Eco-informatics challenges 2017-2027

Tim Clancy

AeRO Thriving in the data-driven research world

May 4, 2017

Task

- Research challenges → 2-10 years
- eResearch drivers/needs
- Capabilities to thrive
- Participation in an open, shared, international data world

Global Trends/Grand Challenges

Anthropocene Era and Ecosystem science

- 1) Human systems and Sustainability
- 2) Natural systems and impacts

Current Approach

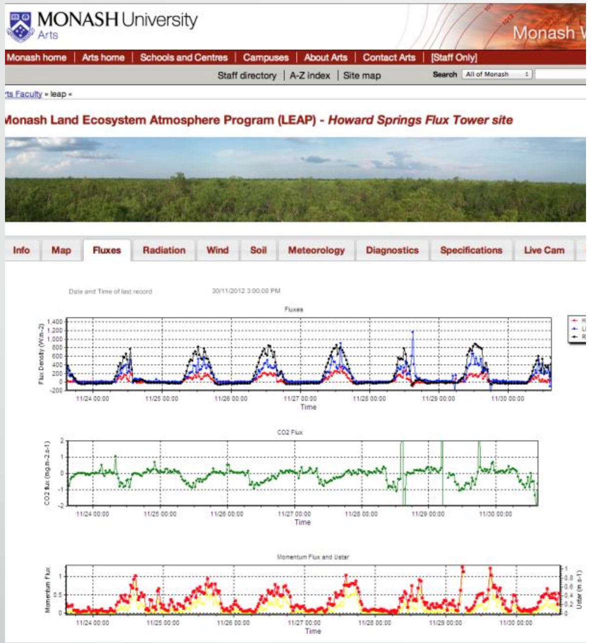
- Climate Science > Marine >>Terrestrial

What’s Needed:

- *Geniune Global, Collaborative, Integrated, Open, Science Driven Approach*
- *With Quality Data at its core*

Ecology and the data deluge

Sensor Technologies




Trends in Ecology & Evolution

Cell PRESS

Volume 27, Issue 2, February 2012, Pages 121–129

Ecological and evolutionary informatics

Review

Staying afloat in the sensor data deluge

John H. Porter[1], Paul C. Hanson[2], Chau-Chin Lin[3]

[1] Department of Environmental Sciences, University of Virginia, 291 McCormick Road, Charlottesville, VA 22904-4123, USA

[2] Center for Limnology, University of Wisconsin, Madison, WI 53706, USA

[3] Taiwan Forestry Research Institute, 53 Nan-Hai Road, Taipei, Taiwan

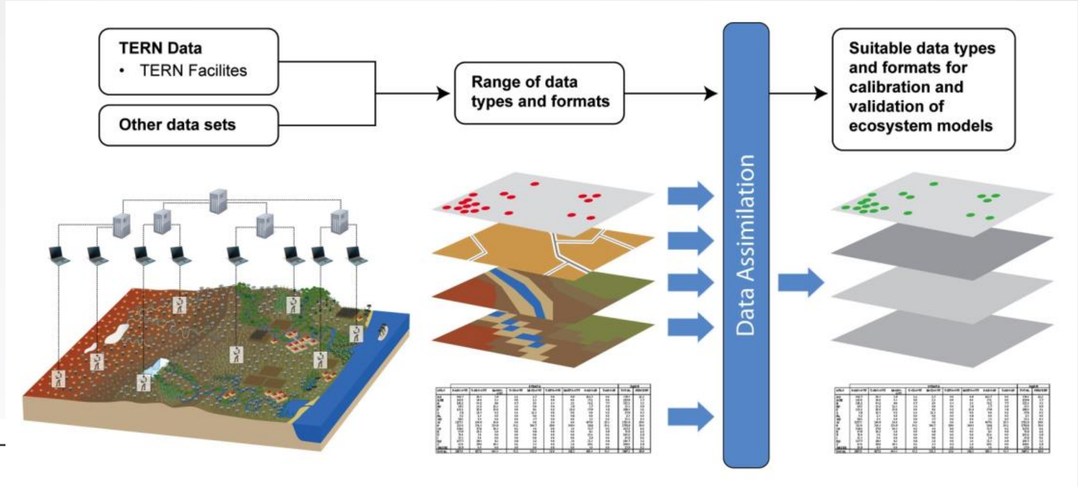
http://dx.doi.org/10.1016/j.tree.2011.11.009, How to Cite or Link Using DOI

