Meeting Highlights Webinar

Theme: *Data to Action: Increasing the Use and Value of Earth Science Data and Information*

By the Numbers:

- 328 Attendees + more online!
- 11 Plenary Speakers
- 53 Posters, 23 Demos, & 8 Art pieces
- 90 Sessions
- 96 First-time Attendees

Official Meeting Logo

Fun Meeting Sticker

Find & Access Meeting Content

SCHED

Multi-sensor data integration for cryosphere and hydrosphere monitoring

In keeping with this year's Summer Meeting theme of "Increasing the Use and Value of Earth Science Data and Information," this session aims to explore different data streams used for monitoring of the hydrosphere and cryosphere. Earth science data for water resources monitoring has existed as field collected data, remote sensing, modeled and in situ data for decades but relatively recent increases in computational capabilities (e.g. cloud computing platforms), data storage and integration and processing methods like machine learning have allowed researchers to ask a suite of questions that rely on data from multiple sources and typologies to answer complex questions about water resources critical to humans and ecosystems. To emphasize the 'use and value of earth science data' this session will incorporate presentations on data generation and processing methods as well as applied uses of data products for water resources monitoring.

Presenter: Eric Sproles

Presentation Title: ~~Bridging the Scaling Issues~~ of Earth Observations

Slides: https://doi.org/10.6084/m9.figshare.8980400

Session Take-Aways

1. NRCS plans to convert long-term snow courses to SNOTEL, continue to pursue tech upgrades, develop new methodologies to improve accuracy
2. Machine learning can integrate satellite observations and in situ measurements to create a more complete measurement
3. UAV provide higher density albedo measurements, remote locations, multiple field sites
4. Creating an integrated system for the future to track cryospheric changes
5. Arctic Data Committee has technical and semantic guidance for integrating cryospheric data

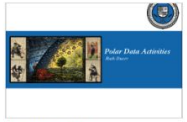
View the Recording on YouTube

figshare

National Resources Conservation Service SNOTEL Network
Scott Oviatt 01/08/2019

Integrating Satellite Observations and In Situ Measurements to Stud...
Yuhan Rao 01/08/2019

New Data, Old Problems: Integrating Novel Data Sources for Study & M...
Jeff Deems 01/08/2019



Polar Data Activities
Ruth Duerr 01/08/2019

Location, Location, Location: Enabling Data Discovery by Place
John Porter 01/08/2019

Google Colaboratory for HDF-EOS
Hyokyung Lee 01/08/2019

Detailed ecology survey data can be captured using a general purpose ...
Simon Cox 01/08/2019

CUAHSI Tools for Data Management
Martin Seul 01/08/2019

Planet data, Applications, and Interoperability
Kelsey Jordahl 31/07/2019

The Critical Zones: Supporting Place Based Research
Collin Bode 31/07/2019

Five Frontend Libraries for Visualizing Your Time-Series Data
Connor Scully-Allison 31/07/2019

The Information Management Code Registry: Software Solutions for In...
Colin Smith 31/07/2019

U.S. Energy Infrastructure: 'What's Past is Prologue'
Fred Beach 31/07/2019

The Scale and Value of Earth Observation Infrastructure
Jason Gallo 31/07/2019

Maintaining and Advancing Data Publishing at Dryad
Daniella Lowenberg 31/07/2019

Culture, Kindness, and Care: Commoning for Earth Knowledge ...
Bruce Caron 31/07/2019

By Session:
https://2019esipsummermeeting.sched.com/

Browse Presentations:
https://esip.figshare.com/ESIP_Summer_2019

Plenary Highlights

Day 1 Plenary

- Welcome, Meeting Overview, Sponsors & Lanyards
  https://youtu.be/W8-fKg4jNY8?t=639
- Plenary Speaker - Local Partner - IRIS - Robert Casey
  https://youtu.be/W8-fKg4jNY8?t=2854
- Plenary Speaker - Raskin Scholar - Kai Blumberg
  https://youtu.be/W8-fKg4jNY8?t=4292
- Plenary Speaker - Charles S. Falkenberg Award Winner - Rebecca B. Neumann
  https://youtu.be/W8-fKg4jNY8?t=5708

From Data Lakes to Rivers: Improving the Value and Reach of a Seismic Data Archive

Rob Casey, IRIS

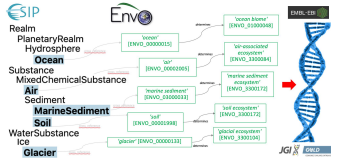
Addressing key challenge of managing and providing access to massive quantities of unstructured seismic data - Curation and Exchange, Accessible Web Services, Data Quality Assurance. Focus on credit for data producers.

