Enhancing Quality in Official Statistics: A European Perspective of Open Source Technologies

Ciprian Alexandru[1], Ana-Maria Ciuhu[2]

[1]*Ecological University of Bucharest, Romania*

[2]*Romanian National Institute of Statistics; Institute of National Economy, Romanian Academy, Bucharest, Romania*

Abstract

This presentation explores the pivotal role of open-source technologies, notably R and Python, in enhancing the quality of official statistics in Europe. We emphasize how these technologies are revolutionizing data collection, analysis, and dissemination, crucial for quality improvement in statistical processes.

Central to our discussion is integrating open-source tools within the European statistical system. We illustrate how leveraging these technologies can lead to more efficient, transparent, and robust methodologies in statistical data handling. The adaptability and versatility of open-source tools are key in meeting the complex requirements of contemporary data science in official statistics, thereby directly contributing to the improvement of statistical quality.

Additionally, we explore into the broader implications of open-source technology in the European context. This includes its role in fostering innovation, educational development, and building a skilled workforce. We argue that open-source tools are instrumental in enhancing the capabilities of statisticians and data scientists, a vital aspect for maintaining and improving the quality of statistics in Europe.

Furthermore, we examine the impact of open-source technologies on the operational strategies of statistical agencies. This encompasses a detailed look at how these technologies streamline operations, enhance statistical methodologies, and lead to more effective practices. In conclusion, our submission will not only highlight the current applications of open source in European official statistics but also propose strategies for its broader integration in the future. We aim to provide insights into the effective utilization of open-source technologies for advancing the quality, reliability, and accessibility of statistical data in Europe, thereby contributing to the ongoing discourse on the strategic importance of innovative methodologies in the field of official statistics.

Keywords: Official Statistics, Open-source, Quality, Data science, R and Python

1. Introduction

1.1 Objective

The primary objective of this paper is to explore and articulate the transformative impact of open-source technologies, specifically R and Python, on the enhancement of quality in official statistics within Europe. This discussion is timely and significant as the European statistical community continues to seek more efficient, transparent, and robust methodologies to improve data collection, analysis, and dissemination processes.

The motivation behind focusing on open-source technologies arises from their proven capability to foster innovation and efficiency in statistical practices globally. These technologies are not only important for handling complex datasets but also enhance the accessibility and

reproducibility of statistical analysis, which are critical components in the context of official statistics.

The paper will explore the integration of these technologies within the European Statistical System (ESS), examining how they contribute to achieving the high standards of quality required for official statistics. It will highlight specific instances where R and Python have been successfully implemented to streamline operations and improve outcomes in data-related tasks across some national statistical offices within Europe.

The use of open source software in official statistics is aligned with the broader strategic goals of digital transformation in the public sector, as endorsed by several European directives and initiatives. This includes the European Commission's Digital Single Market strategy, which emphasizes the importance of open data and open-source software as tools for innovation and economic growth (European Commission, 2015).

Through this exploration, the paper aims to contribute to the ongoing discourse on statistical methodology and practice by showcasing how open source technologies can significantly uplift the quality of statistical outputs. It seeks to provide actionable insights and recommendations for statistical agencies contemplating the integration of these tools into their operational frameworks.

In summary, this paper seeks to bridge the gap between technological capability and statistical needs, providing a comprehensive overview of how open source tools like R and Python can be implemented to enhance the quality, reliability, and accessibility of statistical data in Europe. The ultimate goal is to inform and inspire further adoption and innovation in the realm of official statistics.

1.2 Importance of Quality in Statistics

Quality in statistics is multifaceted, generally encompassing aspects such as accuracy, timeliness, comparability, coherence, and accessibility. The ESS provides a comprehensive quality framework that serves as a guideline for ensuring that these quality dimensions are met in all statistical outputs (Eurostat, 2017).

In Europe, the quality of official statistics is safeguarded through rigorous methodologies and standards set by Eurostat and individual national statistical institutes. These standards are influencing the maintaining of data integrity which, in turn, supports economic and social policies across the continent.

