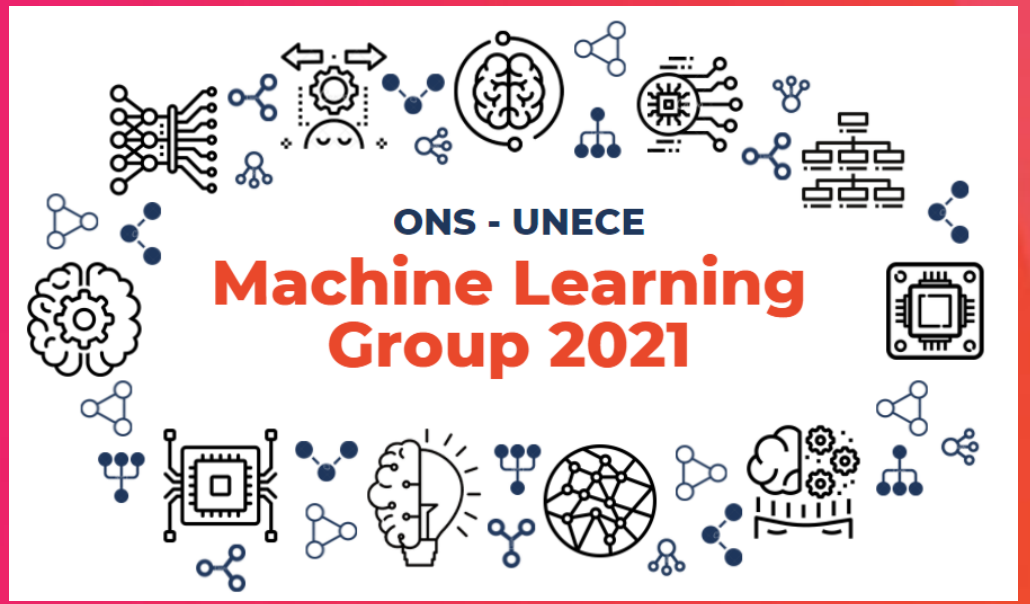
Data Science Campus

UNECE



2021 Workshop on the Modernisation of Official Statistics

Eric Deeben, ONS, UK

16 November 2021

ML: A tool for transformation

- Potential for more accurate, timely and efficient production
- ML is still an experimental technology in many NSOs
- Expertise is varied and scattered
- Need for trusted solutions

Objectives

Platform to facilitate

- Research to modernise official statistics
- Building capacity in machine learning
- Sharing knowledge (data, methods, etc.)

Community driven

- Projects and activities are led by members
- Supported by Group membership
- Decision informed by consensus (i.e. polls)

Public Good

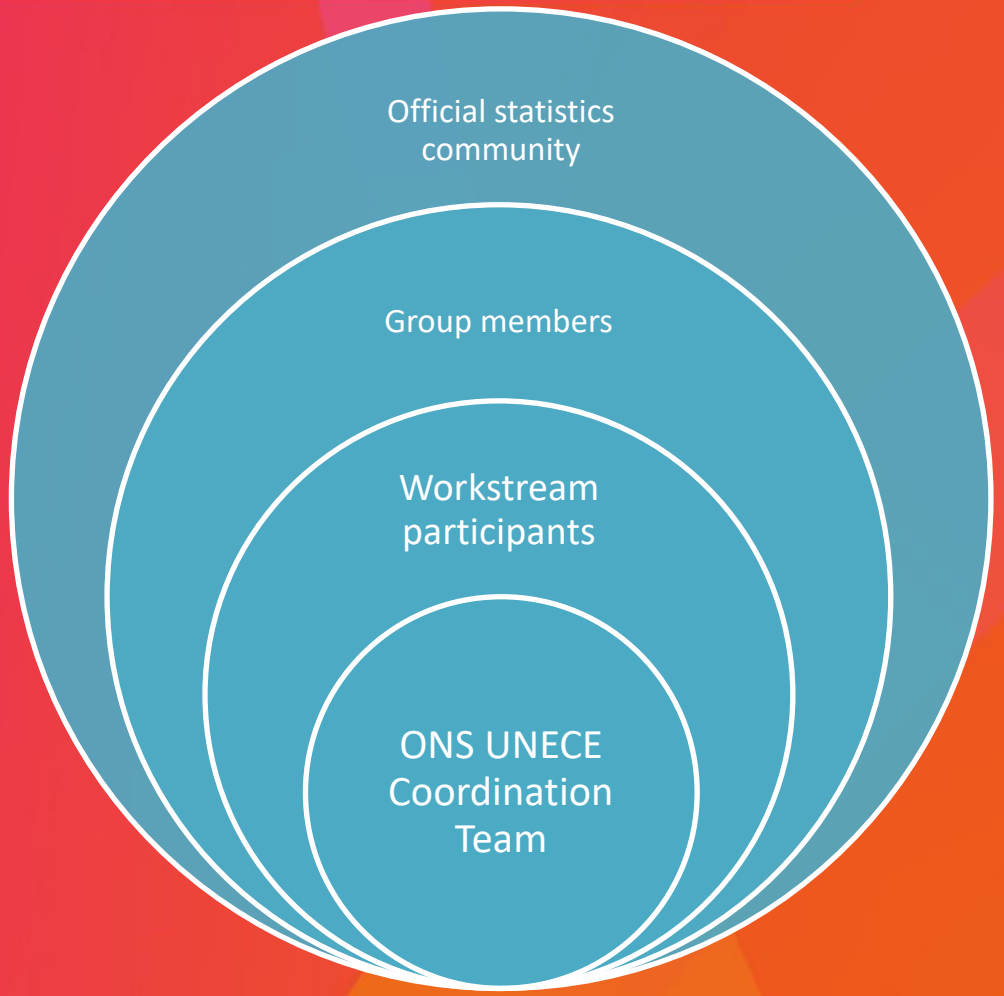
- Open to all official statistical bodies
- Accessible to various levels of expertise.
- Resources available to the wider community.