Permissions & Reprints

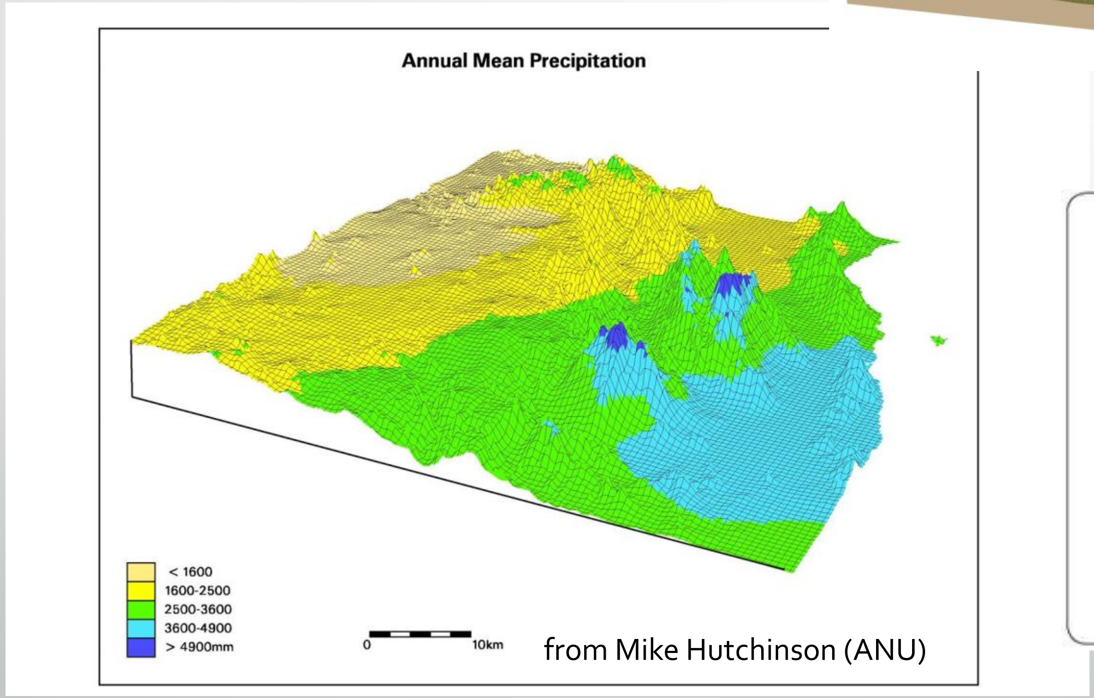
Developments in sensor design, electronics, computer technology and networking have converged to provide new ways of collecting environmental data at rates hitherto impossible to achieve. To translate this 'data deluge' into scientific knowledge requires comparable advances in our ability to integrate, process and analyze massive data sets. We review the experience of one large project in ingesting and analyzing sensor data from global lakes and provide a synopsis of innovative approaches being used to confront the information management and analytical challenges posed by massive volumes of data.

