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

Siddeswara Guru, Gerhard Weis, Wilma Karsdorp, Andrew Cleland, Jenny Mahuika, Edmond Chuc, Javier Sanchez Gonzalez, Mosheh Eliyahu, Tina Parkhurst

s.guru@uq.edu.au
Outline

• Introduction to TERN
• Requirements
• Approach
• Summary
TERN Purpose\textsuperscript{1}

- National infrastructure for collecting, collating, storing and sharing Australia's terrestrial ecosystem data sets and knowledge.

\textsuperscript{1}TERN is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy from 2009
TERN in Operation

- Satellite remote sensing products
- Land cover dynamics and phenology
- Vegetation composition and structure
- Fire dynamics and impacts
- Continental Soil & Landscape data

Plot-based surveillance monitoring
- Soil sample, leaf tissue samples, LAI, Basal area

Carbon, energy, water fluxes
- Phenocams
- Acoustic sensors
- Flora population
NATIONAL DATA COLLECTION: FIELD, AIRBORNE, AND SATELLITE

TERN's national infrastructure includes on-ground, airborne and satellite data collection with data integration and delivery infrastructure that is designed to deliver information, knowledge and tools that are meaningful at local, regional, continental and global scales.

DATA INTEGRATION, ANALYSIS, AND DELIVERY

more than 600 ecosystem observing sites
more than 2500 open datasets
more than 50 national and international partners
more than 90 year continuity for datasets
more than 1000 peer-reviewed papers using TERN data
Users Requirements

• One place to submit and access data
• Need robust search capability
• Need to visualize data before downloading
• Need programmatic access to data
• Improve usability of data
• Ease of access and use
• Analytical capabilities
TERN data landscape analysis

– Continental scale gridded data products: Remote sensing, Soil and landscape products
– Plot-based surveillance monitoring: Soils, vegetation
– Intensive monitoring
  • Flux tower - sensors
  • Phenocam - sensors
  • Acoustic monitoring - sensors
  • Plot-based vegetation monitoring – human observation
  • Calibration and validation data for remote sensing - sensor
– Institution survey data (state government agencies) – Human Observation
Approach

• **Objective 1: Build an integrated platform for plot-based ecological data**
  
  – Build a scalable platform to host all plot-based ecological data
  
  – Provide an integrated view of data and ability to query parameters across different surveys and institutes.
  
  – Support data exchange between different initiatives.
• Objective 2: spatial and sensor data infrastructure to host TERN platform data
  – Effective management of Spatial data with Cloud-native solutions.
  – Data are described in a metadata standard
  – Use standard web services for data access
  – Storage resource to host large volume of data
Objective 3: Improve discovery and access of data

- Use ontology and controlled vocabbs to describe platform, data providers, observed parameters, taxonomy.
  - Align to Semantic Sensor Network Ontology, extended to represent ecology plots and associated data collection.

- Access to metadata authoring tool
  - ISO 19115-3 compliant metadata

- Data search and indexing
- Data visualization
- Data access via API
- Build a trust model for TERN data
• **Objective 4: compute access closer to data**
  – Make data cloud-enabled (data accessible to compute infrastructure)

• **Objective 5: data skills program**
  – Develop tutorials on accessing and using of data
  – Proactive engagement with users to address their needs

• **Objective 6: data impact and user engagement**
  – Measure the usability of data
Data integration platform for plot-based ecology data
High level Architecture of Overall System
high-level infrastructure for information flow

**Data Source**
- water, energy, carbon flux
- ecological plot: long term, intense, inventory monitoring
- ecological sensors
- soil, landscape
- remote sensing
- climate variables
- ecological: images and videos

**Application Services**
- THREDDS
- ERRDAP
- Geonetwork
- GeoServer
- semantic data platform
- Solr
- Open data cube

**Standards**
- Data Standards
  - WMS
  - WCS
  - WFS
  - OpenDAP
  - Veg data exchange
  - WebDAV
- Metadata
  - ISO 19115-3

**Access and Discovery**
- TERN Portal
- External APIs
- Virtual Labs
- cloud-enabled Analytical Platform
- ARDC RDA
- DataOne
- GEOSS Portal
- cloud-enabled Platform

**Governance**
- security
- service level agreement
- licensing policy
- usage and reporting
- authentication and authorisation
Engagement with different initiatives
Harmonisation

- Ontology and Vocabularies are key
  - Platform -- based on SOSA (align with GCMD)
  - Instruments – based on SOSA (align with GCMD)
  - Spatial regions – Australia’s Bioregions (IBRA), Ecoregions, States and Territories
  - Spatial resolution -- RDF (align with GCMD terms)
  - Temporal Resolution – RDF (align with GCMD terms)
  - Content type – RDF (align with GCMD terms)
  - Project – based on schema.org
  - UoM – QUDT ontology
  - Observed properties – Building our terms (align with EnvThes)
  - Methods/procedures – Building Vocabulary
  - Organisations – based on schema.org
  - People – based on schema.org
Information model should represent all the aspect data
Analytical Platform

TERN is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy.
Cloud-based virtual desktop (coesra.tern.org.au)

• A Cloud-based virtual desktop environment accessible via a web browser
• Tools available
  – Data Science - Jupyter labs, RStudio
  – Programming tool - Canopy
  – Geospatial – QGIS
  – Scientific Workflows: Kepler, KNIME
  – NetCDF viewer – Panoply
  – Data cleaning – Google refine
  – Ecology related – Biodiverse, MacroecoDesktop
  – Data sharing - Dropbox, ownCloud
Overarching Operational Goal

For every data collected and collated by TERN

- Where it was collected – place and platform
- How it is collected - procedure
- Who collected – organization/people
- What was collected – observed property, UoM
- How can user access – web access
- How can user use data – Data Skill
- platform for analysis – managed cloud analysis platform
Acknowledgement:
TERN Data Services Team

TERN data Access: https://portal.tern.org.au
TERN Vocabs: https://linkeddata.tern.org.au
CoESRA VDI: https://coesra.tern.org.au
TERN GitHub: https://github.com/ternaustralia/
TERN Skills: https://ternaus.atlassian.net/wiki/spaces/TERNSup/overview