EDS SEMINAR Record Linkage in statistical analysis

On **Monday 15 May 2023**, 11:00-11:35 AM we will have an EDS Seminar organised by the **BigStatistics section of Mark van de Wiel**.

Kayané Robach will tell us all about:

Record linkage in statistical analysis

Combining different data sources allows us to explore innovative questions including counting casualties or conducting healthcare monitoring studies. Identifying information belonging to the same person is tricky however, as often no unique identifier is available. The aim of our work is to produce causal effect estimates that incorporate the uncertainty coming from the merging process. In particular, we work in a setting where one dataset may contain covariates and treatment allocation, whereas the other contains the outcome, e.g. survival.

The aim of record linkage is to uncover the matching indicators between observations collected on different occasions using partially identifying variables. In our approach, we draw the rules governing the underlying data generating process, taking into account typos, mistakes or missing data in the registers to construct the linkage model.