Harnessing the Geospatial Science Revolution to Advance Sustainability Research

Campbell D. Watson (cwatson@us.ibm.com), Hendrik Hamann, Komminist Weldemariam, and Johannes Schmude (IBM Research, USA)
Blair Edwards and Anne Jones (IBM Research, UK)

Background/Objectives. The ballooning volume and complexity of geospatial data is one of the main inhibitors for advancements in climate and sustainability research. Researchers often create bespoke and time-consuming workflows to harmonize datasets, build and deploy AI and simulation models, and analyze the output. It is increasingly evident that these workflows and the underlying infrastructure are failing to scale and exploit the massive amounts of data residing in multiple data centers. While there have been attempts to consolidate relevant geospatial data and tooling into a single cloud infrastructure, we argue that the future of climate and sustainability research must rely on networked and federated systems. These federated systems will be key to building interoperable digital twins for climate and sustainability.

Approach/Activities. Here we introduce a geospatial software system capable of efficient federation between data nodes. It is built upon IBM PAIRS (Hamann and Liu, 2021) and the Climate Impact Modeling Framework (CIMF; Edwards et al 2022) and has four salient features: i) a scalable cloud-based deployment capable of handling hundreds of Petabytes of data; ii) “in-data” analytics and computation to minimize data movement; iii) composable and reusable modules to simplify modeling workflows; and iv) efficient data harmonization to make heterogeneous data homogeneous. Each of these features is critical for building the “Climate Network”—large-scale intelligent federation between distributed data nodes that drastically accelerates sustainability and climate resiliency research.

Results/Lessons Learned. We demonstrate how the system architecture and associated tooling for data discovery and visualization supports these features, and end-to-end examples of the system performing data-level federation, AI-based flood detection from Sentinel-1, and AI-based wildfire risk projections in future climates.

Hamann, H, and Liu. S. (2021) IBM PAIRS: Scalable big geospatial-temporal data and analytics as-a-service. In: Werner, M., Chiang, YY. (eds) Handbook of Big Geospatial Data. Springer, Cham.
Edwards, B., Fraccaro, P., Stoyanov, N., Bore, N., Kuehnert, J., Weldemariam, K., and Jones, A. (In Press) CIMF: Climate Impact Modeling Framework. Fragile Earth, AI for Good, ACM KDD 2022.