



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Using cloud based flexible environment for training on open-source tools and development of new statistics using new data sources

Author: Mátyás Mészáros

Presenter: Ricciato Fabio

Eurostat, Luxembourg



eurostat 

The conference is partly
financed by the European Union



The role of open-source tools in Official Statistics

- Flexibility in adapting to emerging data sources
- Transparency, collaboration, and accessibility
- Encouraging peer review and innovation
- Sharing code and re-using existing solutions
- Enhancing the quality of official statistics in the ESS



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

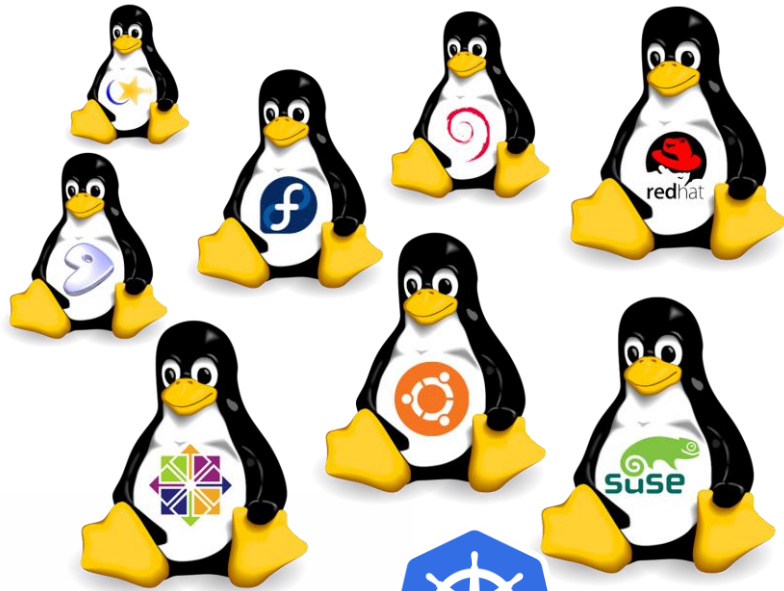


The conference is partly
financed by the European Union

The wild west of open-source tools and platforms



Studio®



Grafana



Open for Innovation

KNIME



python



Apache

Airflow



elastic



Metabase



APACHE
Spark™



PostgreSQL

MINIO



Superset



mongoDB

kubecost



kubernetes



Karpenter



PostGIS





Challenges with open-source tools in enterprise IT environment

- Providing solutions for a wide range of needs
- What tools to support? How to select them?
- Ensure security
- Onboard/training users for the continuously changing tools



Cloud Agnostic Datalab at Eurostat

- **Cloud:** using the public cloud enables flexibility in adapting resources to needs
- **Agnostic:** using Kubernetes as baseline to enable independency from the cloud provider
- **Datalab:** using Docker images and Helm charts to easily launch and combine data science tools



Open-source tools available in Eurostat Datalab (1/2)

IDE



Jupyter
Notebook



RStudio

Search analytics



ElasticSearch



Kibana

Data exploration & visualization



Apache
Superset



Metabase

Big Data analytics



Apache Spark



H2O



Open-source tools available in Eurostat Datalab (2/2)

Data storage, Databases, Linked data



PostgreSQL
PostgreSQL



Virtuoso



MongoDB



MinIO

Low-code data analytics, Annotation tool



Knime



Doccano

Workflow scheduling



Apache
Airflow



Mage.ai



Use cases

- **EU Big Data hackathon 2023:** 24 NSI teams and 4 EMOS teams were working intensively to create an early warning system in 4 days using financial transaction data
- **Eurostat Datacamp:** a week intensive use of the financial data internally in Eurostat, to see the potential of the financial transaction data used at the hackathon
- **Python training in Eurostat:** using JupyterLab for introductory course in Python
- **Traffic and Mobility Experimental Statistics:** using open data to create new and augment existing statistics
- **OJA Datalab:** the datalab in the Web Intelligence Hub to access and analyze Online Job Advertisement (OJA) data



Conclusions

- **Trainings** provide improved skillset to use new open-source tools with new data sources
- **Using open-source tools** gives the possibility to share and reuse transparent methods and workflows
- **Sharing common toolset in the ESS** gives a common knowledge base for new tools and new sources

**It contributes to improved
harmonization and quality of
European statistics**



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL



Thank you!



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat 

The conference is partly
financed by the European Union