Ecology and the data deluge

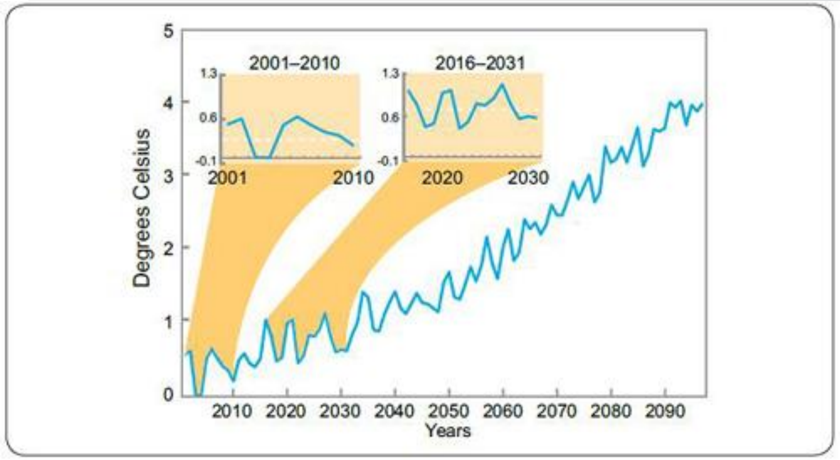
Modelled Data



High-resolution climate surfaces

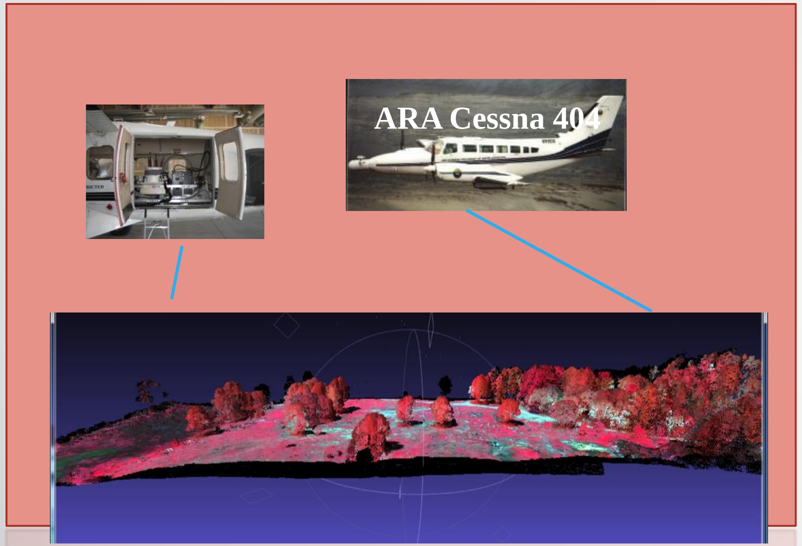


from Mike Hutchinson (ANU)



Ecology and the data deluge

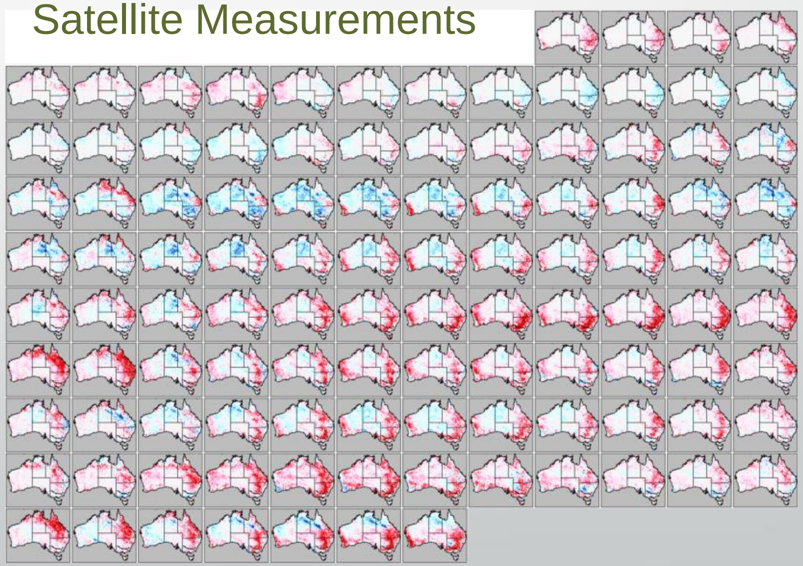
Assembly of Multi-Scale Remote Sensing Datasets



Remote Sensing /High Res Mapping

Airborne Systems and new sensor technologies

Low-spatial resolution, high temporal resolution
Satellite Measurements



Continental Dynamics in Green Cover

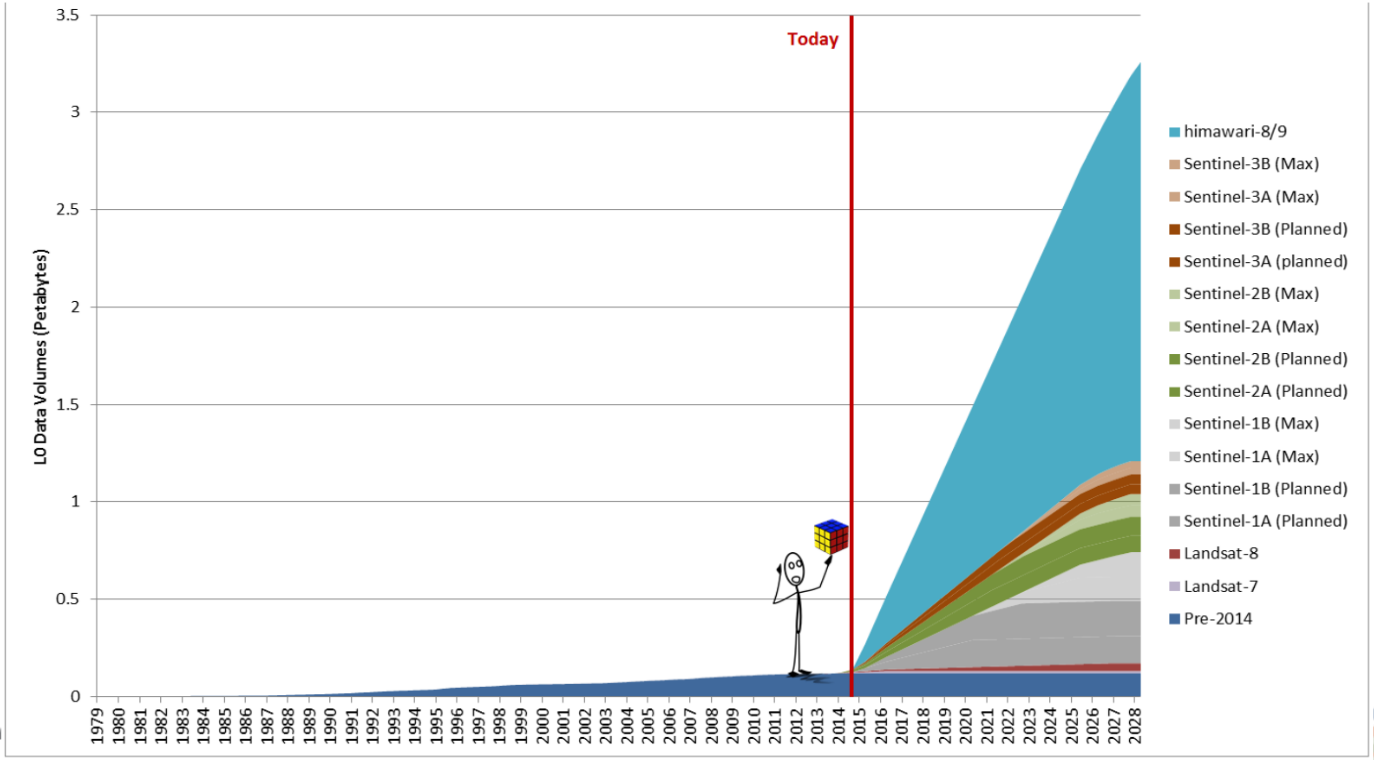
High-Resolution, Field-based Measurements

