Stewardship and Access: Activities of the Federation of Earth Science Information Partners

Ruth Duerr
National Snow and Ice Data Center

Federation of Earth Science Information Partners

Fostering connections to make data matter!

Federation of Earth Science Information Partners

History

- Formed in 1998 by NASA
  - National Academy report recommended the creation of a federation
  - 24 original “Working Prototypes-ESIPs”
    - 12 research
    - 12 applications
    - Later expanded to include NASA DAACs
- “Constitutional Convention” (1999-2000)
  - Constitution and Bylaws
- Foundation for Earth Science (2001)
  - Nonprofit 501(c)(3) corporation
  - The Federation’s secretariat

Federation of Earth Science Information Partners

Vision Pillars (20110712 DRAFT)

- ESIP provides the **Earth science informatics intellectual commons** to drive innovation
- ESIP is the **trusted community authority** that supports the integration of science and data into mainstream use
- ESIP leads the **development of the science data information profession**
- ESIP achieves sustainability through diversification, global partnerships and partner recognition

Federation of Earth Science Information Partners

Partner Types

- Type I: data centers
  - NASA DAACs
  - NOAA (NGDC, NODC, NCDC)
- Type II: researchers and tool developers
  - Academia
  - Government labs
- Type III: application developers
  - Commercial
  - Nonprofit
  - Educational
- Type IV: strategic partners
  - NASA
  - NOAA
  - EPA

Federation of Earth Science Information Partners

Tao of ESIP

- ***We are***...
- **→ Community-driven**
  - Members are the authority
  - Voluntary
    - No requirements
    - No remuneration
    - “For the good of the order”
  - Distributed
    - Geographically
    - Topically
    - Functionally
  - Open
    - Collegial
    - Neutral forum

- ***We value***...
  - Participation
    - Share your expertise
    - Leverage others’ expertise
    - Encourage free flow of ideas
    - Exposure → opportunities
  - Collaboration
    - “Communities of practice”
  - Innovation
    - No institutional barriers
    - Results for $5K!
- ***Hybrid virtual and ‘real’ organization***

Federation of Earth Science Information Partners

Governance

- Assembly
  - 1 partner, 1 vote
  - Annual meeting in Jan.
  - Leadership elected from Assembly representatives
- Committee
  - Chair elected by Assembly
  - Chair serves on Executive Committee
- Working group
  - Created by Assembly or Committee
  - Task-oriented
- Cluster
  - Self-forming
  - For any reason
  - Ends when the last person hangs up

Federation of Earth Science Information Partners

2012 Winter Meeting Activities

- Preservation and Stewardship
  - General assembly voted to make the Data Stewardship cluster a full fledged committee
    - Curt Tilmes of GSFC is the chair
  - Assembly approved Stewardship principles and practices
    - Assembly approved data citation guidelines
    - Next steps on identifiers work and provenance and context content standard (PCCS) agreed upon
- Access
  - Earth Science Collaboratory path forward set
  - Discovery cluster progress

ESIP Statement on Data Stewardship Principles and Recommended Practices

- "a living document" under the auspices of the newly formed committee
- Currently housed at http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Principles
- Will be dated, versioned, and moved to the new ESIP site when that is ready to go
- Principles and practices organized into three sections:
  - Data Creators
  - Data Intermediaries (repositories, value-added providers and such)
  - Data Users

Federation of Earth Science Information Partners

ESIP Data Citation Guidelines

- Currently housed at http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations/provider_guidelines
- Will be dated, versioned, and moved to the new ESIP site when that is ready to go
- Mandatory Content:
  - Author
  - Release date
  - Title
  - Archive and/or Distributor
  - Version
  - Locator, Identifier, or Distribution Medium
  - Access Date and Time
- Suggested content as needed

Federation of Earth Science Information Partners

ESIP Identifiers work

- First identifiers paper was published in the Journal of Earth Science Informatics (www.springrlink.com/content/52760gq3h200gw38/full text.html)
- Identifiers testbed is being used to assess the usability of and determine what issues there are with each of the schemes identified in the paper
- Topic of the next identifiers paper has been agreed to - other than data, what other artifacts need identifiers or locators and why. Also, what are the predominate technologies for each type of artifact
- Report by NASA on their implementation plans for DOI's

Federation of Earth Science Information Partners

ESIP Provenance & Context Content Standard work

- The "Problem" - While many provenance related standards exist they generally do not address the needs for the kinds of provenance and contextual information needed by scientists to use data in climate studies
- A 1998 USGCRP'S workshop report defines these additional information needs
- ESIP is working to eventually create a more fully fleshed out standard based on the USGCRP report
- Currently a spreadsheet exists at http://wiki.esipfed.org/index.php/Provenance_and_Context_Content_Standard
- NASA has adopted much of the PCCS as guidance for their new missions at http://earthdata.nasa.gov/about-eosdis/requirements