Demonstration of the value and use of the archive through two case studies that make use of the archive's data services

- Los Alamos National Lab - scalable cloud-based analysis of seismic noise from wind generation driven by remote data access
- LIGO Ground Motion Prediction for assessing potential impacts from seismic events using machine learning and seismic data for the entire continental US

Toward Interoperable Microbiome Data: Bridging Earth-Systems and Life-Science Semantics

Kai Blumberg, 2019 Raskin Scholar, University of Arizona



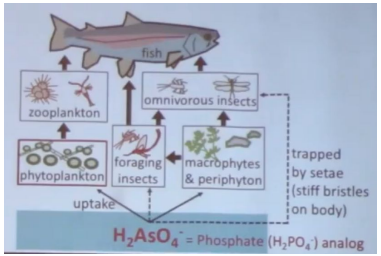
Demonstrated a model for using linked disciplinary ontologies to enable automated cross-disciplinary data integration - abiotic and biotic systems in the metagenomics example he used

Highlights a linkage between SWEET, and OBO ontologies (basic, gene, environment, and relation)

Ultimately demonstrates an enabling capacity that can provide powerful linkage of data collected in different contexts through semantic models

Legacy arsenic contamination in freshwater ecosystems: the unique vulnerability of shallow weakly stratified lakes

Dr. Becca Neumann, 2018 Falkenberg Winner, Univ. of Washington



Presentation that demonstrates through a highly effective collection of data visualizations the dynamics of arsenic uptake into lake food webs in a region subject to pollution from smelting activities.

Multiple lines of evidence: sequestration processes, physical chemistry, interaction between biological systems and arsenic species, dissolved oxygen and temperature profiles in lake waters, sediment-water interface temperature, upward and downward fluxes of arsenic, lake depth.

=> policy outcome is to focus on small, shallow lakes for contamination

Day 2 Plenary

- Plenary Speaker - Adam Mansur, Smithsonian Institution
- Plenary Speaker - Dawn Wright, Esri
- Plenary Speaker - Judy Twedt, University of Washington
- Plenary Speaker - Jennifer Hennessey, Washington State Government

https://www.youtube.com/watch?v=k8wvPJX13EE

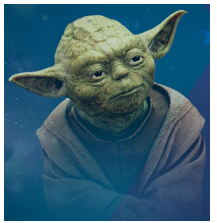
Increasing the Impact of the Smithsonian's Geological Collections

Adam Mansur, Smithsonian Institution

- 155 Million objects are managed by the Smithsonian Institution (98% biology and paleobiology), 500K of them are mineral/geoscience objects
- The Smithsonian Institution has two primary goals in maintaining these collections: 1) to educate the public and 2) to support research
- They measure their success in supporting research based on 1) the number of loans of objects and 2) the use of specimens in publications
- Metadata can be challenging as many specimen records are very old and handwritten, but they have been using International Geo Sample Numbers (IGSNs) since 2015 to track specimens
- For publications, they use GeoDeepDive to mine publications for mention of their specimens
- Adam was excited to be at ESIP to learn about advanced topics such as ontologies and Linked Data.

Ease Leads to Exposure, Exposure Leads to Adoption

Dawn Wright, Esri



- ESRI has found that ease of use leads to exposure and exposure leads to adoption
- One way to simplify access is to provide “Viewers”, for Landsat data, for example
- An “Explorer” has all the capabilities of a viewer but also allows one to compare imagery among different sources
- “Templates” allow users to build upon prior work
- “Dashboards” give insight into your data from a single screen via configurable panels
- Dawn highlighted the integration of the ESRI LivingAtlas (livingatlas.arcgis.com) and EarthCube data
- Using Responsive Design to allow apps to work across various platforms
- The FAIR data theme will become FAIRER - adding Ethical and Revisable

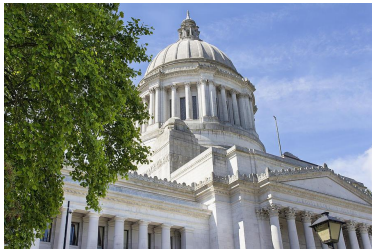
Have you heard this data? New approaches to science communication with data sonification and music