Data Management Challenges

Emerging EO informatics

Estimated storage and processing needs and opportunities for the next decade

- Increasing volume: spatial, temporal and spectral resolution and number of sensors
- Increasing velocity: near real time applications and automation
- Increasing variety: best practice processing and integration of disparate sources



Data Facility

Ecosystem science

- Inter-relationship among the living organisms, physical features, bio-chemical processes, natural phenomena, and human activities in ecological communtiies[1]

- Focusing on Terrestrial Ecosystem
  - Terrestrial Ecosystem Research Network
  - Atlas of Living Australia

- Data is heterogeneous: wide variety from different domain
  - Observation (human, in-situ sensors and satellite remote sensing)
  - Variety of scale: spatial and temporal
  - Different data formats used in the community

Ecosystem research Informatics

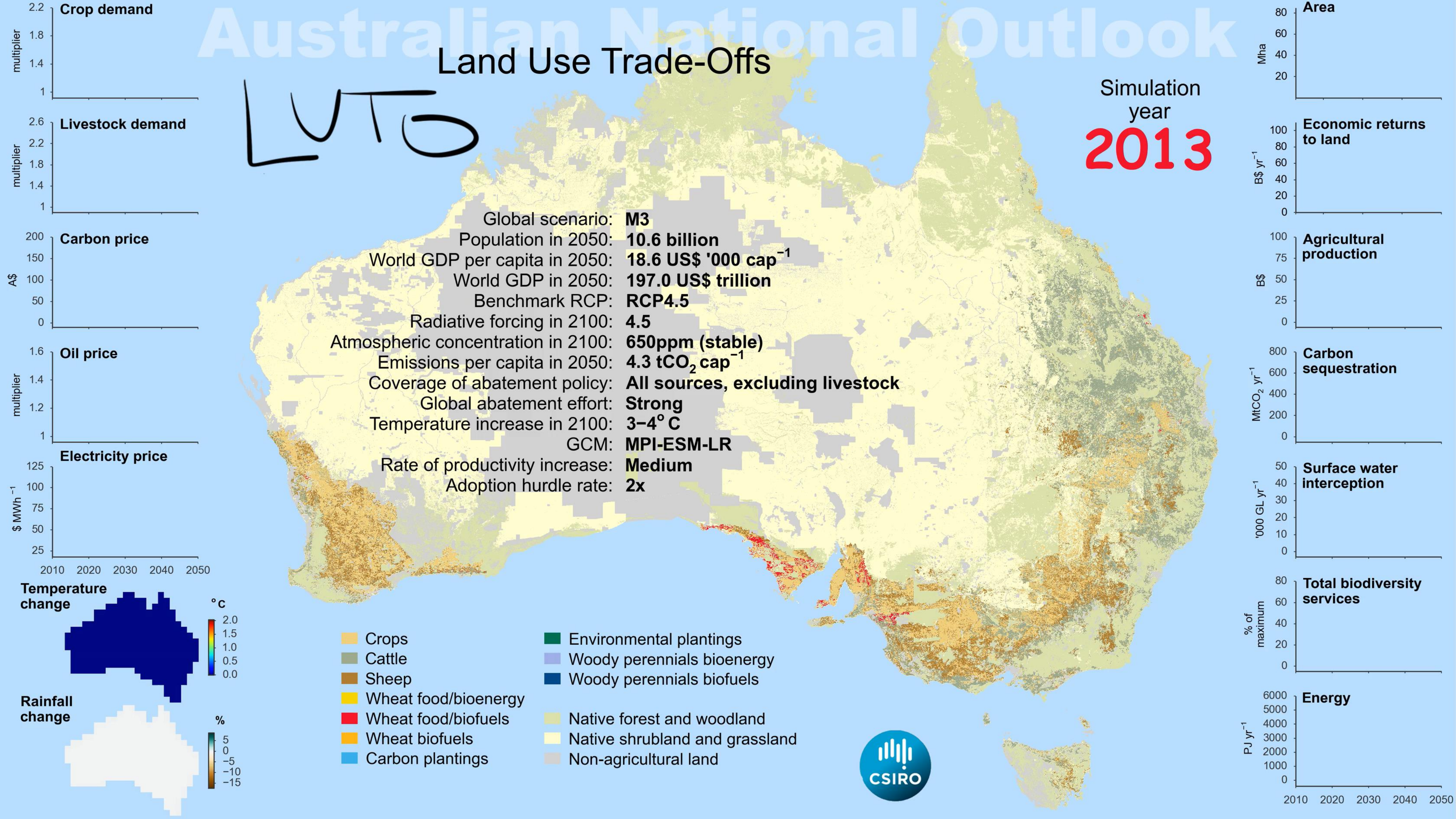
1. Small (but connected) Lab
2. Large Data and integration
3. Mega – Data and Compute