A Growing And Dynamic Community

- 30+ countries; 248 members
- Membership doubled in past year
- High attendance at meetings
- Demand for knowledge sharing and capacity building activities

Organisation

- ONS-UNECE partnership
- A core of active members drives programme
- Majority of members are observers
- Engagement with wider community



Activities

Knowledge Sharing	• Monthly meetings with expert presentations • External engagement at international conferences • Regular updates of ML news and opportunities
Research Collaboration	• Five workstreams + 18 research projects exploring issues from design to implementation • Findings shared on public website
Capacity Building	• Coffee and Coding sessions • Learning and training resources

Knowledge sharing

Presentation Highlights	Organisation
Fair Algorithms	CBS, Netherlands
Satellite Imagery for Wealth Index	BPS, Indonesia
Census Comment Classification & Homomorphic Encryption	Statistics Canada
Linking work injuries when names are missing	BLS, United States
AIS Data Task Team and Global Platform	UN SD
MLOps	ATO, Australia

"This group provides a much-needed space for connecting around ML for official statistics."

"The project has created the opportunity to spread the knowledge of ML among other members of the institution, benefiting the institution as a whole."

Evaluation Survey of members
October 2021

ML 2021 Group

Work Stream 1
From idea to valid solution (demonstration)

Activity 1.1 (Brazil)
Classification and aggregation for web-scrapped price data

Activity 1.2 (Turkey)
Classification of COICOP

Activity 1.3 (Denmark)
Multilingual Transformer based model for classification

Activity 1.4 (Chile)
Automated coding as shared service

C&C Theme

Activity 1.5 (Poland)
Multiple imputation

Activity 1.6 (Malaysia)
Estimating rubber production using satellite imagery

Activity 1.8 (Indonesia)
Use of satellite imagery for wealth index

Imagery Theme

Activity 1.9 (USA-BLS)
Expenditure estimate

Activity 1.10 (Belgium-Flanders)
Transferring knowledge

Activity 1.11 (Chile)
Route optimisation

Work Stream 2
From valid solution to production (operationalization)

