Mapping global cultural diversity

Building a free and open database of the world's languages

Department: Copernicus Institute of Sustainable Development
Research group: Environmental Sciences
Supervisor: Dr. Sietze Norder
Email address: s.j.norder@uu.nl

Project description

There are nearly 7000 languages spoken across the world, but nearly half of them are endangered (Bromham et al. 2022). While cultural diversity is valuable in itself, a plurality of cultural perspectives and indigenous worldviews are also crucial for biodiversity conservation and sustainable development (Sustainable Development Goal 4.7; Díaz *et al.*, 2019). While earlier studies have noted the spatial overlap between cultural diversity and biodiversity (Garnett et al., 2018; Hua et al., 2019), a deeper understanding of these patterns is hampered by limited data availability. At the moment, there are two data sources that provide geospatial representations of language locations, both having their limitations. The first database, Glottolog, is used in most projects but only includes point locations, limiting the kinds of analyses that can be conducted. A second database, Ethnologue, does contain polygon data, but these data are locked behind an expensive paywall, raising ethical and legal issues regarding data accessibility (for language speakers and academics alike) and reproducibility.

With this Bright Minds project, we strive to make language distribution data FAIR – Findable, Accessible, Interoperable, and Reproducible (Wilkinson et al., 2016). In a previous project, numerous (historical) maps and atlases of language distributions have been gathered, digitized, georeferenced, and vectorized. In the Bright Minds project, the student will build on these efforts and contribute to building a global spatially-explicit database of language distributions, and cross-validate it.

Specifically, the student will carry out three tasks. First, manually vectorize historical maps using QGIS in collaboration with research assistants from the university of Zurich. Secondly, adapt an existing function in the glottospace R package to match language names to a unique identifier in Glottolog. Third, assess the degree of spatial congruence of the newly created polygons and interpolated point data.

Job requirements

I am looking for a student who has a background in GIS (Geographic Information Systems) and (geospatial) data analysis, or the desire to become more skilled in this field. Most of the GIS work will be done in QGIS. Further (geospatial) data processing and analysis will be done in R, and can build on earlier work (Norder et al. 2022). The work can either be done remote or on campus, regular meetings will be held with the supervisor.