Earth and environmental science information partners
- an Australian community of practice in earth and environmental science informatics

Simon Cox - CSIRO Land and Water
Erin Robinson - ESIP
Lesley Wyborn - ANU, ANDS
The fabric of science is changing, driven by a revolution in digital technologies that facilitate the acquisition and communication of massive amounts of data. This is changing the nature of collaboration and expanding opportunities to participate in science. If digital technologies are the engine of this revolution, digital data are its fuel. But for many scientific disciplines, this fuel is in short supply.

Kansas Agricultural Smoke, April 12, 2003

Ag Fires

Fire Pixels

Organics

35 ug/m³ max

SeaWiFS, Refl

AOT Blue
<table>
<thead>
<tr>
<th>Use</th>
<th>Data Access</th>
<th>Understanding</th>
<th>Discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Data Access</td>
<td>Understanding</td>
<td>Discovery</td>
</tr>
<tr>
<td>Use</td>
<td>Data Access</td>
<td>Understanding</td>
<td>Discovery</td>
</tr>
<tr>
<td>Quality</td>
<td>Data Access</td>
<td>Understanding</td>
<td>Discovery</td>
</tr>
<tr>
<td>Use</td>
<td>Data Access</td>
<td>Understanding</td>
<td>Discovery</td>
</tr>
<tr>
<td>Quality</td>
<td>Data Access</td>
<td>Understanding</td>
<td>Discovery</td>
</tr>
<tr>
<td>Use</td>
<td>Data Access</td>
<td>Understanding</td>
<td>Discovery</td>
</tr>
<tr>
<td>Quality</td>
<td>Data Access</td>
<td>Understanding</td>
<td>Discovery</td>
</tr>
</tbody>
</table>

[https://upload.wikimedia.org/wikipedia/commons/6/6c/Hurdles_%28Scenes_from_a_Track_Meet%29.jpg](https://upload.wikimedia.org/wikipedia/commons/6/6c/Hurdles_%28Scenes_from_a_Track_Meet%29.jpg)
The Old Way:

Pre-Science:
- Find data
- Retrieve high volume data
- Learn formats and develop readers
- Extract parameters
- Perform spatial and other subsetting
- Identify quality and other flags and constraints
- Perform filtering/masking
- Develop analysis and visualization
- Accept/discard/get more data (sat, model, ground-based)

DO SCIENCE:
- Exploration
- Initial Analysis
- Use the best data for the final analysis
- Derive conclusions
- Write the paper
- Present @ AGU

Jan
Mar
Jun
Sept
Dec

Adapted from Leptoukh, 2012
“The provider cannot document the data for discovery; If she can document it, she cannot share it; If she can share it, she doesn't know how to explain the quality; If she finds them good, she does not know how merge them with other data”

_The Users View of IT, NAS 1989_
Pre-Science

- Find data
- Retrieve high volume data
- Learn formats and develop readers
- Extract parameters
- Perform spatial and other subsetting
- Identify quality and other flags and constraints
- Perform filtering/masking
- Develop analysis and visualization
- Accept/discard/get more data (sat, model, ground-based)

DO SCIENCE

- Exploration
- Initial Analysis
- Use the best data for the final analysis
- Derive conclusions
- Write the paper
- Present @ AGU

More time to DO

Minutes
Days for exploration
Use the best data for the final analysis
Derive conclusions
Write the paper
Submit the paper

Adapted from Leptoukh, 2012
A multi (inter)-disciplinary approach to addressing a range of inter-related concerns to enable data sharing and reuse:

- **Social** – attitudes, identity, behaviour, privacy, ethics, social licence
- **Institutional** – legislation, policy, governance, KPIs, incentives
- **Economics** – value of data & business models

Supplements and informs Technical and Information Architectures
ESIP Community

Multiple Science Domains

data providers

data archives

users

application developers

tool & system developers

Academia

Corporate

Cross-Agency
Information Interoperability Stack

Shared Knowledge
Joint Decisions

Collaborative Infrastructure

Social Software
Connecting Humans

Data Sharing
Data Standards, Service
Oriented Architecture

Hardware

Software

TCP/IP
Neutral Convener

been enough to spook the whole herd," said Curt.

McClellan added that "it helps to have a neutral convener—not a stakeholder with a particular interest—but one that could be trusted by the broad base of stakeholders we think help make these efforts a success." Such a neutral convener can "create the safe, legal harbor needed for collaborations to occur," he said. Woosley concurred, and said "that honest broker component to precompetitive sharing is an essential element because if you are asking people to share, but you are in it for yourself, or you are trying to make money or want to hold the IP, it is just not going to work."
Three Approaches

In-Person Workshops | Virtual Collaboration Space | Virtual Lab Space
STRUCTURE

**Standing Committees**
- Data Stewardship
- Education
- Information Technology and Interoperability
- Semantic Technology

**Administrative Committees**
- Constitution and Bylaws
- Finance and Appropriations
- Nominations
- Partnership

**Clusters**
- Agriculture & Climate
- **CLEAN – Climate Literacy**
- Cloud Computing
- Data Management Training
- Disaster Lifecycle
- Discovery
- Documentation
- Drones
- Earth Science Data Analytics
- Energy & Climate
- Envirosensing
- Info Quality
- Science Software Citation
- Sustainable Data Management
- Data Models
- IM Code Registry
- Web Services
- ... and yours?