Accurate and reliable statistics are the backbone of effective policy making. For instance, high-quality labor market statistics inform employment policies, while accurate economic data guide fiscal and monetary decisions. The ESS report on the quality of statistical data discusses how data accuracy impacts policy decisions across the EU (Eurostat, 2020).

Despite the high standards, statistical systems face challenges such as resource limitations, rapidly changing information needs, and the increasing complexity of data sources. Addressing these challenges is essential for maintaining the quality of statistical outputs. The European Commission's communication on the data economy underlines the need for robust statistical systems to adapt to digital transformation, ensuring that data quality is not compromised as data volumes and sources grow (European Commission, 2018).

The Role of Quality Assurance: Quality assurance mechanisms play a critical role in this context. They involve regular reviews and audits of statistical processes and outputs, as outlined in the Code of Practice for European Statistics (Eurostat, 2017). These practices ensure that the statistical products meet the required quality standards and remain relevant and trustworthy.

1.3 Overview of Technologies

In the landscape of official statistics, the deployment of advanced software technologies, particularly open source tools such as R (R Core Team, 2023) and Python (Van Rossum, G.,

& Drake Jr, F. L., 1995) has been transformative. These technologies are increasingly recognized for their powerful capabilities in statistical analysis, data manipulation, and visualization.

Originally developed for statistical computing, R has a rich ecosystem of packages (Templ & Todorov, 2015), that make it an invaluable resource for statisticians looking to implement sophisticated and tailor-made data analysis and graphical techniques with ease and efficiency.

Known for its versatility, Python excels in data manipulation, machine learning, and large-scale data processing. Its ability to integrate with other technologies and ease of learning makes it a popular choice not only in academia but also in industry.

Both R and Python support a wide range of statistical techniques important for quality assurance in official statistics— from simple descriptive statistics to complex predictive models. They enable statisticians to handle large datasets more efficiently, conduct real-time data analysis, and develop interactive visualizations that make data more accessible.

The integration of these technologies into the European Statistical System has facilitated improvements in various statistical processes. For instance, automated data collection and processing workflows have been streamlined, which enhances the timeliness and accuracy of statistical outputs.

Supporting policies and initiatives include European measures like the Public Sector Information Directive (European Parliament, 2019). This directive advocates for the adoption of open technologies to enhance transparency and foster innovation in the management of public data.

R and Python are at the forefront of a technological revolution in the field of official statistics. Their ability to process large volumes of data and perform complex analyses with high levels of transparency and reproducibility makes them indispensable tools for modern statisticians. This section sets the stage for further discussions on how these technologies are specifically applied to improve the quality of statistical data within the European context.

2.   Integration of Open Source Tools in the European Statistical System

2.1 Case Studies

The integration of open source technologies like R and Python into ESS has faced varying degrees of implementation across different countries. These technologies have empowered statistical agencies to enhance their data processing capabilities, leading to more efficient and transparent statistical practices. Below we present some case studies from Romania (INS), Italy (ISTAT) and Austria, which illustrate the effective use and implementation of R and Python in the official statistics.

Romanian INS has adopted R since 2012 (Dobre & Adam, 2014) and introduced it for Small Area Estimations applied on international migration (Ciuhu & Caragea, 2015). Current uses of R include sampling, calibration, and calculation of quality indicators in household surveys, data validation, and generation of data tables for dissemination. Additionally, R is employed for analyzing paradata, managing databases from statistical and administrative sources, and conducting Small Area Estimations. It is also used for web scrapping (Oancea & Necula, 2019), development of the Tempo package for extracting data from the Romanian INS Tempo database (Necula et al, 2019), use of remote sensing data (Necula et al, 2023) and text classification (Oancea, B., 2023). As part of its R implementation process, Romanian INS is organizing yearly the international conference Use of R in Official Statistics (uRos). The uRos conferences (www.urosconf.org) illustrate a successful model for cooperation between official statistics and academia, as well as among different NSOs. The initiative began as an academic workshop in 2013, reflecting its strong roots in the educational sector. Over time, it has grown to become a prominent conference that not only bridges the gap between academia and official

statistics but also fosters collaboration between different NSOs. The core organizing committee of the uRos conferences includes representatives from both universities and NSOs (Romania, Austria and Netherlands), ensuring a balanced perspective and a comprehensive approach to the challenges and advancements in the field of statistics. By alternating the conference location annually between various NSOs, since 2018, the uRos conferences facilitate an international exchange of knowledge and best practices.

