# Open and reproducible global land use classification

Daniel Nüst (1), Tomáš Václavík (2), and Benjamin Pross (1)

(1) 52°North Initiative for Geospatial Open Source Software GmbH, Münster, Germany (d.nuest@52north.org), (2) Department of Computational Landscape Ecology, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany

Researchers led by the Helmholtz Centre for Environmental research (UFZ) developed a new world map of land use systems based on over 30 diverse indicators (http://geoportal.glues.geo.tu-dresden.de/stories/landsystemarchetypes.html) of land use intensity, climate and environmental and socioeconomic factors. They identified twelve land system archetypes (LSA) using a data-driven classification algorithm (self-organizing maps) to assess global impacts of land use on the environment, and found unexpected similarities across global regions. We present how the algorithm behind this analysis can be published as an executable web process using 52°North WPS4R (https://wiki.52north.org/bin/view/Geostatistics/WPS4R) within the GLUES project (http://modul-a.nachhaltiges-landmanagement.de/en/scientific-coordination-glues/).

WPS4R is an open source collaboration platform for researchers, analysts and software developers to publish R scripts (http://www.r-project.org/) as a geo-enabled OGC Web Processing Service (WPS) process. The interoperable interface to call the geoprocess allows both reproducibility of the analysis and integration of user data without knowledge about web services or classification algorithms. The open platform allows everybody to replicate the analysis in their own environments.

The LSA WPS process has several input parameters, which can be changed via a simple web interface. The input parameters are used to configure both the WPS environment and the LSA algorithm itself. The encapsulation as a web process allows integration of non-public datasets, while at the same time the publication requires a well-defined documentation of the analysis. We demonstrate this platform specifically to domain scientists and show how reproducibility and open source publication of analyses can be enhanced. We also discuss future extensions of the reproducible land use classification, such as the possibility for users to enter their own areas of interest to the system and generate summary statistics relating the particular area to the land system archetype. Such an extension demonstrates the advantages of open geoprocesses, because the user does not need to replicate the whole workflow, which included considerable data preparation steps, and can still access an analysis result tailored to his needs.

The LSAs are the basis for science-based policy recommendations for sustainable land management and yield improvement at a global scale. The reproducibility of the study strengthens the scientific work and the open source platform allows scientists to adapt and extend it to provide new original contributions to sustainable land use management.