



# Parallelised analysis of large-scale bibliometric data (with Dask in Python, DuckDB and DBeaver in SQL) Using example of ORCID 2019 XML files

**Aliakbar Akbaritabar**

Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany  
akbaritabar@demogr.mpg.de

# Recorded Video of Tutorial



- ▶ **Repository of codes/data:**

<https://github.com/akbaritabar/dask-duckdb-dbeaver>

- ▶ **Video:** <https://youtu.be/pYDVrBcIuYI>

- ▶ **00:00 - 02:45;** Introduction
- ▶ **2:45 - 11:04;** Requirements and installation (skip this if you have already installed everything based on instructions in the ReadMe on repository)
- ▶ **11:05 - 38:00;** Steps in using Dask/Python and DuckDB/DBeaver and results