Judy Twedt, University of Washington

- Her focus is on building public trust in climate science, how do we help students and public audiences think about climate change without ignoring it?
- People see plots of data and kind of go numb or feel hopeless and frightened, not really appreciating the change of a phenomena over time
- To increase data-informed action, 1) people need ways to emotionally process environmental information and 2) it is best to link environmental data with impacts on people
- For 1), her creative solution was to take time series data and put it to music for piano. With the left hand repeating the rhythm of the seasons, while the right hand plays each month point of sea ice extent.
- For 2), She asks the question "What are your associations to arctic sea ice loss?" and played back actual stories from people living in the arctic who are seeing the changes first-hand
- To get the true impact of her work, **you must go here and listen!** https://tinyurl.com/ArcticSealceForPiano

Climate Change and Ocean Impacts: Washington's use-case for earth science data

Jennifer Hennessey, Washington State Government

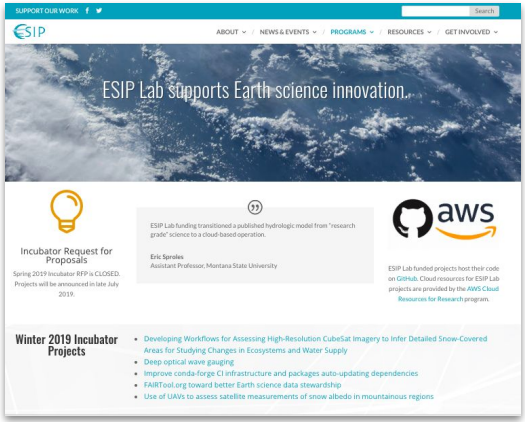


- A conversation at a high school reunion reminded Jennifer that it is important to relate science to actual impacts on people
- The "Washington Way": using innovation, science and collaboration to solve problems
- State data and Information relating to climate include: 1) Snowpack and Water Data, 2) Geological and Hydrological Data, 3) Weather and Atmospheric Data, 4) Climate Information
- The Washington State Governor has initiated many policies aimed at reducing carbon dioxide emissions
- Jennifer works mostly on the oceans, there is an ocean warming off the northwest coast known as the "Blob", 2-4 degrees F warmer which triggered harmful algal blooms and led to fishery closures - Washington is the #1 producer of farmed clams and oysters.
- The also develop and support ocean acidification forecast models via the Northwest Association of Networked Ocean Observing System, nanoos.org
- Decision-makers and the general public lack an understanding of the value and use of observational data and this is a needed focus

ESIP Lab Plenary Session

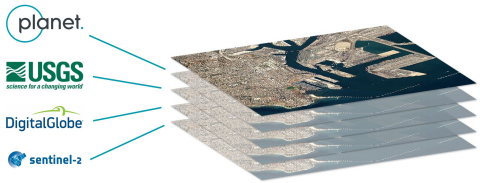
Funded Projects - esipfed.org/esip-lab/funding-opportunities

- Ziheng Sun (Geoweaver - ML workflow management)
- Abdullah Alowairdhi (FAIRtool.org)
- Eric Sproles (UAS Snow Albedo)
- Amanda Tan (CubeSats + Snow Covered Area)
- Rich Signell (Conda Forge + XrVis)



Invited Speakers

- James Bednar: PyVis and HoloVis
- Kelsey Jordahl: Planet data, applications, and interoperability

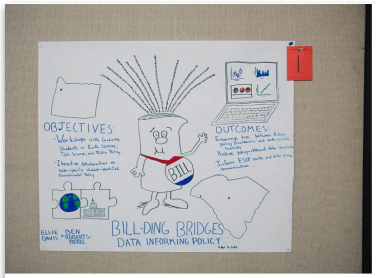


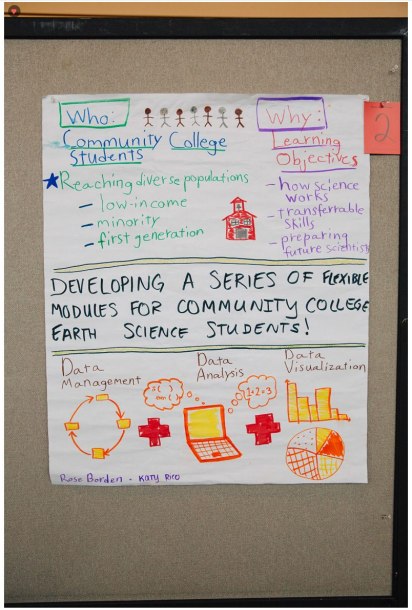
Watch talks at:
http://bit.ly/labplenary19