Activity 2.1 (IMF)
Automated production tool

Activity 2.2 (Mexico)
Data lake architecture

Activity 2.3 (Mexico)
Workflow

Activity 2.4 (Sweden)
Technical platform for support ML

Work Stream 3
Data ethics and governance

Activity 3.1 (UK)
Data ethics and governance

Work Stream 4
Quality of training data

Activity 4.1 (Finland)
Quality of training data

Work Stream 5
Quality Framework for Statistical Algorithm

Activity 5.1 (Mexico)
Quality Framework for Statistical Algorithm

Data Science Campus

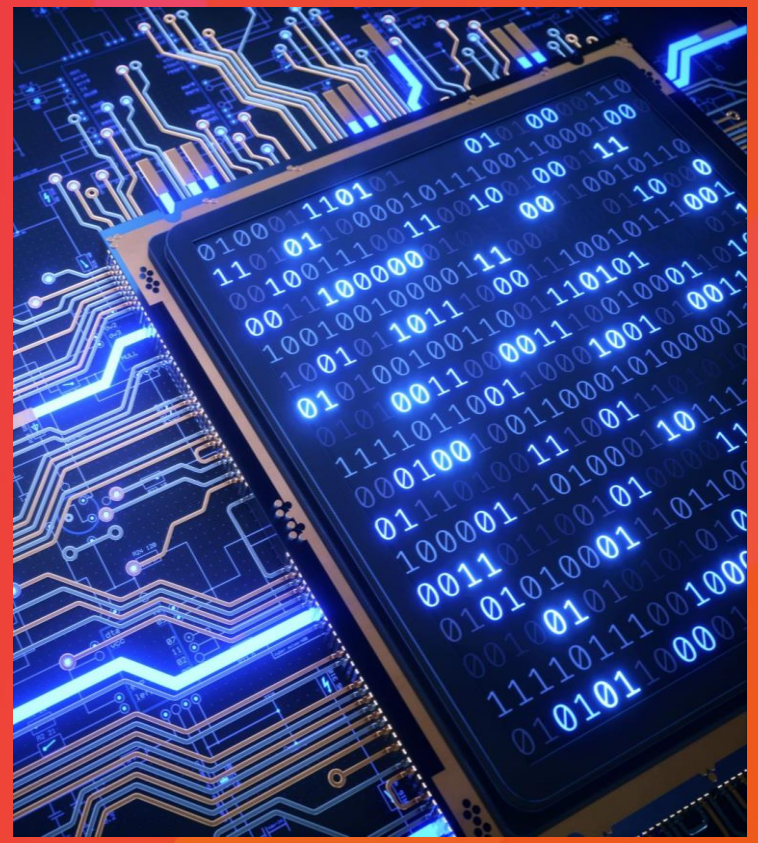
UNECE

Coffee and Coding

- Around 100 from 20 countries attended
- Good interaction, engagement and feedback
- Outcomes include:
  - ✓ increased confidence and knowledge in ML
  - ✓ higher awareness of how to build generalisable open source applications and code;
  - ✓ promotion of ONS and ML2021 group work
- Recording available for study and learning

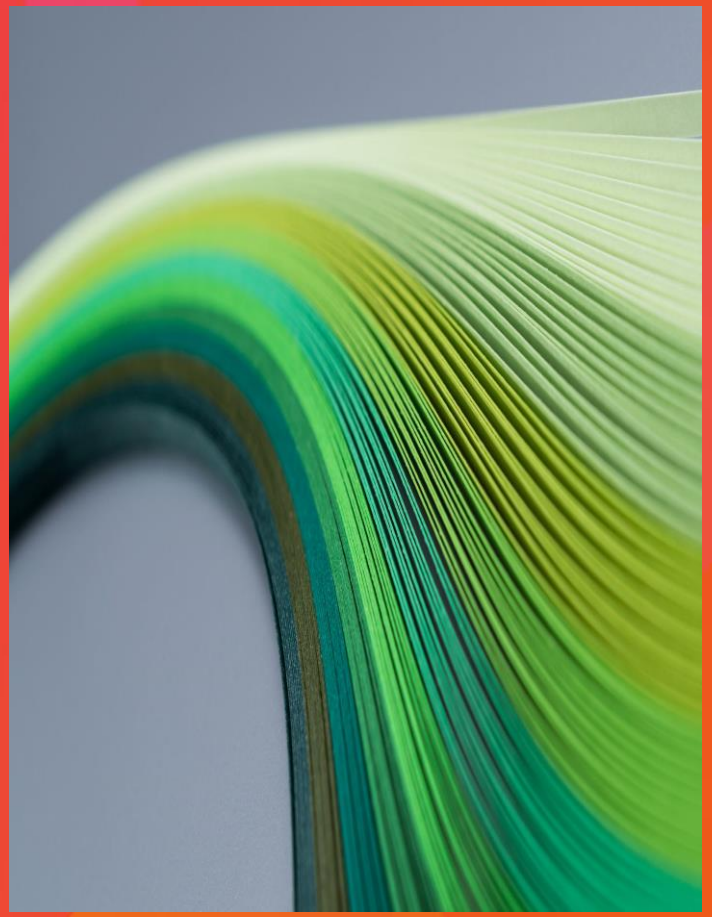
Impact on Production

- Advice on how to overcome challenges
- Knowledge of new tools and methods
- Exploring and testing ideas
- More targeted use of ML resources



Lessons learned

- ML is essential for analysis of big data sources
- Ways to keep up with the pace of change
- Systems need to be developed for addressing privacy and ethics concerns
- We have a better understanding of production challenges



Plans for 2022: Ambition and Objectives

- Bolder, bigger and more collaborative
- Focus on moving from proof of concept to production
- Other key areas: ethics, quality of training data.
- Increased emphasis on joint research
- Build capacity to meet ML requirements
- A hub for ML news and networking

Operational model

- Strategic guidance from lead data scientists
- New ‘Global Squad’ for research projects
- Workstreams → knowledge sharing, networking + discussion
- More technical expertise + interdisciplinary input
- Training programme targeted at different skill levels

Discussion Points

- Your thoughts on strategic priorities for 2022
- Operational model – how do we engage more data scientists?
- Invitation to join global research squad
- Call for proposals: research projects

Any other questions / comments?

Q & A

ML 2021 Webinar: 1300-1700 CET, Friday 19 November

For more information and to register contact ML2021@ons.gov.uk