

The European One-Stop-Shop for Artificial Intelligence and Machine Learning for Official Statistics (AIML4OS)

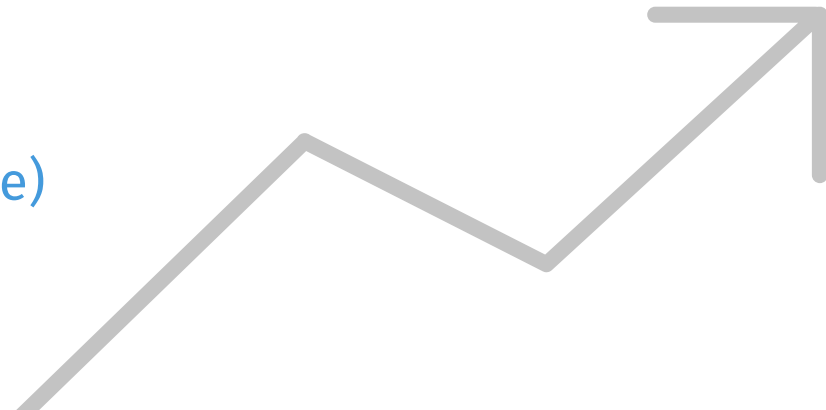
WP8 Use Case focused on data editing

UNECE Expert Meeting on Statistical Data Editing 2024, Vienna

Steffen Moritz (German Federal Statistical Office)

Katja Bürk (German Federal Statistical Office)

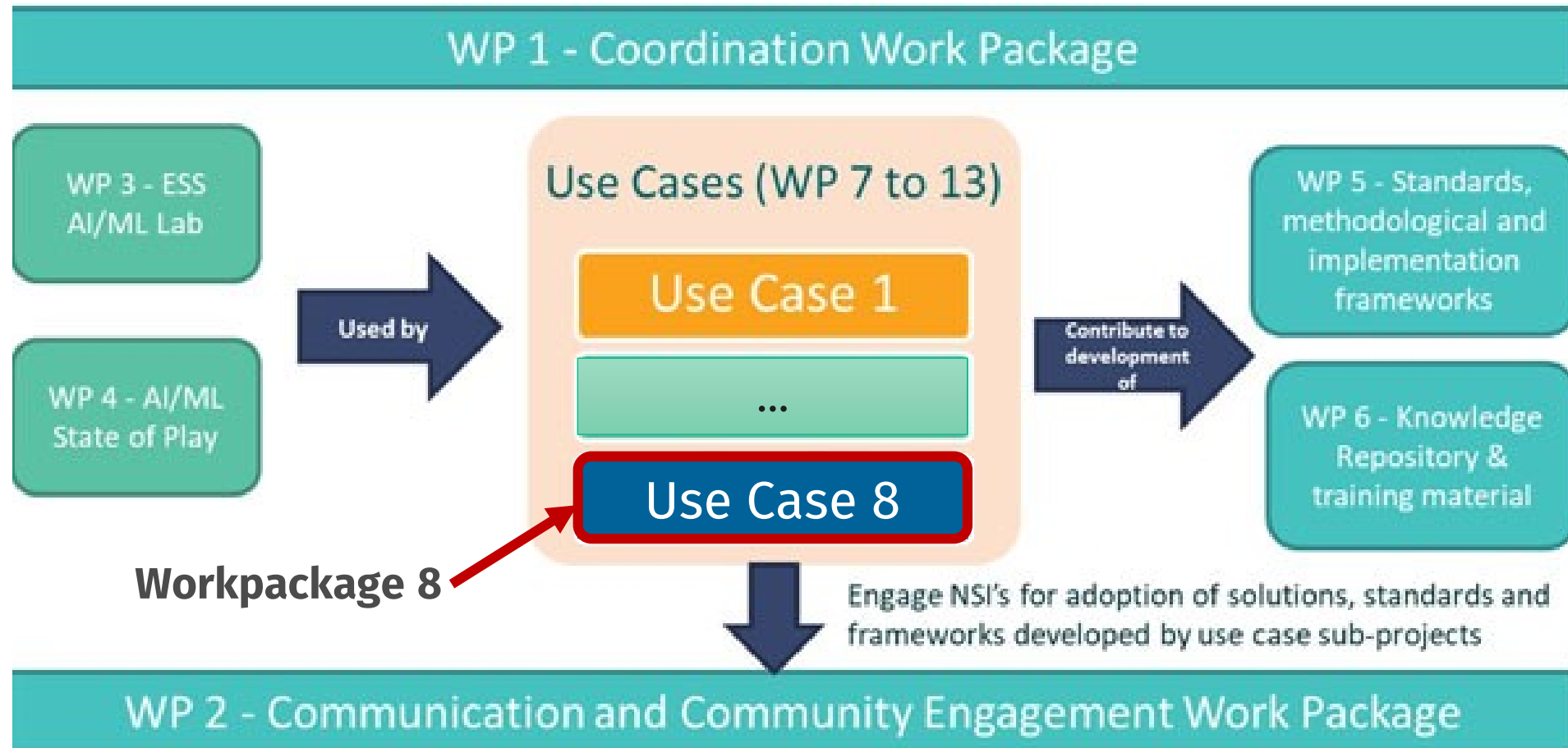
Florian Dumpert (German Federal Statistical Office)