1. Small Lab – Access and Publication

- Data heterogeneity: wide variety from different domain
- Variety of scale: spatial and temporal
- Data formats
- Metadata standards and quality control
- Common data exchange format
- Aligned domestic and international data delivery platforms
- Accessibility, attribution and usage statistics

2. Large Data and Integration

- E.g. Bryan CSIRO



Australian National Outlook
Land Use Trade-Offs
LUTO
Simulation year
2013
Global scenario: M3
Population in 2050: 10.6 billion
World GDP per capita in 2050: 18.6 US$ '000 cap^{-1}
World GDP in 2050: 197.0 US$ trillion
Benchmark RCP: RCP4.5
Radiative forcing in 2100: 4.5
Atmospheric concentration in 2100: 650ppm (stable)
Emissions per capita in 2050: 4.3 tCO_2 cap^{-1}
Coverage of abatement policy: All sources, excluding livestock
Global abatement effort: Strong
Temperature increase in 2100: 3−4° C
GCM: MPI-ESM-LR
Rate of productivity increase: Medium
Adoption hurdle rate: 2x
Crop demand
multiplier
2.2
1.8
1.4
1
Livestock demand
multiplier
2.6
2.2
1.8
1.4
1
Carbon price
A$
200
150
100
50
0
Oil price
multiplier
1.6
1.4
1.2
1
Electricity price
$ MWh^{-1}
125
100
75
50
25
2010 2020 2030 2040 2050
Temperature change
°C
2.0
1.5
1.0
0.5
0.0
Rainfall change
%
5
0
−5
−10
−15
Crops
Cattle
Sheep
Wheat food/bioenergy
Wheat food/biofuels
Wheat biofuels
Carbon plantings
Environmental plantings
Woody perennials bioenergy
Woody perennials biofuels
Native forest and woodland
Native shrubland and grassland
Non-agricultural land
CSIRO
Area
Mha
80
60
40
20
Economic returns to land
B$ yr^{-1}
100
80
60
40
20
0
Agricultural production
B$
100
75
50
25
0
Carbon sequestration
$MtCO_2$ yr^{-1}
800
600
400
200
0
Surface water interception
'000 GL yr^{-1}
50
40
30
20
10
0
Total biodiversity services
% of maximum
80
60
40
20
0
Energy
PJ yr^{-1}
6000
5000
4000
3000
2000
1000
0
2010 2020 2030 2040 2050