Category Descriptions (1 of 2)

1. Category	2. Content Item	3. Definition/Description
Preflight/Pre-Operations Calibration	Instrument Description	Documentation of Instrument/sensor characteristics including pre-flight or pre-operational performance measurements (e.g., spectral response, instrument geometric calibration (geo-location offsets), noise characteristics, etc.).
	Preflight/Pre-operational Calibration Data	Numeric (digital data) files of Instrument/sensor characteristics including pre-flight or pre-operational performance measurements (e.g., spectral response, instrument geometric calibration (geo-location offsets), noise characteristics, etc.).
Science Data Products	Raw Data and Derived Products	Raw data are data values at full resolution as directly measured by a spaceborne, airborne or *in situ* instrument. Derived products are higher level products (level 1b through 4) where calibration and geo-location transformations have been applied to generate sensor units, and/or algorithms have been applied to generate gridded geophysical parameters.
	Metadata	Information about data to facilitate discovery, search, access, understanding and usage associated with each of the data products.
Science Data Product Documentation	Product Team	Names of key science team leads and product team members (development, help desk and operations), roles, performing organization, contact information, sponsoring agencies or organizations and comments about the products.
	Product Requirements	Requirements and designs for each science data product, either explicitly or by reference to the requirements/design documents. Product requirements and designs should include content, format, latency, accuracy and quality.
	Processing and Algorithm Version History	For all products held in the archive, documentation of processing history and production version history, indicating which versions were used when, why different versions came about, and what the improvements were from version to version. For all products held in the archive, the versions of source code used to produce the products should be
	Product Generation Algorithm	Detailed discussion of processing algorithms, outputs, error budgets and limitations. Processing algorithms and their theoretical (scientific and mathematical) basis, including complete description of any sampling or mapping algorithm used in creation of the product , geo-location, radiometric calibration, geophysical parameters, sampling or mapping algorithms used in creation of the product, algorithm software documentation, & high-level data flow diagrams.
	Product Quality	Description of the impact to product quality due to issues with computationally intensive operations (e.g., large matrix inversions, truncation and rounding). Documentation of product quality assessment (methods used, assessment summaries for each version of the datasets). Description of embedded data at the granule level including quality flags,
	Product Application	Useful references to published articles about the use of the data and user feedback received by the science and instrument teams about the products. Includes reports of any peculiarities or notable features observed in the products.

Category Descriptions (2 of 2)

1. Category	2. Content Item	3. Definition/Description
Mission Data Calibration	Calibration Method	The methods used for instrument/sensor radiometric and geometric calibration while in operation (e.g., in orbit). The source code used in applying the calibration algorithms. Documentation of in-line changes to calibration or to instrument or platform operations or conditions that occur throughout the mission.
	Calibration Data	Instrument and platform engineering data collected during operations (e.g., on orbit), including platform and instrument environment, events and maneuvers; attitude and ephemeris; aircraft position; acquisition logs that record data gaps; calibration look-up tables; and any significant external event data that may have impacted the observations.
Science Data Product Software	Science data product generation software and software	Source code used to generate products at all levels in the science data processing system. Software release notes, including references to versions of operating systems, compilers, commercial software libraries used in the code. Versions of science data product software should be archived for each major product release. A major product release is
Science Data Product Algorithm Inputs	Ancillary data and documentation	Complete information on any ancillary data or other data sets used in generation or calibration of the data set or derived product, either explicitly in data descriptions or by reference to appropriate publications. Ancillary data should be stored with the products unless it is available from another permanent archive facility.
Science Data Product Validation	Datasets and documentation	Accuracy of products, as measured by validation testing, and compared to accuracy requirements. Description of validation process, including identification of validation data sets, measurement protocols, data collection, analysis and accuracy reporting.
Science Data Software Tools	Software and documentation	Product access (reader) tools. Software source code that would facilitate use of the calibration data, ancillary data and the data products at all levels. Includes software source code useful for creating programs that will read and display the calibration data, ancillary data and product data and metadata values. Commercial tools should be identified with

Federation of Earth Science Information Partners

ESIP Provenance & Context Content Standard work

- The "Problem" - While many provenance related standards exist they generally do not address the needs for the kinds of provenance and contextual information needed by scientists to use data in climate studies
- A 1998 USGCRP'S workshop report defines these additional information needs
- ESIP is working to eventually create a more fully fleshed out standard based on the USGCRP report
- Currently a spreadsheet exists at http://wiki.esipfed.org/index.php/Provenance_and_Context_Content_Standard
- NASA has adopted much of the PCCS as guidance for their new missions at http://earthdata.nasa.gov/about-eosdis/requirements