction to social science methods*. Rowman & Littlefield Publishers.

Panagiotidis, K., & Veglis, A. (2020). Transitions in journalism—Toward a semantic-oriented technological framework. Journalism and Media, 1(1), 1-17. https://www.mdpi.com/2673-5172/1/1/1 [accessed 30 April 2024].

Porlezza, C. (2024). The datafication of digital journalism: A history of everlasting challenges between ethical issues and regulation. *Journalism, 25*(5), 1167-1185.

Rogers, S. (2008). Turning official figures into understandable graphics, at the press of a button. Inside the Guardian Blog, https://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata [accessed 29 April 2024].

Siapera, E., & Veglis, A. (Eds.). (2012). *The handbook of global online journalism*. John Wiley & Sons.

Silverman, C. (2014). Verification Handbook. European Journalism Centre http://verificationhandbook.com/downloads/verification.handbook.pdf [accessed 29 April 2024].

Simon, B. (2021). *Coding With Data in the Newsroom*. In Bounegru, L., & Gray, J. (eds). The Data Journalism Handbook: Towards a Critical Data Practice (p. 124-127). Amsterdam University Press.

Snijders, C. C. P., Matzat, U., & Reips, U. D. (2012). " Big Data": big gaps of knowledge in the field of internet science. *International journal of internet science, 7*(1), 1-5.

Spence, I. (2005). No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics, 30*(4), 353-368.

Stone, M. (2014). *Big Data for Media*. Report. Reuters Institute for the Study of Journalism.

Thienthaworn, E. (2021). What is Data Journalism? Medium. https://medium.com/@ekapon.th/what-is-data-journalism-17afbec6e30a [accessed 29 April 2024].

Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2, p. 9). Cheshire, CT: Graphics press.

Veglis, A. (2013). Education of Journalists on ICTs: Issues and Opportunities. Journal of Applied *Journalism & Media Studies, 2*(2), 265-279.

Veglis, A. (2009). *Cross Media Communication in newspaper organizations*. In A. Poulymerakou, N. Pouloudi, & K. Pramatari (eds), 4th Mediterranean Conference on Information Systems, Athens, Greece.

Veglis, A., & Bratsas, C. P. (2021). *Data Journalism: Definition, Skills, Difficulties, and Perspectives*. In Encyclopedia of Information Science and Technology, Fifth Edition (pp. 1140-1151). IGI Global.

Veglis, A., Bratsas, C. (2017a). Journalists in the age of Data Journalism: The case of Greece. *Journal of Applied Journalism & Media Studies, 6*(2), 225-244.

Veglis, A., & Bratsas, C. (2017b). Towards a taxonomy of data journalism. *Journal of Media Critiques, 3*(11), 109-121.

Veglis, A., and Maniou, T. (2019). Chatbots on the Rise: A new Narrative in Journalism. *Studies in Media & Communication, 7*(1), pp. 1-6. DOI:10.11114/smc.v7i1.3986.

Veglis, A., Saridou, T., Panagiotidis, K., Karypidou, C., & Kotenidis, E. (2022). Applications of Big Data in Media Organizations. *Social Sciences, 11*(9), 414.

Weber, W., Engebretsen, M., & Kennedy, H. (2018). Data stories: Rethinking journalistic storytelling in the context of data journalism. *Studies in communication sciences*, (1), 191-206.

Wu, S. (2013). A review on coarse warranty data and analysis. *Reliability Engineering and System*, 114, 1-11.