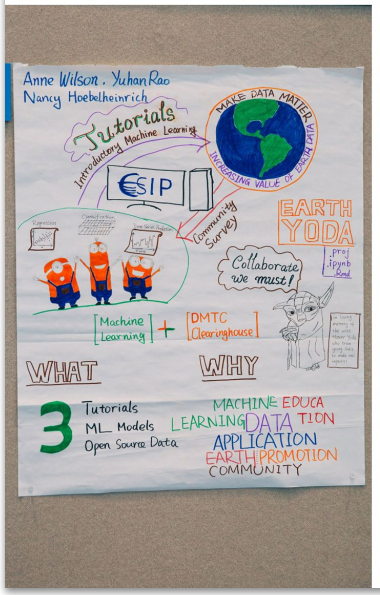
Tags:
Innovation, Funding, Collaboration

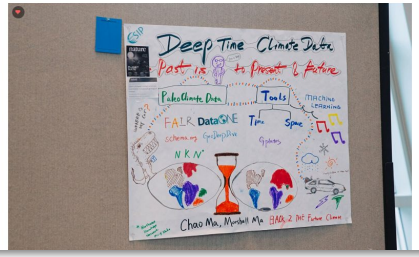
Want to learn more?
Contact: lab@esipfed.org

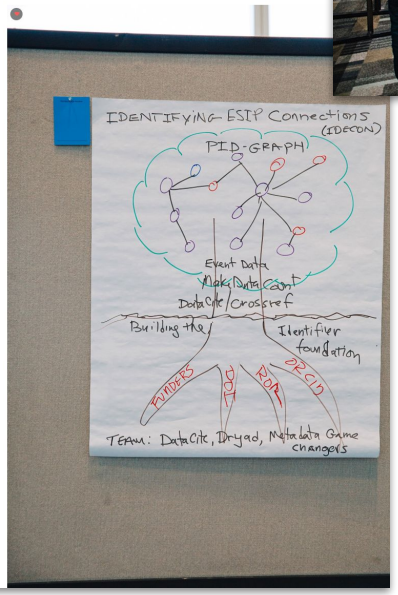
ESIP FUNding Friday 2019











3 Member Awards
3 Student/Teacher Awards

Posters will be available on the wiki soon!

Breakout Session Highlights

Bridging The Gap Between Discovery and Use (Data and Tools)

ABOUT:

To increase the value and use of Earth science data, having tools and services that can utilize data is crucial for doing scientific research. This session conveyed how metadata repositories are attempting to help users start working with their data immediately through the use of metadata modeling and intuitive discovery tools. There were talks from NASA EOSDIS, NOAA One-Stop, and DataOne/NSF.

Source: https://www.go-fair.org/resources/internet-fair-data-services/

Take-Aways:

- The “Search, Order, and Download” paradigm is changing
- Users want to start working with data more quickly and easily
- There is a gap between data and a Tool/Service to work with that data
- There is a need to have knowledge between data and a tool/service that works with that data
- Users are and want to do data analysis online without downloading anything
- NASA, NOAA, and DataOne are working on how to bridge the gap
  - NASA EOSDIS is developing ‘ToolMatch’ to match data to appropriate tools that can be discovered
  - NOAA OneStop is working to improve tool and data metadata
  - DataOne/NSF is developing ‘WHOLE Tale’ as a platform for reproduceable research
- We need your help to continue to bridge the gap and develop best practices for connecting data to tools

Tags:

Data Analysis, Data Use, Metadata, Tools, Services, Software

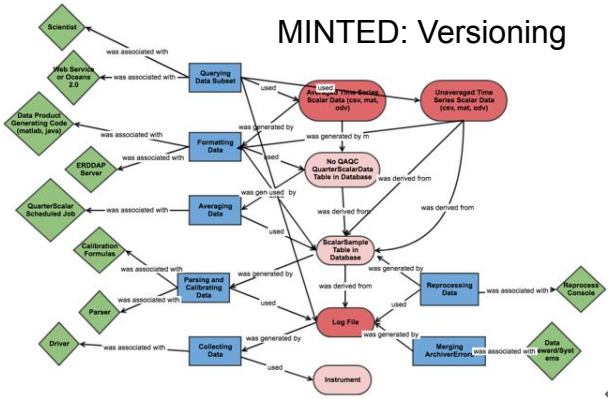
Want to learn more?

