ESIP Federation: A Model for Community Collaboration & Coordination in Big Data

Leveraging Community Expertise in Data, Technology and Use

Stefan Falke
Carol Meyer
ESIP

December 11, 2012
Federation of Earth Science Information Partners (ESIP) Vision

To be a leader in promoting the collection, stewardship and use of Earth science data, information and knowledge that is responsive to societal needs.
About ESIP

ESIP Federation has far-ranging expertise and provides a neutral space for Earth & environmental science data & technology practitioners to share their knowledge, experience & technology.

ESIP Strategic and Vision Goals (subset)

- ESIP is the trusted community authority that supports the integration of science and data into mainstream use.
- ESIP increases use and value science data and information.
- ESIP articulates and measures the value and benefit of Earth science data.
- ESIP leads the development of the Earth science data and information field to:
  - recognize data management as a profession;
  - provide opportunities for continued professional development
  - train the research community on data management
  - engage the next generation of science data information professionals
ESIP Core Values

- Agility
- Collaborative
- Collegial
- Community-driven
- Innovative
- Neutral
- Open
- Participatory
- Voluntary
Things ESIP Does

- Community-generated Best Practices (e.g. Data exchange, Citation)
- Testbed (e.g. Identifiers, Ontology)
- Community Conventions (e.g. Discovery, Data Web Services)
- Professional Development
  - Technical Workshops
  - Non-technical Workshops (e.g. Evaluation, Communication)
  - Data Management Short Course/Workshops
- Outreach
  - Education (e.g. annual teacher workshop on climate change)
  - Professional Societies (e.g. AGU, GSA)
  - International Efforts (e.g. GEO, ISRSE)
- Provide Venue for Collaboration and Connections
  - Both virtual and in-person
  - Support with suite of collaboration tools
Problem: *Data producers get little professional recognition.*

- Traditional academic rewards system recognizes journal publications
- Good data are not always permanently available or traceable
  & when found, *attribution is not always given*
- No single permanent identifier scheme supports all data types

Case Study: Data Persistence and Publication
Case Study: Activities & Outputs

Activities:
- Identifiers analysis and testbed
- Community-generated consensus for data citation

Outputs include:
- Recommendations for identifiers
  - http://springerlink.com/content/52760gq3h200gw38/?MUD=MP
- Data citation guidelines
  - http://commons.esipfed.org/node/308
- Data stewardship principles
  - http://commons.esipfed.org/node/419

On the utility of identification schemes for digital earth science data: an assessment and recommendations

Ruth E. Duerr, Robert R. Downs, Curt Tilmes, Bruce Barkstrom, W. Christopher Lenhardt, Joseph Glassy, Luis E. Bermudez, Peter Slaughter
ESIP exists because of big data related challenges. ESIP was created in 1998 as a response to a 1995 National Research Council Report aimed at increasing the usefulness of NASA data.

ESIP members contribute to the growth of data and the development of new ways to gain understanding from those data.

Q1: Has your organization been impacted by the growth of data?
Q2: What has changed in your information landscape in the last 2-3 years?

- No substantial changes from a big data perspective: continued activities and trends that started at ESIP’s inception
- Increased and more diverse ESIP membership
- New types of data and ways to conduct analysis, e.g. apps, Air Twitter

Robinson et al., 2010

S. Cal Fires, Aug '09
Australian Dust, Sep '09
Arizona Dust, Dec '09

Robinson et al., 2010
Q3: What are your biggest challenges?

- Getting right information to right person at right time in right context
- Advancing community-oriented approaches (cross-organization, cross-discipline)
- Making interoperability easier
Q4: What changes do you see in the next five years?

- Simpler ways to share data and conduct analysis
- Increased role of citizen scientists
  - Volunteered geographic information
- Big data is not just about the Data
  - Data processing and analysis
  - Data lineage and provenance
- Big data challenges don’t end after the ‘first phase’ of big data analysis

Image source:
“The Data Deluge”, The Economist, Feb. 25, 2010