At Romanian INS, Python was introduced in 2020 for European Health Interview Survey in households in data validation and generation of data tables for dissemination. Furthermore it was used during the 2021 round PHC to advance its capabilities in developing applications for external users and for validation processes.

At Istat (Italy), the transition from SAS to R and Python began in response to high costs and a government directive favoring open-source software (Barcaroli, 2019). Initially, Istat relied heavily on proprietary software, with SAS being a significant component. Due to high costs and government directives favoring open-source solutions, Istat began integrating R, developing R packages for various statistical tasks such as sampling, data integration, and selective editing. Over time, a growing community within Istat also began using Python, particularly for text processing and machine learning. Both R and Python now coexist at Istat, each serving different needs and complementing each other in statistical computing and data science tasks.

Statistics Austria migrated from RStudio Desktop to RStudio Server to overcome the difficulties in maintaining and updating individual R installations by centralizing the environment (Meindl, Kowarik, 2018). The preparation phase in 2017 involved collaborating with the IT department to set up Linux servers for testing and production and evaluating RStudio Server Pro. In 2018, they developed training courses, transitioned users from desktop to server, and installed RStudio Connect for deploying Shiny apps and reports. By 2019, they aimed to potentially remove the last desktop installations. The hardware setup included virtualized servers with Ubuntu 16.04, Xeon 8 Cores, and 128GB RAM, while the software setup included RStudio Server, Shiny Server, and RStudio Connect, utilizing internal Bitbucket for version control. Users were added to a Jira project for support, and a test environment ensured stability before full deployment. Achievements included the successful migration of approximately 140 R users and the conduction of several training sessions, with ongoing fine-tuning to aid user adaptation. Lessons learned emphasized the importance of realistic planning, gradual implementation, IT cooperation, user involvement, anticipating roadblocks, and dedicating time for training and feedback. The phased approach and careful planning ensured a smooth transition to RStudio Server.

2.2 Benefits

The integration of open source tools such as R and Python into the ESS offers significant benefits that enhance the operational capabilities of statistical agencies.

Enhanced efficiency and productivity are achieved through open-source tools, which automate routine data processing tasks, significantly speeding up workflows and allowing statisticians to focus on analysis. These tools – R and Python - perform in handling large datasets efficiently, which is important for managing increasing data volumes.

Increased transparency and reproducibility are facilitated by the inherent design of open-source software for sharing and collaboration. Statistical methodologies encoded in R and Python can be easily shared and reviewed, enhancing transparency and facilitating reproducibility. The communities around these tools continuously contribute to their development, providing a rich resource of plugins, libraries, and modules that keep the tools at the cutting edge of statistical analysis techniques.

Cost-effectiveness is realized through reduced licensing costs, as open-source tools are generally free to use, significantly reducing overhead costs. Savings on software licenses allow

agencies to reallocate resources to other critical areas such as data collection or human capital development.

Improved data security and privacy are achieved by giving agencies control over the software, allowing them to implement robust security measures and comply with data privacy regulations. The open-source community rapidly addresses security vulnerabilities, providing updates and patches that help secure the software.

Adaptability and customization are facilitated by the ability to tailor R and Python to meet specific needs. Custom scripts and modules can be developed for unique statistical methods or data types. These tools can be seamlessly integrated with other data management systems and platforms, enhancing data interoperability across different agencies.