Contact: esip-documentation@lists.esipfed.org

Data Citations:What Makes a Good Citation?

ABOUT: Discuss what goes into registering PID and elements of a data citation. What are current good practices at the various repositories when it comes to dealing with the dynamic nature of data, like adding extra data, versions, general heterogeneity.

MINTED: Versioning



- **Citationing data is important for credit, reproducibility, provenance, etc.**
- **Useful to know when a DOI/PID should be produced (different file formats, versioning, data appended to a file,...)**
- **Guidelines on when data should be cited**
- **Bring comments back to the Research Objects Citation cluster**

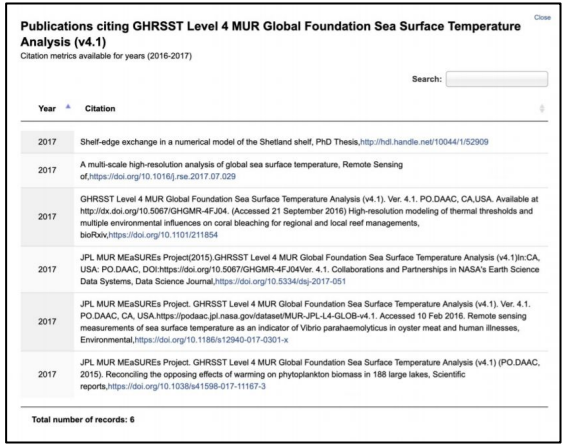
Research Object Citations Cluster
3rd Thursday at 11 Mountain Time

Tags:
PID, DOI, Citation, Data, Research Object Citations

Want to learn more?
Contact: Jessica.Hausman@jpl.nasa.gov

Current Approaches for Tracking and Exposing Research Object Usage Metrics

Three data repositories presented how they track data citations and Make Data Count representatives discussed the COUNTER Code of Practice



Publications citing GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1)

Citation metrics available for years (2016-2017)

Search:

Year	Citation
2017	Shelf-edge exchange in a numerical model of the Shetland shelf, PhD Thesis,http://hdl.handle.net/10044/1/52909
2017	A multi-scale high-resolution analysis of global sea surface temperature, Remote Sensing of,https://doi.org/10.1016/j.rse.2017.07.029
2017	GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1). Ver. 4.1. PO.DAAC, CA,USA. Available at http://dx.doi.org/10.5067/GHGMR-4FJ04. (Accessed 21 September 2016) High-resolution modeling of thermal thresholds and multiple environmental influences on coral bleaching for regional and local reef managements, bioRxiv,https://doi.org/10.1101/211854
2017	JPL MUR MEaSUREs Project(2015).GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1)In:CA, USA: PO.DAAC, DOI:https://doi.org/10.5067/GHGMR-4FJ04Ver. 4.1. Collaborations and Partnerships in NASA's Earth Science Data Systems, Data Science Journal,https://doi.org/10.5334/dsj-2017-051
2017	JPL MUR MEaSUREs Project. GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1). Ver. 4.1. PO.DAAC, CA, USA.https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1. Accessed 10 Feb 2016. Remote sensing measurements of sea surface temperature as an indicator of Vibrio parahaemolyticus in oyster meat and human illnesses, Environmental,https://doi.org/10.1186/s12940-017-0301-x
2017	JPL MUR MEaSUREs Project. GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1) (PO.DAAC, 2015). Reconciling the opposing effects of warming on phytoplankton biomass in 188 large lakes, Scientific reports,https://doi.org/10.1038/s41598-017-11167-3

Total number of records: 6

Takeaways:

- **Need to look in nontraditional places in pubs and use nontraditional tools (e.g. GeoDeepDive, podaac_metrics)**
- **Automated techniques are less extensive than manual techniques, but way more efficient for repos with limited resources**
- **Everyone needs to contribute data metrics to "Make Data Count" so we can get the whole picture of data usage**

MAKE DATA COUNT

Want to learn more?

Contact: Madison Langseth mlangseth@usgs.gov

How to Build Your Data Groups to Optimize Discovery

Workshop session to experiment with ways to determine the best level of aggregation (LoA) to optimize discovery, access, interoperability and citability.

