



EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL





EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Quality Assurance of the Re-coding to NACE Rev. 2.1, Combining Model and Manual Coding

Jacob Kasche – jacob.kasche@scb.se

Wictoria Widén – wictoria.widen@scb.se

Kira Gylling – kira.gylling@scb.se

Gustaf Strandell – gustaf.strandell@scb.se

Kristina Strandberg – kristina.strandberg@scb.se



eurostat 

The conference is partly
financed by the European Union



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Outline

- Background
- Coding methods
- Quality assurance
- Discussion



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat 

The conference is partly
financed by the European Union



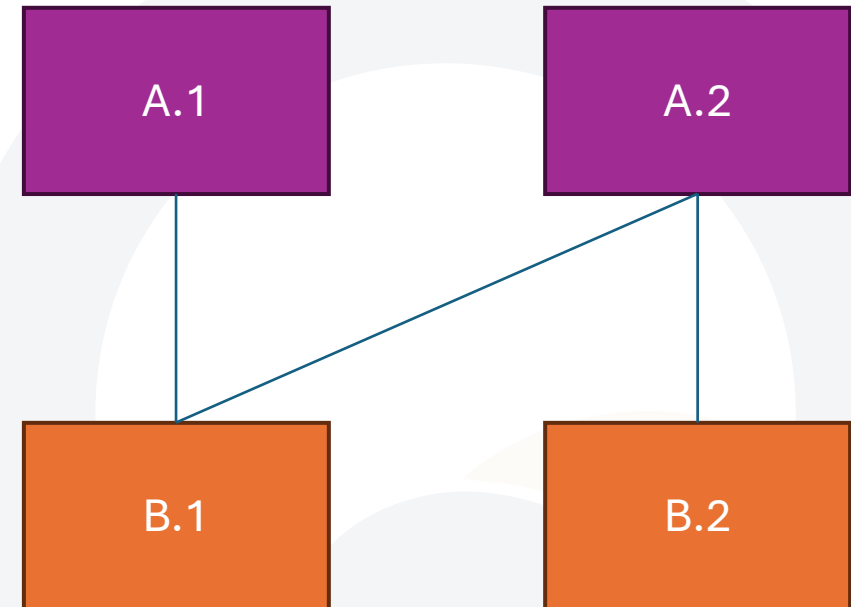


Background

NACE is a European classification of economic activity for statistical purposes

Eurostat provides a mapping between NACE Rev. 2.0 and NACE Rev. 2.1

Re-coding is problematic for one-to-many cases i.e., $A.2 \Rightarrow (B.1, B.2)$





EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Coding methods

Because of budget restrictions we suggest to use more automatic coding methods

Suggested methods:

