



New Approaches to Technology Development: How USGS and ESIP are working towards comprehensive Earth data provenance

Annie Burgess, Earth Science Information Partners



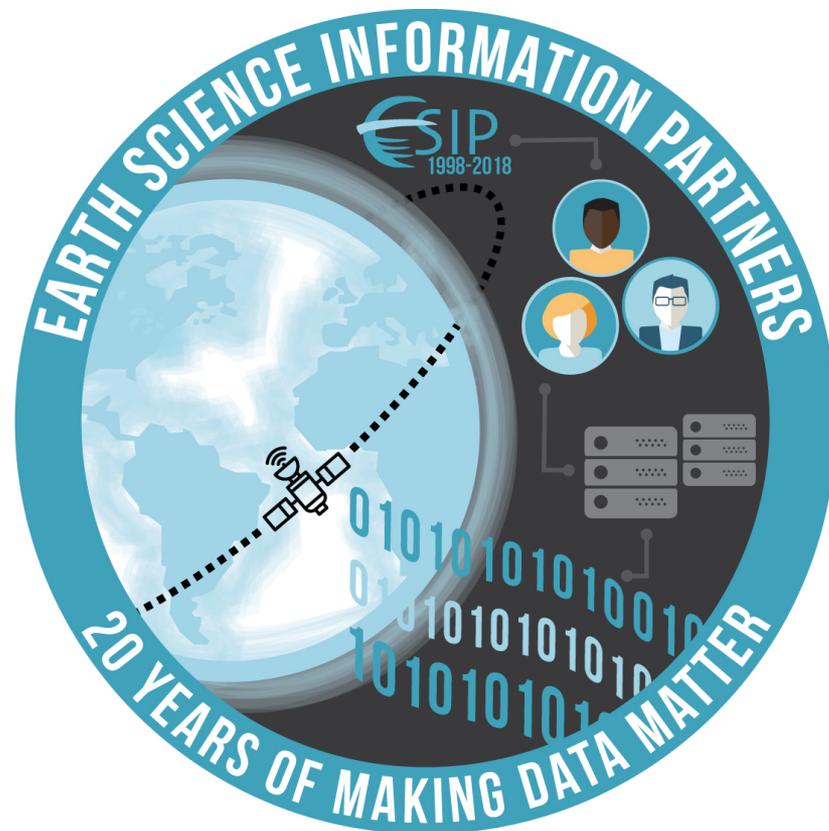
What is ESIP?

Earth Science Information Partners (ESIP)

ESIP brings people together who manage Earth science data to create and share best practices around making data FAIR.

1000 active participants across 110 member organizations that span the public, private and academic sectors.

Primarily funded by NASA, NOAA, and USGS.



What is the ESIP Lab?

The ESIP Lab supports the development of innovative, applied Earth science technologies.

What is the ESIP Lab?

The ESIP Lab supports the development of innovative, applied Earth science technologies.

With ESIP Lab support we:



"We transitioned a hydrologic model from 'research grade' to cloud-based operations for watersheds on three continents."



"We can now produce high-quality crop maps by creating deep learning workflows in web browsers."



"We are developing a sensor network to calibrate hydrology altimetry data for airborne and satellite applications."



"We can now incorporate deep learning in Hurricane modeling to save lives and reduce damages."

How do we support innovation?

- Small-grant funding
 - 10K over 6 months.
- Strategic outreach
 - Who *else* could benefit/contribute.
- Community input
 - Get early feedback from potential users.
- Organizational support
 - ESIP staff and resources.



How do we support innovation?

- Small-grant funding
 - 10K over 6 months.
- Strategic outreach
 - Who *else* could benefit/contribute.
- Community input
 - Get early feedback from potential users.
- Organizational support
 - ESIP staff and resources.



How do we support innovation?

- Small-grant funding
 - 10K over 6 months.
- Strategic outreach
 - Who *else* could benefit/contribute.
- Community input
 - Get early feedback from potential users.
- Organizational support
 - ESIP staff and resources.



How do we support innovation?

- Small-grant funding
 - 10K over 6 months.
- Strategic outreach
 - Who *else* could benefit/contribute.
- Community input
 - Get early feedback from potential users.
- Organizational support
 - ESIP staff and resources.



Why small grants and open innovation work.

Small grants are keenly suited to innovation.

- Culture of idea sharing and producing discrete deliverables that can be built upon quickly.

Community input accelerates innovation.

- Learn from the failures/successes of peers.

Sharing improves likelihood of sustainability.

- Solving common problems creates partnerships and continued support.



Experiment with USGS: Innovation and Provenance

How can USGS leverage the ESIP Lab to move the ball forward around Earth data provenance?



What is Provenance?

Provenance describes the origin and history of a dataset, from collection through any analysis or synthesis that resulted in its current state.

Experiment with USGS: Innovation and Provenance

Step 1: Gather Ideas

Step 2: Prototype an Idea

Step 3: Come Together

Experiment with USGS: Innovation and Provenance

Step 1: Gather Ideas

17 Ideas Posted

27 Comments

35 Votes

41 Individual Participants



How would YOU
improve how PROV
systems interoperate
across agencies or
institutions to enable a
more complete picture
of provenance?

Experiment with USGS: Innovation and Provenance

Step 1: Gather Ideas

Step 2: Prototype an Idea



PROVISIUM

PROVISIUM.IO

Develop tools that show the lifecycle of a dataset through web-based visualization. (Tom Narock)

github.com/ESIPFed/provisium

Experiment with USGS: Innovation and Provenance

Step 1: Gather Ideas

Step 2: Prototype an Idea

Step 3: Come Together



Goal 1: Design use of data provenance for targeted personas that reflect the provenance life-cycle phases.

Goal 2: Articulate a list of the barriers for implementation of provenance systems.

Goal 3: Determine a going forward strategy towards cross-agency and cross-persona interoperability.

Broad Outcomes

- USGS found 'great value' in the ideas, content, and connections generated through the three stages.
- Building the Provisium prototype was invaluable.
 - Identified roadblocks/insights to future-system functionality.
 - It is being reused. We will "revisit and refactor the code based on what we learned and apply to these real world driving use cases."
- A simplest-possible-use-case scenario of cross institutional provenance would be a valuable exercise.
- **Provenance is uniquely suited for a community approach.**

Provenance is uniquely suited for a community-approach

Is provenance for all Earth data a technology-solvable problem?

“The main pain point is NOT technology. The tougher challenge is how to coordinate.”

“The technology exists, but it will take buy in from the practitioners.”

“This is a problem like all the other data sharing challenges. It’s easier to convince people that it’s a good idea than to do it.”

“The software is easier than the social/cultural part.”



Thank you.

- [ESIPFED.ORG/LAB](https://esipfed.org/lab)
- [GITHUB.COM/ESIPFED](https://github.com/esipfed)

Acknowledgments: Sky Bristol, USGS and all those who submitted ideas and/or attended the Earth Data Provenance Workshop, graciously hosted by the eScience Institute at the University of Washington.

Current RFP:

Collaboration is key. Four page proposal.
\$10K | 6 months

bit.ly/f2018rfp