Takeaways:

- "It's more than science, it takes a lot of experience and a little bit of art!"
- Definite interested in developing guidance with useful questions
- This will be useful for national/international catalogs
- A lot of the questions are common for different datatypes, but best LofA still varies in different communities

Tags:

Metadata, Data Management

Want to learn more?

Contact:esip-documentation@lists.esipfed.org

FAIR Metadata

The ESIP community can develop a set of automated checks to test compliance of metadata in multiple dialects with the FAIR Principles. We are starting with EML, ISO 191*, and DataCite

We are collecting community comments and suggestions at:

F - https://github.com/NCEAS/metadig-checks/labels/Findable
A - https://github.com/NCEAS/metadig-checks/labels/Accessible
I - https://github.com/NCEAS/metadig-checks/labels/Interoperable
R - https://github.com/NCEAS/metadig-checks/labels/Reusable

We Need You!

Tags:
Documentation, FAIR, Metadata Evaluation

Want to learn more?
ted.habermann@gmail.com
jones@nceas.ucsb.edu
margaret.obrien@ucsb.edu

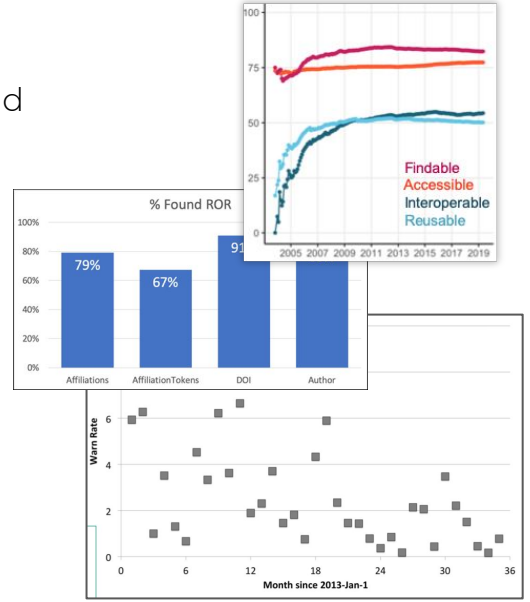
Metadata Evaluation -Tools and Results

Three repositories described their approaches to metadata evaluation and improvement:

Matt Jones, Peter Slaughter, Ted Habermann
Tools: Using the MetaDIG Quality Engine to evaluate 687,126 EML and ISO records.
Result: Generally improving over time, lots of variation in detail.

Margaret O'Brien
Tools: EDI Metadata Congruence Checker
Result: Community buy-in, warning rate has decreased significantly since 2013

Daniella Lowenberg, Ted Habermann
Tools: Using Dryad DOIs to search CrossRef, PLOS, and others
Result: 41,000+ standard affiliations and identifiers added



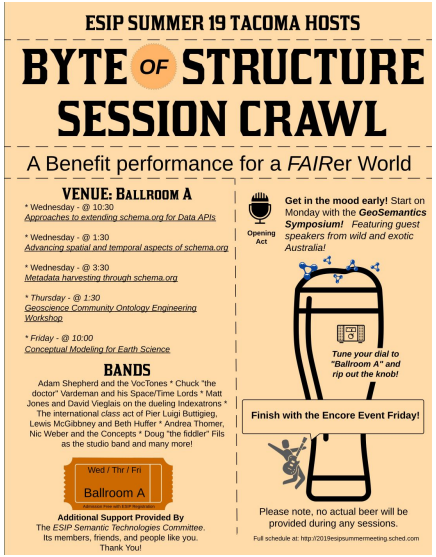
Tags:
Documentation, FAIR, Metadata Evaluation

Want to learn more?
jones@nceas.ucsb.edu
margaret.obrien@ucsb.edu
Daniella.Lowenberg@ucop.edu

Approaches to extending schema.org for Data APIs

ABOUT: schema.org can describe static Datasets, but it's difficult to accurately describe services and APIs that provide access to data. This session will bring together data API managers and curators, conceptual modelers and ontologists to model and develop a schema.org extension address accessing data through APIs and services.