---

http://esipfed.org/get-involved/collaborate
How to Start an ESIP Cluster... they’re like workshops, only instant and free.

**Step 1:** Have an idea? Ping ESIP community (via mailing list or Slack). Get feedback.

**Step 2:** Contact ESIP Vice President with Cluster name - get approved.

**Step 3:** Get access to ESIP resources [Slack, Listserv, wiki, GoTo Meeting, GitHub, AWS].

**Step 4:** Schedule first telecon and get to work!
ESIP Lab Mission

The ESIP Lab empowers the scientific community by supporting ideation, incubation and evaluation of Earth sciences cyberinfrastructure.
ESIP Backbone

20 years - Core funding through NASA, NOAA, USGS
ESIP Backbone
2017 Cutting edge science symposium - Linking environmental data & samples

- Canberra, May 2017 - 3 days
- ~70 participants, 18 international, ESIP co-chair
- Geoscience, marine, ecosystems, biodiversity, climate
- Field trip, presentations, un-conference (breakouts), anti-conference, up-goer challenge
Pre-conference workshop - [http://www.c3dis.com/1725](http://www.c3dis.com/1725) - 09:00–18:00
- ~40 in the rooms (6 international), ~5 remote (2 international)
- ESIP on organizing committee
- 6 NCRIS facilities + CSIRO + GA + CRCSI + 2 state govts + ...
E2SIP @ C3DIS - 6 breakouts

- Vocabularies & vocabulary services
- FAIR publishing
- FAIR services & software
- Big data/analytics
- Cloud platform for X-domain science
- Drones
Do the participants want more of this?

Yes!
Institutional context

- NASA’s initial requirements
- Ongoing support from NASA, NOAA, USGS, Foundations
  - 20 years history
- Scale
- AGU
- NASA, NOAA, USGS grants to researchers tied to ESIP participation

- NCRIS facilities
  - ALA, AuScope, IMOS, TERN, NCI
- Cross-disciplinary institutions
  - CSIRO, ANDS
    - Neutral players?
- Govt participation, not grants
  - DoEE, GA, BoM
- Smaller community
  - We all know each other?
More context/exemplars?
Summary

- Coordinated information infrastructure technologies and practices creates more time for science
- ESIP community is enabling this in US, variety of activities
- Two successful workshops → the Australian E2 community wants a similar platform
Next steps

- Semi-annual workshop - alongside C3DIS & eResearch
  - include labs/hackathons with concrete goals
- E2SIP as an ESIP Cluster
  - Leverage ESIP collaboration tools and logistics support
- E2 community get involved, or re-shape, ANDS activities in vocabs etc.
- ANDS-v2 to provide local coordination support?
Thank you

Simon Cox - simon.cox@csiro.au
Erin Robinson - erinrobinson@esipfed.org
Lesley Wyborn - lesley.Wyborn@anu.edu.au
Data Stewardship - Wrote a white paper as a response to the grass-roots "data rescue" events to provide data facility professionals' perspectives on those activities.

Next Tech Dive - 10 May 2018: "NetCDF-CF Simple Geometries": Dave Blodgett (USGS) and Tim Whiteaker (UT Austin)

Summary: Simple geometry (points, lines, and polygons) has now been accepted as part of the Open Geospatial Consortium’s NetCDF-CF specification. This a major enhancement to a widely used standard whose utility has previously been limited to time-series of point or (raster) coverage data only.
Features

- Chair elected by Assembly
- Chair serves on Executive Committee
- Budget*

Standing Committees
- Data Stewardship
- Education
- Information Technology and Interoperability
- Semantic Technology

Administrative Committees
- Constitution and Bylaws
- Finance and Appropriations
- Nominations
- Partnership

Formality

Active Participation
Opportunities

Committees

Least
Most
COLLABORATION AREA

STRUCTURE

- Agriculture & Climate
- CLEAN – Climate Literacy
- Cloud Computing
- Data Management Training
- Disaster Lifecycle
- Discovery
- Documentation

- Drones
- Earth Science Data Analytics
- Energy & Climate
- Envirosensing
- Info Quality
- Science Software Citation
- Sustainable Data Management
- Data Models
- IM Code Registry
- Web Services
- ... And yours?

Active Participation
Opportunities

Formality

Least

Most

Cluster

Step 1: Have an idea? Ping ESIP community (via mailing list or Slack). Get feedback.

Step 2: Contact ESIP Vice President with Cluster name - get approved.

Step 3: Get access to ESIP resources [Slack, List serv, wiki, GoTo Meeting, GitHub, AWS].

Step 4: Schedule first telecon and get to work!