Virtual Laboratory for Earth Observation

What's the use?

- Ocean dynamics
- Agronomy
- Ocean productivity
- Soil conservation
- Marine biogeochemistry
- Fire history
- Fish ecology
- Climate change
- Forecasting

Sensor

Day-time

Night-time

Local

Terrestrial

Annual

Historical Record

Timeliness

Historical Record

Continental

Marine

Daily
Divided, we struggle on…

Data Volume + Process specialisation = Community Fragmentation

- Acquisition, Storage & Processing In Common (e.g., Unpacking, Geolocation, Calibration)
- Specialised Processing (e.g., NDVI, LAI, SST, Chl, Compositing…)
- Generic Tools: Subsetting, process control, visualisation

NCRIS/EIF Paradigm Change is an Opportunity

1. Funding science domain research infrastructure
   - Marine – IMOS has a National focus
   - Terrestrial – TERN has a National focus
   - Participants in BOTH domains are thinking at large scale AND getting to know each other…

   - National Computational Infrastructure (NCI) provides
     1. The scale to handle these data sets (12k CPU + storage)
     2. The shared access - It is outside all our institutional firewalls so it overcomes one of the biggest practical obstacles to collaboration

3. Seeding uptake through NeAT (ARCS+ANDS)
   - MACDDAP Crawler+Aggregator, Bioflow, AusCover Workflows (RS-YABI)

4. A roadmap in the future – RDSI + NeCTAR
Phase 1 - National Datasets for Key Sensors

Concept:
- Improve storage efficiency
- Establish basic processing tools
- Build selected specialised tools
- Start with MODIS + Landsat

Partners:
- NCI + IMOS + TERN
- Geoscience Australia
- CMAR + CLW
- CDU + Curtin

---

Phase 1a. MODIS (2000+)

Ocean
Vegetation
Land-Air
Weather
Fire
Coastal
Phase 1b. Landsat (1979+)

- 300+ TB Australasian archive, every 16 days
- Geoscience Australia leads this work
- See Poster “Unlocking the Landsat Archive”
- ~2 years to build this

Coming Soon (3-6 months)

- NOAA/AVHRR (1980s+)
  - 10 TB, 200000 scenes

- SeaWIFS (1997-2010)
  - 0.5 TB
  - re-uses MODIS software
Phase 2 – Community Tools

- **Data services (DAP, OGC protocols, for Google Earth)**
  - http, Hyrax, Thredds Data Server…
- **Catalogues and data access**
  - IMOS/AODAAC + NeAT MACDDAP Aggregator
- **Generic processing management**
  - NeAT Bioflow + NeAT AusCover Workflows
- **Presentation, query and analysis tools (Twinkleware)**

IMOS/AODAAC – Automatic Cataloguing & Access for Gridded Data

- **IMOS Project, NeAT aggregator cpt.**
- **Deployed to support**
  - IMOS portal
  - Tern/Auscover
- **Interactive/batch mode**
- **Supports**
  - mass data access
  - individual scenes

![Diagram of IMOS/AODAAC system](image)
Workflow Tools: RS-YABI

- RS workflows typically linear & per-scene: big automation opportunity
- YABI engine originally developed in Bioinformatics domain to exploit an abstraction of HPC processing and storage (NeAT Bioflow)
- Adapted for remote sensing under a NeAT project with TERN/AusCover

- Strong HPC capability
- Easy to incorporate heterogeneous processing steps (different languages, code etc)
- User friendly interface
- Developed @iVEC, being installed @NCI
- See 2 Posters – YABI & RS-YABI

Tools to query large spatial time series (example, currently twinkleware)

These rely on standards-based ways to represent data, such as netCDF/CF conventions and the (emerging) raster storage architectures and the National nested grid

Credit: Leo Lymburner @ GA
Where have we got to?

According to NeCTAR, virtual laboratories are…

“Connecting researchers with existing and new research facilities, data repositories and computational tools, NeCTAR’s Virtual Laboratories will streamline research workflows and enable new opportunities for research innovation. The Virtual Laboratories will enable more collaborative research and new research efficiencies on a national scale by connecting researchers with each other and with research infrastructure.”

This doesn’t look too different from what we are building.

Key Points

- Enabled by NCRIS/EIF support across multiple research sectors and evolution of research community attitudes
  - IMOS + TERN + P/C + NeAT
- No specific funding for this initiative but now attracting support through projects - we need a plan!
- Enable trans-domain science (ESS) by bringing together data sets not normally used by only one community
- Share with partners: data assembly - data production - software maintenance
- Easier to collaborate
- Consistency and standards
- VIIRS – launched 11 days ago – will provide a good test to see if we can do a better job nationally than we did with MODIS
- Reduce duplication
- Reduce duplication