- **Discussed the distinction in data accessed by direct download and data mediated by services with examples of various patterns**
- **Discussed the goals of how to do discovery across data and data services**
- **Several people and groups provided their POV & experiences**
- **Leveraging schema.org/DataFeed and also the DCAT services Reference:**
  https://github.com/earthcubearchitecture-project418/p419dcatservices/issues/6
- **Initial implementation work with CHORDS and IRIS**
- **Join the new cluster… the schema.org cluster!**
  - **And in slack on channel: sci-schemaorg**

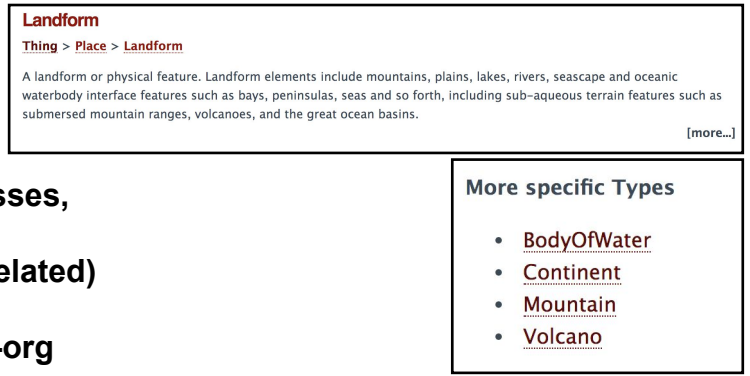


Tags:
schema.org

Want to learn more?
Contact: dfils@oceanleadership.org
ashpeherd@whoi.edu

Advancing Spatial and Temporal aspects of schema.org

ABOUT: schema.org is currently inconsistent with standards organizations (W3C, OGC) representations of spatial and temporal information. This session will bring together data curators, conceptual modelers and ontologists to formulate solutions for extending schema.org's approach to spatial and temporal descriptions.



- **Ongoing discussion about how to improve spatial https://github.com/schemaorg/schemaorg/issues/1375**
  - **within, intersects, equals, touches, contains, coveredby, crosses, overlaps, disjoint**
- **Subclasses added to schema.org/Place (starting point for all Geo-related)**
  - **https://schema.org/Landform**
- **Open-ended intervals for unbounded temporal coverages (schema-org issue:1365)**
  - **Streaming data, ongoing research w/o known end-date**
- **OWL-Time work**
  - **https://github.com/earthcubearchitecture-project418/p419voctemporal**
  - **Granularity of dates/times: intervals, durations, seasonality (OWL-Time issue: 1140)**

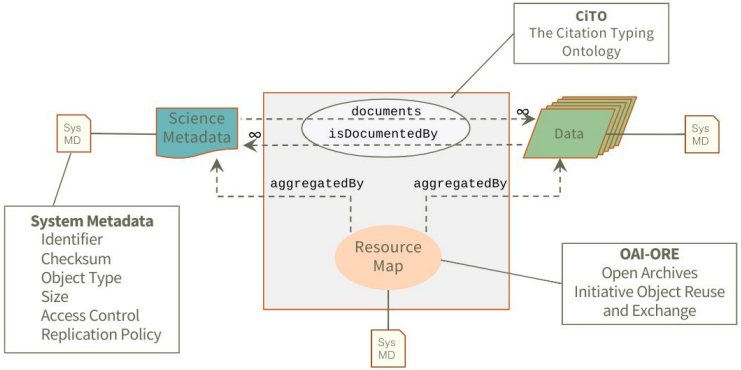
Tags:
schema.org

Want to learn more?
Contact: ashepherd@whoi.edu cvardema@nd.edu dfils@oceanleadership.org

Metadata harvesting through schema.org

ABOUT: Repositories have recognized the benefits of adopting schema.org metadata in their data catalog landing pages to improve discoverability, particularly with the incentive of inclusion in the Google Dataset search. While Google supports broad, general search and discovery, we can also use this mechanism to improve domain-specific aggregated search systems like DataONE ...

- **Prefer JSON-LD for schema.org markup**
- **Follow widely used practices for web introspection**
- **Reuse existing identifiers**
- **Structured metadata as complete as practicable**
- **Reference rich external metadata resource when available**
- **Reference well established vocabularies to reduce ambiguity**
- **Provide explicit references to dataset components**
- **Ref:** https://figshare.com/articles/Metadata_harvesting_through_schema_org/9120509



Tags:
schema.org

Want to learn more?
Contact: jones@nceas.ucsb.edu
dave.vieglais@gmail.com

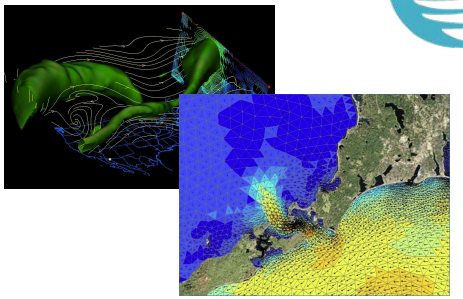
2019 netCDF-CF Workshop

ABOUT:
Community-developed metadata conventions for storing and describing earth system science data in the netCDF binary data format.