Background: One-stop-shop for Artificial Intelligence and Machine Learning for Official Statistics (AIML4OS)

- ESSnet collaborative project involving **16** countries
- **Four-year** project - started in April 2024
- The project is developed through **13 Workpackages**
- Explore the use of **Artificial Intelligence/Machine Learning (AI/ML)** for the production of official statistics
- Implement innovative **solutions** for **statistical products** and **processes**.

Embedding in AIML4OS



WP 8: Statistically valid and efficient editing and imputation in official statistics by AI/ML – with a special focus on editing

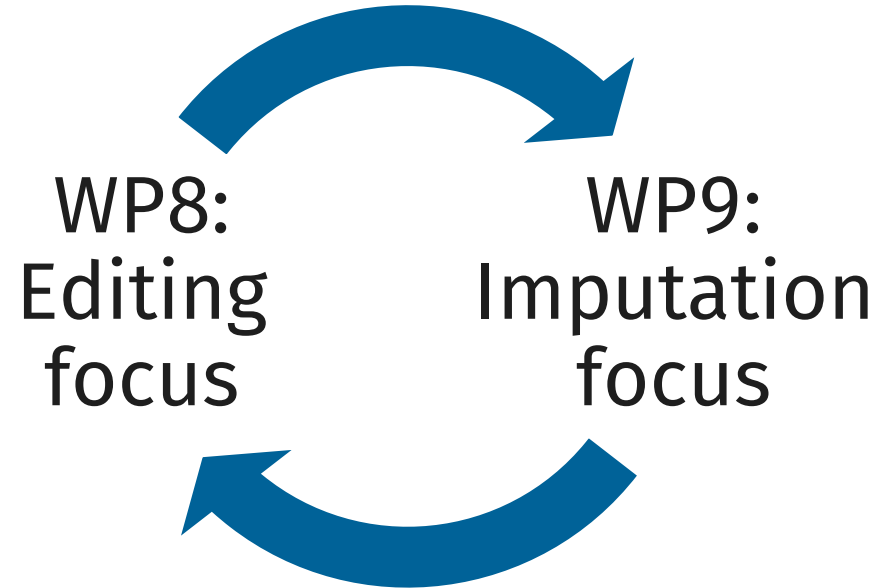
- Data Editing deals with issues that are essential for the quality of official statistics.
- Still lots of manual work and potential for quality improvement



GOAL

Improving the quality of official statistics through the innovative use of AI/ML in data editing and imputation, with a special emphasis on editing.

