Cultivating an Emergent Earth Observation Analytics Ecosystem in the Cloud

Christopher Lynnes*, NASA/Goddard Space Flight Center

System Architect
Earth Observing System Data and Information System

*U.S. Civil Servant
We are in a golden age of Earth Observations

- Dozens of Earth Observation Satellites
- Many agencies with Open Data policies
- Tens of thousands of datasets
Data analysis tools and services abound

- Jupyter
- RStudio
- R
- ArcGIS
- Python
- NumPy
- matplotlib
- scikit-learn
- pandas
- RNotebooks
- qgis
- GRASS
- SciPy
- xarray
- ncdf4
- matlab
- Statsmodels
- IDV
- IDL
- e1071
- GDAL
- SeaDAS
- caret
- SeaDAS
- gdal2gos
- SNAP Toolbox
- GMT
- nco
- Iris
- cartopy
- ggplot2
How Do We Make Full Use of Everything?

● Dependency on research problem

● Evolution of key components
  ○ instruments
  ○ analysis methods

● Tool and service providers in different
  ○ Countries
  ○ Agencies
  ○ Companies
Answer:

It’s Complicated...
Answer:

...or is it Complex?
Complex Adaptive Systems (CAS)

1. Multiple components
2. Varying degrees of heterogeneity
3. Inter-component interactions: signals and rules
4. No central authority
5. Emergent behavior: whole > sum of parts
## Complicated vs. Complex Adaptive Systems

<table>
<thead>
<tr>
<th>Complicated System</th>
<th>Complex Adaptive System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss of one individual component -&gt; system impairment</td>
<td>Loss of one individual -&gt; system adaptation (usually)</td>
</tr>
<tr>
<td>Completely deterministic</td>
<td>Unexpected behavior arising from ensemble interactions</td>
</tr>
<tr>
<td>Example: internal combustion engine</td>
<td>Examples: ecosystems, economies, ant colonies, genetic algorithms...</td>
</tr>
</tbody>
</table>
Earth Observation Analytics as CAS

- Multiple components: datasets, analysis software
- Varying degrees of heterogeneity
- Inter-component interactions:
  - Application Program Interfaces
  - Services
- No central authority
- Emergent behavior?
Can Complexity Theory Help?

1. Derive lessons from “similar” Complex Adaptive Systems
   a. The Internet
   b. The World Wide Web

2. Analogies from work in Complex Adaptive Systems
   a. Bucket Brigade
   b. Recombination
The Internet

Killer App

Email

SMTP | POP | IMAP

TCP/IP

Application Support Standards

Simple Foundational Standards

SMTP: Simple Mail Transfer Protocol (1982)
## The World Wide Web

<table>
<thead>
<tr>
<th>Killer App</th>
<th>Web Browser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application Support Standard</td>
<td>Web Server</td>
</tr>
<tr>
<td>Simple Foundational Standard</td>
<td>HTML</td>
</tr>
<tr>
<td></td>
<td>HTTP</td>
</tr>
</tbody>
</table>

**HTML:** Hypertext Markup Language (1991)  
**HTTP:** Hypertext Transfer Protocol (1991)
Earth Observation Analytics

Killer App

Machine Learning?

WCPS? WPS?

OPeNDAP? OpenAPI?

Application Support Standards

Simple Foundational Standards

WCPS: Web Coverage Processing Service
WPS: Web Processing Service
OPeNDAP: Open-source Project for a Network Data Access Protocol
OpenAPI: Open Application Program Interface
Analogies from Complex Adaptive Systems
Bucket Brigade Algorithm and Provenance

“Bucket Brigade” Credit Assignment to brokers

New role for provenance?

data source → transform → analyze → publication

provenance citation consumer
**Adaptation by Recombination (Genetic Algorithms)**

*Combine syntactic and semantic standards?*

<table>
<thead>
<tr>
<th>Syntactic</th>
<th>Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenSearch</td>
<td>schema.org</td>
</tr>
<tr>
<td>OpenAPI</td>
<td>DCAT</td>
</tr>
<tr>
<td>WCS</td>
<td>SWEET</td>
</tr>
<tr>
<td></td>
<td>ISO 19115</td>
</tr>
</tbody>
</table>

OpenAPI: Open Application Program Interface  
WCS:  Web Coverage Service  
DCAT: Data Catalog Vocabulary  
SWEET: Semantic Web for Earth and Environmental Terminology
Going Forward...

1. Embrace Earth Observation Analytics as a Complex Adaptive System
2. Leverage combinations of simple protocols and standards
3. Ensure proper credit to participating agents
4. (Keep an eye out for dysfunctional emergence)
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Program Interface</td>
</tr>
<tr>
<td>caret</td>
<td>Classification and regression training</td>
</tr>
<tr>
<td>CAS</td>
<td>Complex Adaptive System</td>
</tr>
<tr>
<td>DCAT</td>
<td>Data Catalog Vocabulary</td>
</tr>
<tr>
<td>GDAL</td>
<td>Geospatial Data Abstraction Library</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information System</td>
</tr>
<tr>
<td>GMT</td>
<td>Generic Mapping Tools</td>
</tr>
<tr>
<td>GRASS</td>
<td>Geographic Resources Analysis Support System</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>IDL</td>
<td>Interactive Data Language</td>
</tr>
<tr>
<td>IDV</td>
<td>Integrated Data Viewer</td>
</tr>
<tr>
<td>IMAP</td>
<td>Internet Message Access Protocol</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>ncdf</td>
<td>netCDF</td>
</tr>
<tr>
<td>nco</td>
<td>netCDF Command Operators</td>
</tr>
<tr>
<td>netCDF</td>
<td>network Common Data Form</td>
</tr>
<tr>
<td>OPeNDAP</td>
<td>Open-source Project for a Network Data Access Protocol</td>
</tr>
<tr>
<td>pandas</td>
<td>Python Data Analysis Library</td>
</tr>
<tr>
<td>POP</td>
<td>Post Office Protocol</td>
</tr>
<tr>
<td>SeaDAS</td>
<td>SeaWiFS Data Analysis System</td>
</tr>
<tr>
<td>SeaWiFS</td>
<td>Sea-viewing Wide Field-of-View Sensor</td>
</tr>
<tr>
<td>SMTP</td>
<td>Simple Mail Transfer Protocol</td>
</tr>
<tr>
<td>SNAP</td>
<td>Sentinel Application Platform</td>
</tr>
<tr>
<td>SWEET</td>
<td>Semantic Web for Earth and Environmental Terminology</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>WCPS</td>
<td>Web Coverage Processing Service</td>
</tr>
<tr>
<td>WCS</td>
<td>Web Coverage Service</td>
</tr>
<tr>
<td>WPS</td>
<td>Web Processing Service</td>
</tr>
</tbody>
</table>