Transformational improvements are possible

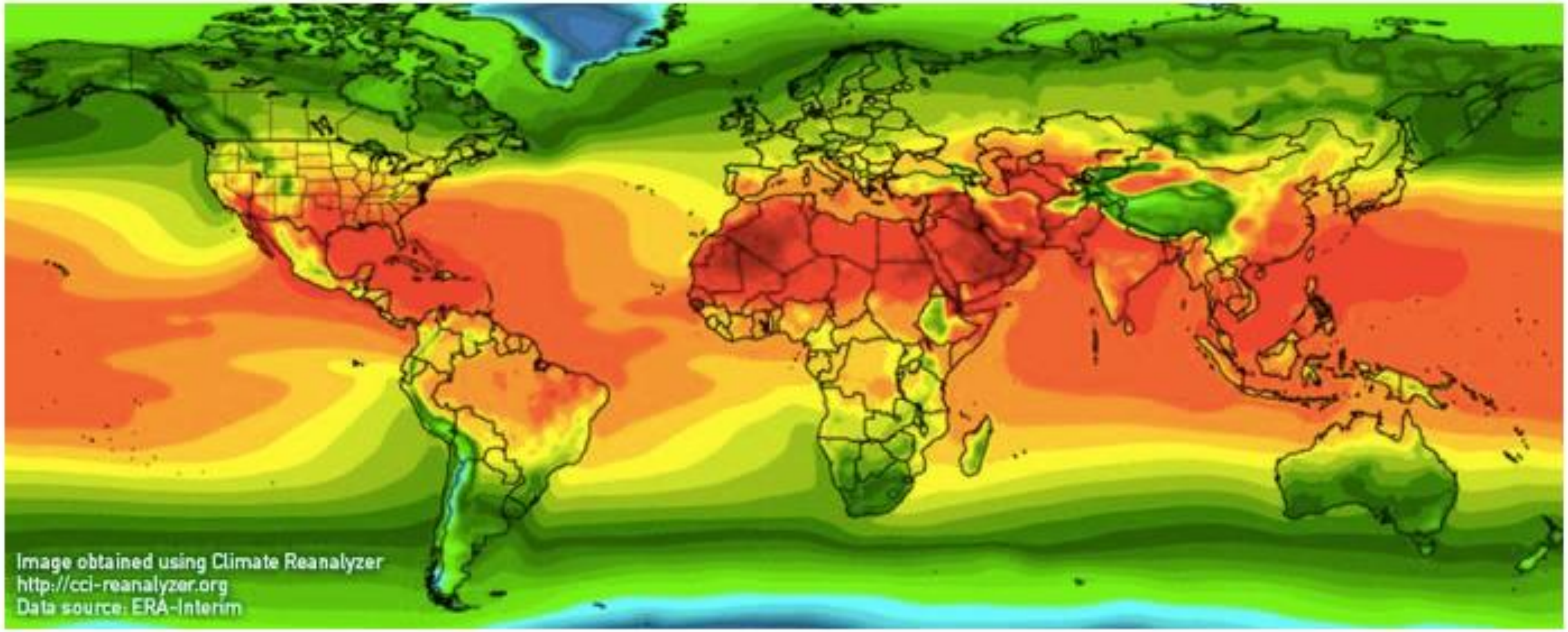
- GIS processing = 109 days or 15.5 weeks
- Python and Numpy offer a substantial advantage
- Further improvements in GPU, ElementwiseKernels, parallelisation

Hardware	Software	Time	Iterations	Effective time for 1,000 iterations	Speedup from AML
1 CPU core	ESRI Arc Macro Language	94,103	10	9,410,300	
1 CPU core	Ipython, Numpy	16,042	100	160,425	59
1 GPU	Ipython, Numpy, PyCUDA	6,388	1,000	6,388	1,473
1 GPU	Ipython, Numpy, PyCUDA elementwise kernel	1,921	1,000	1,921	4,898
256 CPU cores	Ipython, Numpy	187	100	1,865	5,046
64 GPUs	Ipython, Numpy, PyCUDA	293	1,000	293	30,553
64 GPUs	Ipython, Numpy, PyCUDA elementwise kernel	148	1,000	148	63,643

Bryan, B.A. (2013). High-performance computing tools for the integrated assessment and modelling of social-ecological systems. *Environmental Modeling & Software* 39, pp.295 – 303.

3. Mega

HOW NCI'S HIGH-PERFORMANCE DATA IS DRIVING INNOVATIVE CLIMATE CHANGE RESEARCH



After sweltering through Australia's warmest year to date, the National Computational Infrastructure's petascale contribution to Earth System Science has never been more critical.