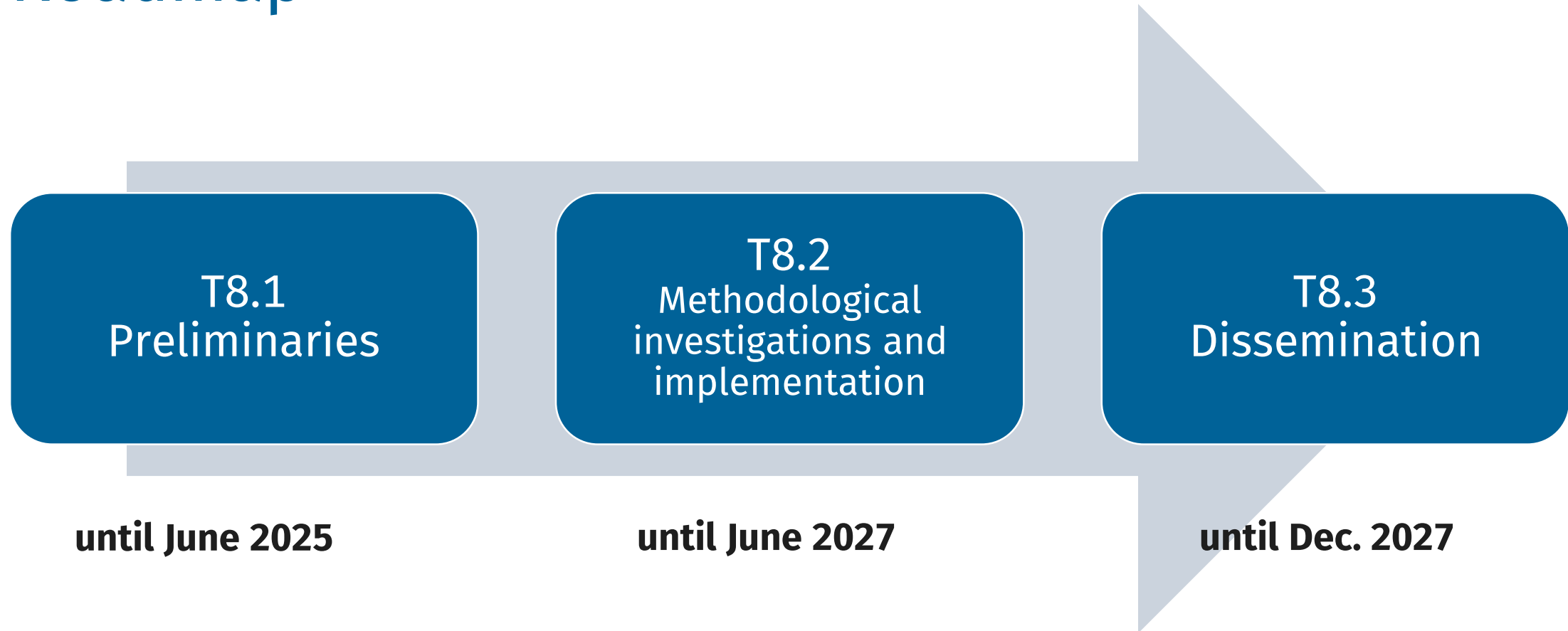
Close Cooperation with WP9



Participating Countries

GERMANY	Statistisches Bundesamt	Lead
AUSTRIA	Statistics Austria	Beneficiary
ITALY	National Institute Of Statistics	Beneficiary
NETHERLANDS	Statistics Netherlands	Beneficiary
NORWAY	Statistics Norway	Beneficiary
PORTUGAL	Instituto Nacional De Estatistica	Beneficiary
CYPRUS	Statistical Service of Cyprus	Associate
DENMARK	Statistics Denmark	Associate
SPAIN	Instituto Nacional de Estadistica	Associate
FRANCE	Institut National De La Statistique Et Des Etudes Economiques	Associate
IRELAND	Central Statistics Office	Associate
EUROPE	European Statistical Office	Observer

Roadmap



Deliverables

D 8.1

Methodological investigations
and implementation

(Prototype)

April 2027

D 8.2

Recommendations and
lessons learned

(Report)

Dec. 2027

Short-term Outlook

Work on Preliminaries

- **Literature review** about the **state-of-the-art**
- Getting to know about the **issues** and **ideas** of other **NSIs**
- **Systematic overview** on the **current experiences** and **challenges**
- **Decision** on which **problem(s)** to tackle from a **methodological** and an **implementation** point of view

Plans to have a paper ready by mid 2025, with the current working title:

“Assessing Data Editing in European Official Statistics: Challenges, Future Directions, and the Role of Machine Learning”

Country Use Cases

Country	Data to be edited	Project status
Ireland	Census data	In development
Italy	Administrative education data	In development
	Labour Force Survey	Idea stage
Norway	Business survey on R&D	In development
	Customs data	In development
Austria	Survey on travel behaviour	Beginning of project
Germany	Survey on structural business statistics	In development
Netherlands	Survey on Regional Employment Statistics	Beginning of project
Portugal	Employment Statistics	In development
France	Administrative data on wage and employment	Completed
	Business short-term statistics	Idea stage