1. Coding using a key
2. Large Language Models (LLM)
3. Imputation e.g. with ML
4. Manual coding



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Quality Assurance Methods

Main goal is to effectively allocate the resources

Measurement methods

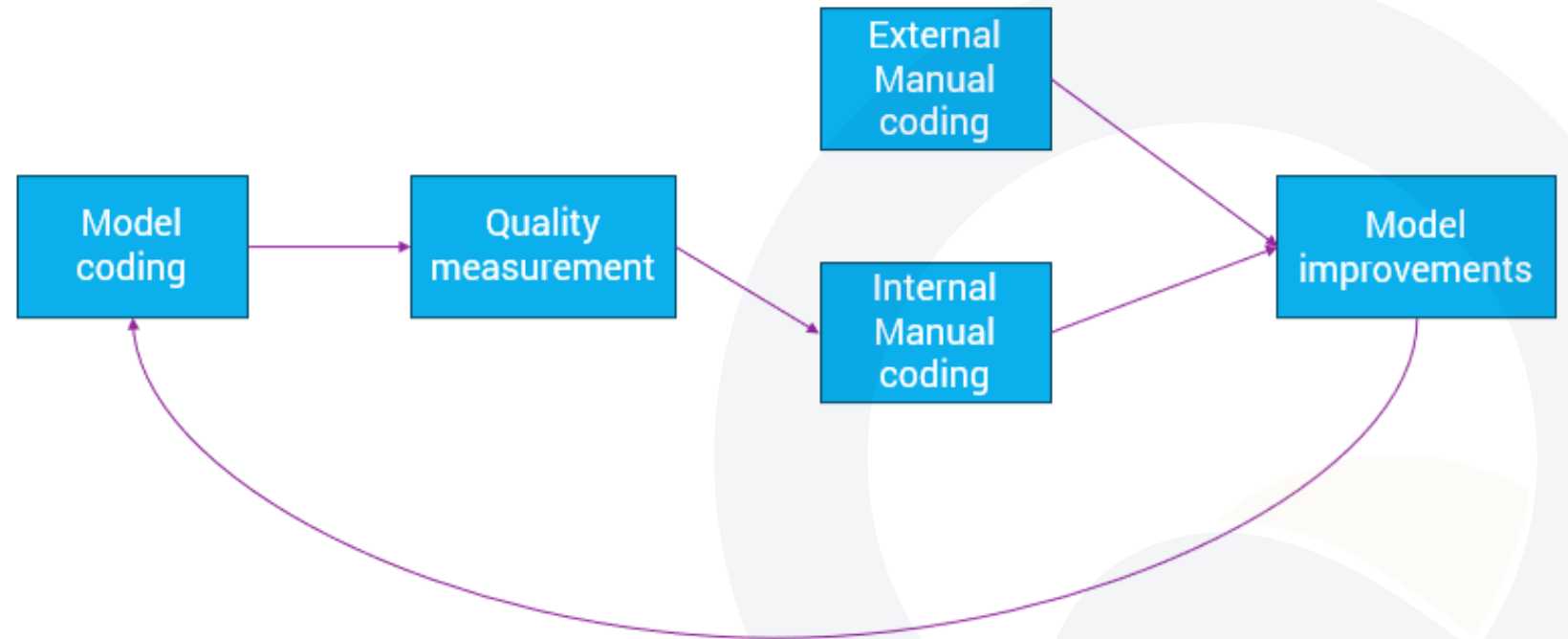
1. Model inference
2. Design inference

Improvements

1. Manual coding supported by models
2. Re-use of manually coded data

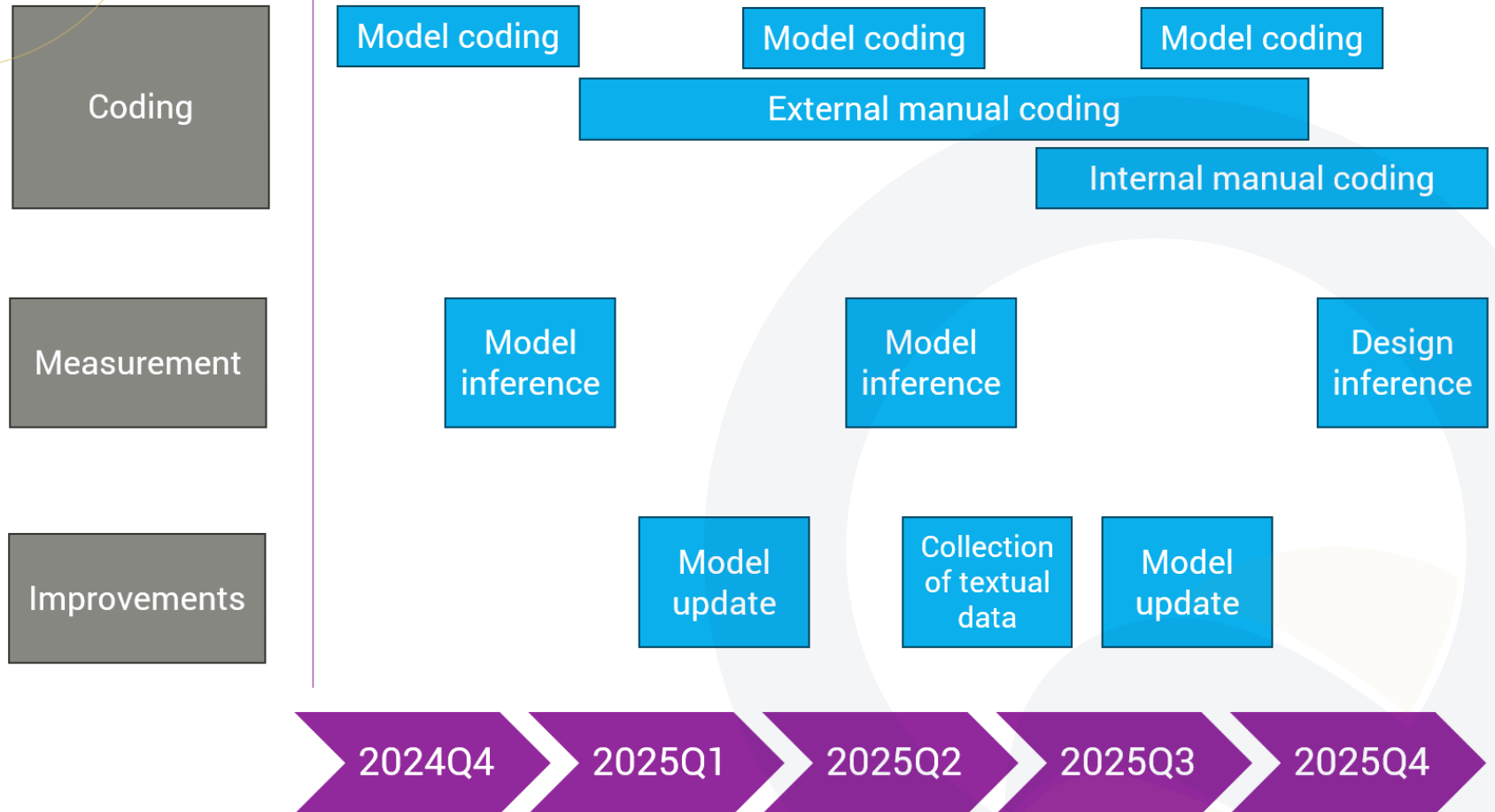


Quality assurance process





Timeline





Discussion

- Risk of transferring cost
- Quality in the coding is dependent of quality in data, coding methods and priority system
- Relate the uncertainty from the re-coding to the total uncertainty



EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL