EARTH CUBE CONCEPTUAL DESIGN
A Scalable Community Driven Architecture

http://earthcube.org/group/scalable-community-driven-architecture

Overview

PI: G. Djorgovski (Caltech)

CO-I: D. Pilone, T. Pilone (Element 84), D. Crichton, E. Law (JPL)

Other key personnel: S. Caltagirone (E84), S. Hughes (JPL),
T. Huang (JPL), A. Mahabal (Caltech)

2016 All Hands Meeting
A high level system blueprint for the definition, construction, and deployment of both existing and new components to ensure that they can be unified and integrated into an evolutionary national infrastructure for EarthCube
Methodology

- Identification of stakeholders, concerns and requirements
- Identification of architectural use cases and drivers
- Selection of an architectural framework
- Development of the architectural principles
- Development of the architectural models
- Capture of the architecture artifacts in a consolidated report
- Generation of recommendations for adopting the architecture for the EarthCube program
# Stakeholders

<table>
<thead>
<tr>
<th>Stakeholder/Actor</th>
<th>Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSF Program Managers</td>
<td>Make decision and provide guidance at the EarthCube program level. Provide sufficient funding to support the EarthCube mission.</td>
</tr>
<tr>
<td>EarthCube Scientists</td>
<td>Use EarthCube resources and services to conduct scientific research. Publish scientific results &amp; curate data as needed.</td>
</tr>
<tr>
<td>EarthCube Developers</td>
<td>Develop technologies and services that can be integrated into EarthCube.</td>
</tr>
<tr>
<td>External Data Users</td>
<td>Use EarthCube resources and services for research, education, and decision-making.</td>
</tr>
<tr>
<td>Curator</td>
<td>Ensure data is properly captured in EarthCube compliant data repositories.</td>
</tr>
<tr>
<td>Data Owner</td>
<td>Responsible for producing the data. Concerned about its distribution and use.</td>
</tr>
<tr>
<td>External Data Facility</td>
<td>Responsible for archiving data at other agencies (NASA, NOAA, USGS, etc); interoperability with the EarthCube Cyberinfrastructure.</td>
</tr>
<tr>
<td>EarthCube Governance Committees</td>
<td>Responsible for generating and monitoring the governance for the system including data curation, access, use case priority, interoperability standards, etc.</td>
</tr>
<tr>
<td>EarthCube Office Staff</td>
<td>Responsible for maintaining the community involvement within EarthCube and communicating changes and how to use the system.</td>
</tr>
</tbody>
</table>
Use Cases

- Big Science – Discovery, Comparison, Provenance, Model & visualization
- Collaborative Science
- Dark Data Contribution
- Tools Contribution
- Data Documentation
- Models Sharing
- High Performance Computing and Storage Resources
- Real Time Data
- Physical Sample Curation
Drivers

- Transform and accelerate research and discovery by turning data into knowledge and enabling interdisciplinary data integration.

- Provide critically needed data, tools, and computational resources and frameworks for cross-domain scientific collaboration, analysis and with long-term geoscience software and data preservation, discovery and use.

- Provide a geosciences cyberinfrastructure and architecture that is scalable, extensible and sustainable.
EarthCube CI

Science Teams

EarthCube Repository

EarthCube Data Analytics Centers

Data Science Infrastructure (Data, Algorithms, Machines)

Other Data Systems (In-Situ, University)

Other Data Systems (e.g. NOAA)

Data Provider

Research

Applications

Decision Support

Data Analysis
Status

- Participation in TAC Architecture WG
- Participation in Architecture & Assessment Workshops
- Conceptual Design Review in progress
- Community engagement
  - Presentations at various venue (AGU, ESIP, EC meetings, demo etc )
  - Concept review
- Next Step
  - Finalize Design Document and Recommendations
Stepping back…
EarthCube is an institutional socio-technical program of NSF, as well as a technology infrastructure supported by the program and its broader community of stakeholders.

- 2016 Architecture Workshop Report Draft
Resource centric vs infrastructure: If EC does not build infrastructure to support science then the data lifecycle, who will?
Feedback

• Needs more focus on how data-driven, science architecture can help transform how science is being done
  • Emerging computing technologies
  • Innovative analytics that support interdisciplinary science (machine learning, statistical methodologies, visualization)

• More focus on information architecture (not just limited to metadata) to enable data/model interoperability across data repositories/registries
How do we make this more concrete?
How are we doing as a program?
Benchmark

- Earth System Grid Federation (ESGF)
- Early Detection Research Network (EDRN)
- NASA’s Earth Observing System Data and Information System (EOSDIS)
Recommended Next Steps in the next 3 months

2. Prioritize the identified Use Cases
3. Develop concrete success metrics for top X Use Cases
4. MEASURE current effectiveness of existing building blocks
5. Develop 6 month plan to address top X UCs and MEASURE AGAIN.
Thank you!