2.3 Challenges

While the adoption of open source technologies such as R and Python offers numerous benefits to statistical agencies, the integration process also presents several challenges. These challenges must be effectively managed to ensure that the potential of these tools is fully realized. Key issues include training needs, data security concerns, and other operational hurdles.

Training and skill development for R and Python involve several challenges. Both languages have steep learning curves, requiring substantial training, which can hinder quick integration. Continuous education is necessary due to the fast-evolving nature of these technologies, and allocating resources for comprehensive training programs can be difficult.

Data security and privacy concerns include vulnerabilities to threats and the complexity of ensuring compliance with regulations like GDPR. Integrating open source tools with existing proprietary systems can cause compatibility issues and complicate dependency management.

Community dependency presents risks, such as reliance on community support and varying quality control. Mitigation strategies include developing tailored training programs, implementing robust security protocols, and encouraging active community engagement.

These challenges, while significant, can be managed through careful planning, strategic resource allocation, and a proactive approach to staff training and community engagement. Addressing these issues is important for statistical agencies to fully imply open source technologies and enhance their statistical capabilities.

2.4 Adaptability and Versatility of Open Source Solutions

The adaptability and versatility of open source solutions are among their most valued attributes, especially when dealing with the complex and evolving data needs of modern statistical agencies. These characteristics not only allow for tailored analytical approaches but also enable statistical organizations to keep pace with technological advancements and changing data landscapes.

Adaptability to diverse data requirements is a significant advantage of open source tools like R and Python. Their customizability allows statistical agencies to modify and adapt the tools to specific needs or unique datasets, offering flexibility beyond the rigid structures of proprietary software. Additionally, these tools can scale to handle growing data volumes and computational demands without requiring significant additional investment, essential for large-scale national data collections and complex econometric models.

Versatility across different data types is another benefit. Open source languages can process both structured data, such as numerical and categorical data in databases, and unstructured data, like text, images, and social media content. Python's libraries, like numpy for numerical data and nltk or BeautifulSoup for text processing, are invaluable for integrating and analyzing diverse data sources. These tools are widely applicable across various statistical domains, including economic statistics, health analytics, environmental studies, and demographic

research, enhancing the capacity to address multifaceted and interdisciplinary data challenges.

Innovative approaches to statistical analysis are supported by open source tools through advanced analytical techniques and algorithms, including machine learning, artificial intelligence, and deep learning. These capabilities enable statisticians to explore new methodologies and improve the precision of their analyses. Furthermore, the dynamic visualization capabilities of tools like ggplot2 in R and matplotlib and plotly in Python allow for the creation of complex visual representations, making statistical data more accessible and understandable to a broader audience.

3.   Advantages of Open Source Technologies on Quality

In official statistics, the quality of data and analyses is critical. This section highlights the specific advantages that open-source technologies offer in enhancing the quality of official statistics, emphasizing the flexibility, transparency, and innovative capabilities of tools like R and Python. These tools allow for customizable and scalable solutions, support advanced analytical techniques, and facilitate dynamic visualizations, contributing to more accurate and accessible statistical data.

Improved accuracy and reliability are achieved through error reduction and reproducibility. Open-source software allows for peer review, significantly reducing bugs in statistical software. The transparency of open-source code facilitates the reproducibility of research findings, for example on platforms such as GitHub.

Enhanced flexibility and responsiveness are clear in the customization and agility of open-source tools. These tools can be tailored to meet specific data analysis needs, essential in official statistics where data types and analytical requirements vary widely. The ability to modify and extend open-source software allows statistical agencies to quickly adapt to new data trends or policy requirements.

Broader educational impact is achieved through skill development and collaboration with academia. The widespread use of open-source statistical tools has influenced the educational sector, leading to a broader base of future statisticians and data scientists proficient in these technologies. Many statistical agencies have partnered with universities to provide training in open-source technologies, enhancing the quality of education and research in statistics.