Trends

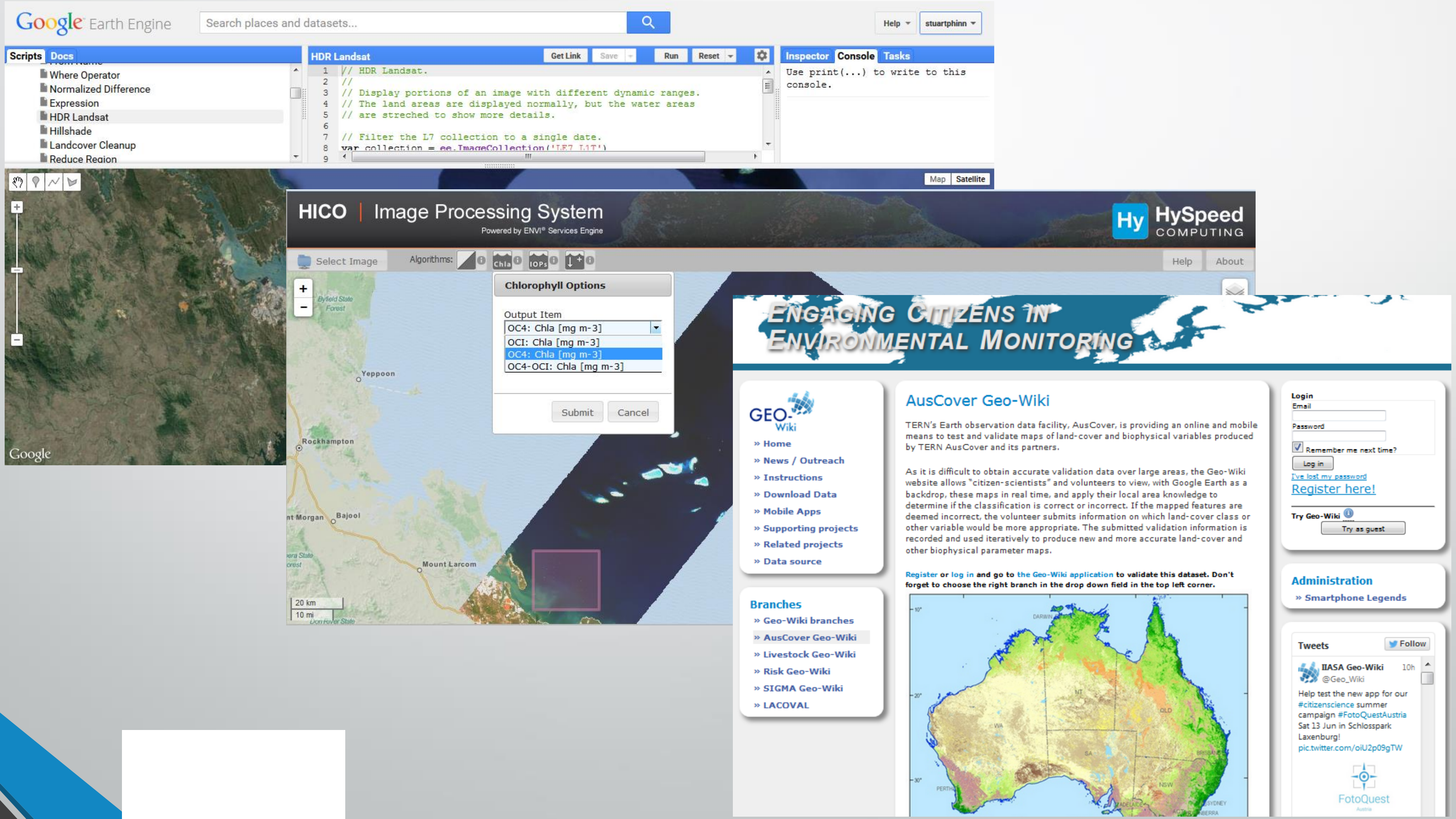
1. More 1 becoming 2.
2. More 2 becoming 3.