Phase of the editing process

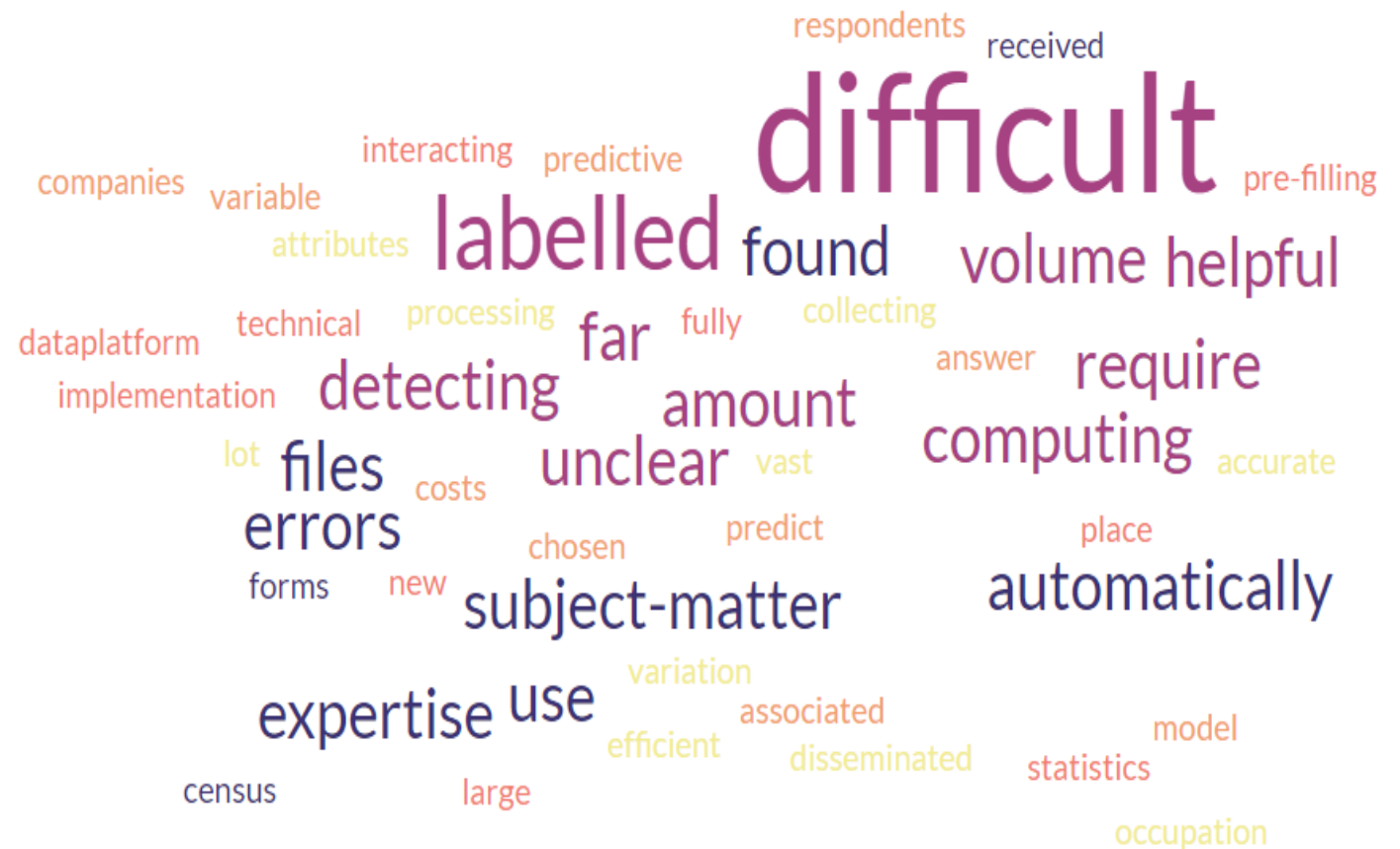
Country	Phase of the editing process: Error ...		
	Identification	Localisation	Correction
Ireland	✓		✓
Italy	✓	✓	
	✓		✓
Norway	✓	✓	
	✓		
Austria	✓	✓	✓
Germany		✓	✓
Netherlands	✓	✓	
Portugal	✓	✓	
France	✓		
	✓		

Reported Difficulties of the Projects

Common Themes

- Often very subject expertise specific problems
- Computing and Data Volume
- Labels and Quality Evaluation

Overall lots of different issues



Conclusions and Outlook

Key Factors for Maximizing Synergies in ML Data Editing Solutions

- **More Collaboration:** Seen on a higher level everyone is working on really similar problems
- **Code Sharing:** Would be great if we can profit more from the work of others
- **Aligned Outputs:** It is important to make sure at the beginning we produce outputs (reports, code), that really enables other to benefit from the experiences
- **Generalizability:** It is already complicated on an NSI level to take care to design generalizable solutions. Designing solutions that work on an EU-level is even more challenging.

Thanks for listening. Questions?

Steffen Moritz
steffen.moritz@destatis.de