4.   Broader Implications for Innovation and Education

4.1 Innovation

Open-source technologies are important in fostering innovation within statistical methodologies, significantly impacting how data is analyzed, interpreted, and used in official statistics. The contribution of open source to statistical and technological innovation can be traced through several key developments. Open source software like R and Python allows statisticians to implement complex statistical models, giving free access to advanced methodologies such as machine learning and predictive analytics. The flexibility of these tools accelerates research by allowing modifications to suit specific needs, fostering rapid advancements in statistical methods. Open source projects thrive on global collaboration, hosted on platforms like GitHub, where statisticians and data scientists contribute and develop software in real time. This collaborative environment speeds up innovation and enhances research robustness through peer reviews and diverse input.

Open source software reduces barriers to entry for statistical analysis by being freely available, allowing institutions with limited budgets to engage in cutting-edge research. The community offers extensive resources, facilitating self-learning and capacity building. The versatility of open source tools encourages interdisciplinary research, with statisticians collaborating with

experts in fields like biology, public health, and environmental science, integrating various data types and methodologies.

Open source software inherently supports reproducible research, ensuring findings can be verified and replicated, which is essential for scientific progression.

4.2 Educational Development

The integration of open source technologies in educational programs, particularly those related to data science and statistics, has an important impact on the development of curricula and the skill set of the workforce. The accessibility and versatility of tools like R and Python are transforming statistical education and research, broadening educational and research capabilities. Open source software is now fundamental in statistics and data science curricula, providing hands-on experience with data manipulation and analysis, and preparing students for modern statistical jobs. This training builds a skilled workforce ready for data-driven roles across various sectors. Educational programs, including partnerships with industry and MOOCs, enhance learning through practical experience and accessible, high-quality courses. Open source tools also support academic research by providing cost-free, sophisticated statistical methods, fostering innovation and collaboration. However, maintaining updated educational resources and adapting to industry changes remain challenges.

4.3 Workforce Development

Open source tools such as R and Python play a critical role in workforce development within the European context, addressing the growing demand for data literacy and advanced analytical skills across various sectors. The integration of these tools into educational and professional development frameworks is equipping the European workforce with the necessary skills to excel in a data-driven economy.

Enhancing data literacy through open source software provides universal access to advanced analytical tools, reducing barriers related to cost and accessibility. This democratization ensures that individuals and organizations across Europe, regardless of size or budget, can develop and enhance their data analysis capabilities. The widespread adoption of tools like R and Python standardizes data science skills across the workforce, simplifying the assessment of competencies and facilitating the transition of employees between roles and industries.

Professional training and continuous learning are supported by numerous European initiatives, such as those backed by the EU's Digital Europe Programme, which emphasize digital skills training, including the use of open source software for data analysis. These programs aim to upskill workers, ensuring they remain competitive in the labor market. Open source communities foster an environment of continuous learning and improvement, essential in the dynamic field of data science. Community forums, webinars, and workshops offer ongoing opportunities for professionals to refine their skills and stay updated with the latest advancements.

Bridging the skill gap involves collaboration with industry, where many European educational institutions align their curricula with real-world needs to ensure that taught skills are relevant and applicable. Internships and applied projects provide students with practical experience using open source tools, better preparing them for real-world data challenges they will face in their careers.

5. Operational Strategies and Impacts

5.1 Streamlining Operations

Open source technologies significantly streamline operations in statistical agencies by enhancing data handling efficiencies, reducing operational overheads, and facilitating better management of statistical processes. These technologies, particularly tools like R and Python,

have transformed the way data is processed, analyzed, and shared, leading to more streamlined operations across various levels of statistical organizations.

Open source tools like R and Python automate routine data processing tasks, such as data cleaning, transformation, and summarization, reducing the need for manual intervention and minimizing human error. For instance, Python scripts can handle data extraction and cleaning from various sources. Automated reporting is facilitated by tools like R Markdown and Jupyter Notebooks, which create dynamic reports that update with new data, saving time and ensuring accuracy.

Efficiency in data analysis is enhanced by advanced analytical tools and real-time data analysis capabilities, allowing complex analyses to be performed quickly and providing timely insights important for decision-making. Open source software also reduces costs related to software licensing fees, enabling budget savings to be redirected to other needs, such as data collection and employee training. These solutions are highly scalable, handling increasing data volumes without expensive upgrades.

Collaboration and flexibility are improved through the collaborative development of open source projects, enhancing software quality and sharing best practices. Open source tools can be customized to meet specific agency needs, allowing adaptation as operational requirements evolve.

5.2 Enhancing Methodologies

The adoption of open source technologies has had a transformative impact on the methodologies used within statistical agencies. Tools such as R and Python have introduced advanced capabilities for data analysis, leading to significant enhancements in the methodologies for collecting, processing, analyzing, and disseminating statistical data.

Open source technologies, such as R and Python, enhance advanced data analysis techniques by facilitating machine learning and AI through libraries like TensorFlow and scikit-learn. These tools enable complex statistical modeling with packages like lme4 and survival, improving data quality through automated validation and consistency checks. They offer customizable workflows and rapid prototyping, allowing researchers to adapt to new data and methods. Open source tools handle large datasets efficiently and support parallel computing, speeding up processing times. However, continuous learning and significant investment in training are essential to fully improve these advanced methodologies.

6.   Future Perspectives and Strategies

6.1 Future Integration

As statistical agencies continue to navigate the evolving landscape of data-driven decision-making, the broader integration of open source technologies into statistical processes presents a strategic opportunity to enhance efficiency, transparency, and innovation. To expand the use of open source tools in statistical methodologies and operations, it is essential to implement a multifaceted approach. Providing comprehensive training programs that span from foundational to advanced levels and offering continuous learning and certification opportunities will ensure proficiency in these tools. Cultivating a culture of innovation and collaboration through the establishment of internal innovation labs and promoting joint projects with external organizations can drive progress. Strengthening infrastructure and support by investing in robust IT systems and dedicated support teams is also important. Adopting 'open source first' policies, with clear guidelines and documentation standards, will facilitate a smooth transition. Additionally, encouraging active participation in open source communities and developing platforms for resource sharing can improve community expertise and resources effectively.

6.2. Integrating AI and Open Source in Statistics: Future Outlook

The integration of open source technologies with artificial intelligence (AI) holds transformative potential for the field of official statistics. This section explores the future implications of the synergy between AI and open source technologies, focusing on how it can drive significant advancements in data analysis, decision-making processes, and overall statistical methodology.

AI algorithms can process vast amounts of data more efficiently and accurately than traditional methods, enhancing analytical capabilities through frameworks like TensorFlow and PyTorch. Predictive analytics integrated with open source tools can lead to more accurate forecasts, providing valuable insights for policy-making.

AI can also improve data quality and integrity by detecting and correcting errors automatically, and enabling real-time data validation. Additionally, AI can streamline operations by automating reporting and optimizing resource allocation. Enhanced decision-making processes can be achieved through AI's scenario analysis and simulation capabilities, supporting data-driven policymaking. However, ethical considerations such as bias, fairness, privacy, and security must be addressed to maintain public trust. Future research directions should include interdisciplinary studies and the establishment of innovation labs to explore the full potential and address the challenges of using AI in official statistics.

7.   Conclusions

This paper has explored the transformative potential of open source technologies within the field of official statistics. By integrating tools like R and Python, statistical agencies have experienced enhanced efficiencies in data collection, processing, and analysis. We discussed various ways these technologies improve operations—from automating routine tasks and enabling real-time analytics to facilitating advanced predictive modeling and machine learning applications. The discussion extended into educational impacts, demonstrating how open source technologies are important in developing a skilled workforce capable of handling complex and voluminous data sets.

Open source technologies have proven essential in advancing the quality, reliability, and accessibility of statistical data. By fostering an environment of transparency, these tools enhance the reproducibility of statistical analyses and ensure that data processes are open to scrutiny and validation. The ability to customize and adapt open source software to specific needs allows for continual improvement in methodologies, ensuring that statistical outputs are not only accurate but also highly relevant to current data challenges. The integration of these technologies supports a more robust statistical framework that can dynamically respond to the evolving landscape of data-driven decision-making.