Needs

1. Data, data and data
2. Skills
3. Methods (Data and Tools) environments
   - Integration

Moving forwards – sustaining long term science

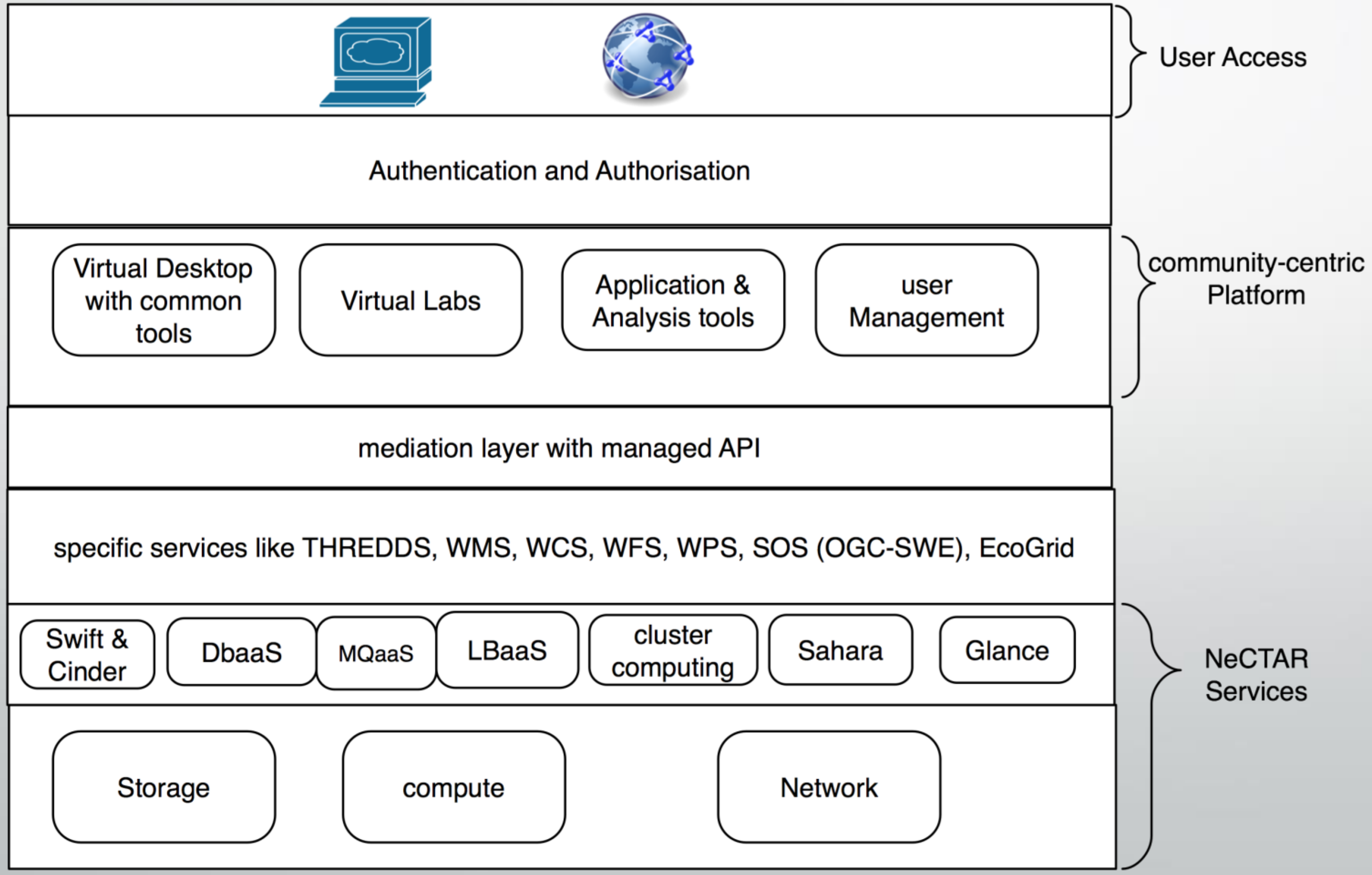
- Global shift to collaborative data , algorithms and participatory resources:



To thrive

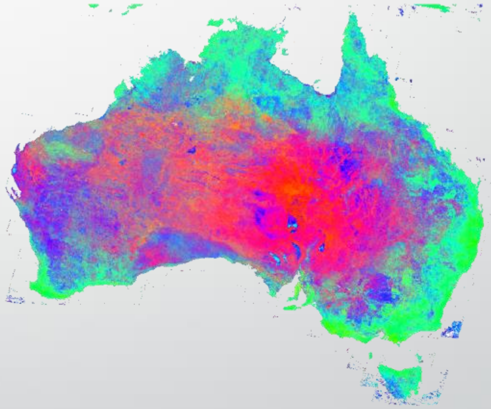
- Learn from other disciplines
- Build on what’s working, ALA, BCCVL, TERN, IMOS etc
- Forward Journey(s) not a fixed destination
- Support skills sharing
- Integration and focus -> Eco-Cloud

High-level conceptual Architecture



Current status

- Setup a Technical Advisory Group advice on the scoping and implementation of the project.
- In the first iteration: reference datasets will be made available
  - Remote sensing reference data (fractional Cover)
  - Long-term ecological monitoring data
  - Climate variables
- Scoping the mediation layer and overall architecture
- Building a coalition of willing for partnership and collaboration



Contributions

- NeCTAR – Major project sponsor
- TERN, ALA – NCRIS Domain Projects, partners
- QCIF - implementation partner
- NCI – collaborator, partners

Thank you

Contact:
t.clancy@uq.edu.au