

## **BUILDING THE DATA REPOSITORY FOR LOCAL-SCALE EVAPOTRANSPIRATION DATA USING DATA SCIENCE**

Yessica Ramirez-Yara <sup>\*1</sup>, Agnieszka A. Malinowska <sup>1</sup>, Ryszard Hejmanowski<sup>1</sup>

<sup>1</sup> *AGH University of Krakow, Faculty of Geo-Data Science, Geodesy, and Environmental Engineering, Al. Adama Mickiewicza 30, 30-059 Krakow, Poland*

\*Corresponding author: [yyara@agh.edu.pl](mailto:yyara@agh.edu.pl)

---

**Keywords:** Data repository, Git version control, Data mining, Evapotranspiration.

Climate warming is gradually drying out the land. Especially in agricultural and forest regions, this leads to a gradual loss of natural and economic potential in many countries. The rapid surface run-off of rivers after sudden but infrequent rainfall makes it difficult to retain moisture. Evaporation phenomena from both soil and plants, i.e. evapotranspiration, play a huge role in the drying out of land around the world. To better understand and model phenomena related to the consequences of climate change, access to considerable volumes of meteorological data is essential. The aim of the study was to build a repository of data indexed in space and time, of meteorological, hydrological, anthropological, and environmental variables. This repository was built for areas around the world, based on available open data libraries. The repository development and data integration relied on web scraping techniques in open-access portals, advanced raster geoprocessing tools and spatial statistical analysis, using specialized libraries in Python and R languages thought Anaconda environments, were used to massively gathering data, resample variables and automate the spatial data mining tasks involved in collecting data for the repository. This process was supported by the CI/CD software concept, utilizing Git version control tools via GitHub.