As we stand on the brink of a new era in statistical analysis and data processing, it is imperative for stakeholders at all levels—government agencies, educational institutions, private sector entities, and international organizations—to invest in and adopt open source technologies. This call to action is not merely about adopting new software tools but embracing a culture of innovation, collaboration, and transparency that open source embodies. Stakeholders should actively participate in open source communities, contribute to the development of these technologies, and promote their use within their spheres of influence.

References

Barcaroli, G. (2019) R at the Italian National Institute of Statistics (Istat): a twelve-year story, presented at the *International Conference Use of R in Official Statistics uRos2019*, https://www.r-project.ro/conference2019/presentations/Keynote_Barcaroli.pdf

Ciuhu (Dobre), A.M., Caragea, N. (2015). Producing small area estimation using R in the Romanian official statistics. *Romanian Journal of Economics.* 40. 115-125, https://revecon.ro/articles/2015-1/2015-1-5.pdf

Dobre, A.M., & Adam, R. (2014) The Progress of R in Romanian Official Statistics, *Romanian Statistical Review*, Volume 2, 45-54. https://www.revistadestatistica.ro/wp-content/uploads/2014/07/RRS_2_2014_a05.pdf

European Commission (2015), Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Digital Single Market Strategy for Europe. https://ec.europa.eu/commission/presscorner/api/files/attachment/8210/DSM_communication.pdf

European Commission (2018), Communication on Building a European Data Economy. https://eur-lex.europa.eu/content/news/building_EU_data_economy.html

European Parliament (2019), Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), https://eur-lex.europa.eu/eli/dir/2019/1024/oj

Eurostat (2017), Quality Assurance Framework of the European Statistical System.

    https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V2.0-final.pdf

Eurostat (2017), European Statistics Code of Practice – 2017 revision edition. https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7?t=1528447068000

Eurostat (2020), Quality report on National and Regional Accounts 2019 Data - 2020 edition, https://ec.europa.eu/eurostat/documents/7870049/12108401/KS-FT-20-007-EN-N.pdf/a2229bc1-ff53-b246-18af-8791f124571d?t=1608301992302

Meindl, B., Kowarik, A. (2018) Migrating to Rstudio Server Experiences from Statistics Austria, presented at the *International Conference Use of R in Official Statistics uRos2018*, https://r-project.ro/conference2018/presentations/Bernhard_Meindel_Migrating.pdf

Necula, M., Țîru, A.M., Oancea, B. (2019) Tempo – an R package to access the TEMPO-Online database, *Romanian Statistical Review*, 3/2019, pp. 61-71, https://www.revistadestatistica.ro/wp-content/uploads/2019/09/A6_RRS-3_2019.pdf

Necula M., Andrei, T., Oancea, B., Păun, M. (2023) Urban Green Index estimation based on data collected by remote sensing for Romanian cities, *Romanian Statistical Review*, 4/2023, pp. 59-76, https://www.revistadestatistica.ro/wp-content/uploads/2023/12/4_RRS-4_2023.pdf

Oancea, B., Necula M., (2019) Web scraping techniques for price statistics – the Romanian experience, *Statistical Journal of the IAOS*, vol. 35, no. 4, pp. 657-667, https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji190529, DOI: 10.3233/SJI-190529

Oancea, B. (2023) Text classification using machine learning methods, presented at *International Conference KNOWCON 2023 Knowledge on Economics and Management*, https://kems.upol.cz/knowcon/programme/

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria., https://www.R-project.org

Templ, M., & Todorov, V. (2015) The Software Environment R for Official Statistics and Survey Methodology. *Austrian Journal of Statistics*, October 2015, Volume 44, 1–28. https://doi.org/10.17713/ajs.v45i1.100

Van Rossum, G., & Drake Jr, F. L. (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.