# Mapping between Norwegian municipalities and dialect regions

'dialect_mapping.csv' provides a mapping between Norwegian municipalities and dialect regions, and can be used, e.g., to infer the dialect region of a speaker in a speech dataset based on their place of birth. This mapping is developed by Phoebe Parsons at the Norwegian University of Science and Technology. The mapping includes both old and new municipalities and counties, and maps to two different dialect classifications:

1. Named dialects: Skjekkeland, Martin. 1997. Dei norske dialektane. Høyskoleforlaget

2. Numeric dialects: A classification from the SpeechDat project.

## Description of the fields in the CSV:

- 'old_muni': Historical municipalities, i.e. municipalities which ceased to exist after the municipality reforms around 2020, as well as some relevant municipalities which ceased to exist at an earlier date. Municipalities which still exist, are also given in 'old_muni'.

- 'new_muni': Current municipalities (as of 2023)

- 'old_county': Counties before 2020

- 'new_county': Counties 2020-2023

- 'new_county_2024': Counties after 2024

- 'named_dialect': Dialect region according to Skjekkeland

- 'numeric_dialect': Dialect region according to SpeechDat

## Phoebe Parson's description of how the mapping was created:

"The alignment between municipalities ("kommuner"), counties ("fylker"), and dialects was developed through manual effort. It began by looking over dialect maps (such as can be found in Skjekkeland's *Dei norske dialektane* (page 276)) and attempting to align the currently existing kommune and fylke to them. Once I was satisfied with the mapping I extended it to historical *kommuner* and *fylker*. To do this I referenced Wikipedia pages for the historial *fylke* and which *kommune* they contained and updated the list. The new 2024 updates were created via programatically parsing information in the table on the Regjering website Nye kommune og fylkesnummer.

The named dialects used are from Skjekkeland's book (with the 4 Eastern regions collapsed into one). Additionally, the names were translated from Norwegian Nynorsk to Bokmål. The numbered dialects are from a survey produced by SpeechData.

The first column in the CSV file contains the historic municipalities and the second columns the modern municipalities. Where possible this creates a mapping between historic and current municipalities. That is, the municipality in column one became the municipality in column two. Occasionally column one is left empty. This is for instances where a new municipality, one without a historical equivalent, was included. The same alignment through time can be found for the three county columns as well."