EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

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

Ciprian Alexandru - Ecological University of Bucharest, Romania

Ana-Maria Ciuhu - Romanian National Institute of Statistics; Institute of National Economy, Romanian Academy, Bucharest, Romania

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Presentation Objectives

- Explore the role of open-source technologies in European official statistics;
- Expose some case studies;
- Highlight the importance of these technologies in fostering innovation and educational development.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Importance of Quality in Statistics

- **Quality Dimensions**: Accuracy, timeliness, comparability, coherence, accessibility.
- **Impact on Policy Making**: Accurate statistics are essential for effective economic and social policies.

The conference is partly financed by the European Union

Overview of Open Source Technologies

- **R**: Developed for statistical computing, rich ecosystem of packages (e.g., ggplot2, dplyr).
- **Python**: Versatile, excels in data manipulation, machine learning, and large-scale data processing (e.g., pandas, scikit-learn).

Benefits: Efficiency, transparency, reproducibility, and cost-effectiveness.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Integration of Open Source Tools in the European Statistical System

Examples of implementation: Romania (INS), Italy (ISTAT), Austria (Statistics Austria).

Contribution: Improved efficiency, transparency, and reproducibility in statistical practices

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Case Study: Romania (INS)

Adoption of R: Since 2012, for Small Area Estimations on international migration.

Furthermore: sampling, calibration, quality indicators in household surveys, data validation, generating dissemination tables.

Other R Applications: Analyzing paradata and managing databases;

- Web scraping; Tempo package for extracting data from INS Tempo database;
- Use of remote sensing data; Text classification.

International Conference Use of R in Official Statistics (uRos): Annual event organized by Romanian INS, fostering collaboration between academia and NSOs.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Case Study: Romania (INS)

- **Adoption of Python**: Introduced in 2020 for the European Health Interview Survey in households, data validation, and generating dissemination tables.
- **2021 Population and Housing Census (PHC)**: Enhanced capabilities for developing applications for external users and validation processes.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Case Study: Italy (ISTAT)

- **Transition from SAS to R and Python**: Driven by high costs and government directives favoring open-source software.
- **R Integration**: Development of R packages for sampling, data integration, and selective editing.
- **Python Integration**: Used for text processing and machine learning tasks.
- **Coexistence of R and Python**: Complementary use in statistical computing and data science tasks.

Case Study: Austria (Statistics Austria)

- **Migration to RStudio Server**: Overcoming difficulties in maintaining and updating individual R installations.
- **Preparation Phase (2017)**:
  - Collaboration with IT department to set up Linux servers for testing and production.
  - Evaluation of RStudio Server Pro.
- **Training and Transition (2018)**:
  - Development of training courses.
  - Transition of users from desktop to server environment.
  - Installation of RStudio Connect for deploying Shiny apps and reports.
- **Full Deployment (2019)**: Migration of approximately 140 R users, ongoing fine-tuning for user adaptation.

INSTITUTO NACIONAL DE ESTATÍSTICA | STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Benefits of Open Source Tools

- **Efficiency and Productivity**: Automation of routine data processing tasks, handling large datasets efficiently.
- **Transparency and Reproducibility**: Open-source design for sharing and collaboration, enhancing transparency.
- **Cost-Effectiveness**: Reduced licensing costs, enabling resource reallocation.
- **Data Security and Privacy**: Control over software, implementation of robust security measures.
- **Adaptability and Customization**: Tailoring R and Python to specific needs, integration with other data management systems.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Challenges in Adopting Open Source Technologies

- **Training and Skill Development**: Steep learning curves, need for continuous education.
- **Data Security Concerns**: Ensuring compliance with regulations like GDPR.
- **Community Dependency**: Reliance on community support, varying quality control.
- **Mitigation Strategies**: Tailored training programs, robust security protocols, active community engagement.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Broader Implications for Innovation and Education

- **Innovation**: Implementation of complex statistical models, global collaboration, reduced barriers to entry.
- **Educational Development**: Integration in educational programs, hands-on experience with data manipulation and analysis.

The conference is partly financed by the European Union

Future Perspectives and Strategies

- **Future Integration**: Comprehensive training programs, innovation labs, robust IT systems.
- **AI and Open Source Synergy**: Advanced data analysis, predictive analytics, improved decision-making processes.
- **Ethical Considerations**: Addressing bias, fairness, privacy, and security.

Conclusions

- **Transformative Potential**: Enhancing the quality, reliability, and accessibility of statistical data.

- **Call to Action**: Embracing innovation, collaboration, and transparency in statistical methodologies.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Thank you for your attention!

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union