Evolution of data infrastructure for effective integration and management of environmental and ecosystem data

Siddeswara Guru\textsuperscript{1}, Gerhard Weis\textsuperscript{1}, Wilma Karsdorp\textsuperscript{1}, Andrew Cleland\textsuperscript{1}, Jenny Mahuika\textsuperscript{1}, Edmond Chuc\textsuperscript{1}, Javier Sanchez Gonzalez\textsuperscript{1}, and Mosheh Eliyahu\textsuperscript{2}

\textsuperscript{1}TERN, University of Queensland, Brisbane, Australia
\textsuperscript{2}TERN, University of Adelaide, Adelaide, Australia

The Terrestrial Ecosystem Research Network (TERN) is Australia's national research infrastructure to observe, monitor and support the study and forecasting of continental-scale ecological changes. TERN data are classified under two themes: Ecology and Biogeophysical.

The Ecology theme relates predominantly to plot-based ecological observations conducted as a one-off, repeated surveys and sensor-based measurements. The Biogeophysical theme-related data collections are inclusive of point-based time-series eddy-covariance based micrometeorological measurements from flux towers; and continental and regional scale gridded data products related to remote sensing, soil and landscape ecology.

Integrating and querying data from different data sources are complicated. Furthermore, the advancement of technology has transformed the mode of data collection. For instance, mobile sensors (drones) of various sizes are used more in recent times to sample the environment. The user-centric data handling mechanisms of different types of datasets are dissimilar, requiring heterogeneous data management practices alongside ease of access to data for users bundled with tools and platforms to interrogate, access, analyse and share analysis pipelines.

TERN is developing data e-infrastructure to support holistic capabilities that not only manage to store, curate and distribute data. But, enable processing based on user needs, linking consistent data to various analysis tools and pipelines and acquisition of data skills. The infrastructure would allow collaboration with other national and international data infrastructures and ingest data from partners including state and federal government institutes by adopting domain standards for metadata and data management and publications.

For effective data management of plot-based ecology data, we have developed an ontology-based on O&M and Semantic Sensor Network Ontology with an extension to support basic concepts of ecological sites and sampling. Besides, controlled vocabularies for observed properties, observation procedures and standard lists for taxa, geology, soils etc. will supplement the ontology.

The biogeophysical data is managed using domain standards in the data and metadata
management. Each of the data products is represented in a standard file format and hosted in an OGC standard web services. All datasets are described and catalogued using ISO standards. An overarching discovery portal allows users to search, access and interact with data collections. The user's interaction with data can be at the collection level, on a spatial map and via web services and Application Programming Interface (API).

TERN has also developed a cloud-based virtual desktop environment, CoESRA, accessible from a web browser to enable easy access to the computing platform with tools for the ecosystem science community. The advantage is that it allows access to all TERN data in a compute environment for performing analysis and synthesis activities from a single managed platform.