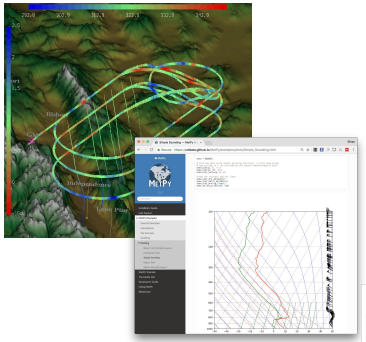
The CF community holds annual workshops to develop, refine, and review enhancements to the CF convention and to manage CF governance and processes.

- **CF Governance and Process:**
  - **Move to use GitHub to better support community involvement**
  - **CF Data Model now part of CF process**
- **Current and proposed enhancements to CF**
  - **Geometries, Groups (Approved)**
  - **Satellite Swath**
  - **netCDF for Linked Data**
  - **Uncertainty**

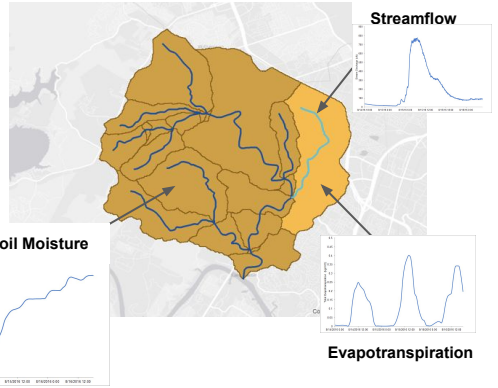
Gridded Data



Observational Data



Geometry: Data over a polyline or polygon



Tags:
netCDF,Community, Metadata Convention

Want to learn more?
Contact: edavis@ucar.edu

Cloud Data Optimization: Emerging Best Practices

ABOUT: When data is shared in the cloud, anyone can analyze it without having to download it or store it themselves, which lowers the cost of new product development, reduces the time to scientific discovery, and can accelerate innovation. However, staging large-scale datasets for analysis in the cloud requires consideration of how data should be prepared and organized to allow fast, efficient, and programmatic access from distributed computing systems.

- **Moving to cloud infrastructure offers a chance to reevaluate best practices, though some of these may not be purely cloud-related (e.g., data formats) but the discussions are coming along for the ride!**
- **It is unclear who will own the cloud-optimized datasets and it will likely be different from dataset to dataset. Until (if/when) cloud-optimized formats become the norm, they may often be provided by other groups (or created on the fly).**
- **There is a lot of focus on datasets in these conversations, but we need to also focus on tooling/services and education.**

Tags:

Cloud, data, analytics, processing

Want to learn more?

Contact: jflasher@amazon.com

Geospatial Data Analytics and Visualization for Sustainability in the Cloud

ABOUT: Sustainability's geospatial processes are complex since environmental, societal, and economic systems are deeply interconnected and extracting timely and meaningful insights often requires large datasets and tools capable of capturing the multidimensional nature of the problem. To address these challenges, many users are exploring the use of cloud computing to leverage its scalable storage and geospatial analytical capabilities.

- **The cloud has made operating at a large scale easier than ever before. As an example, Sinergise's Blue Dot Water Observatory costs a few dollars a month to monitor surface water levels of water bodies across the globe. We are seeing groups move past toy prototypes and into useable systems.**
- **Tools built for the cloud can be used around the world, but also in disconnected environments. Element 84 talked about their use case of running disaster response tooling in environments where there is no active network connectivity.**
- **We need to make sure these tools and algorithms work for everyone!**

Tags:

Cloud, data, analytics, sustainability

Want to learn more?

Contact: jflasher@amazon.com

EnviroSensing:
Sensor Data, Technology, and Best Practices

ABOUT: Our session brought together scientists, information managers, and technologists interested in in-situ environmental sensing for scientific research, and featured a series of short talks on topics ranging from next-gen water observing systems to data acquisition in remote environments to data visualization.

- **Visualization Dashboards:** Sensor users struggle with effective real-time visualization tools and data quality assessment pipelines. Network managers and scientists with deployments need effective dashboards. Viz libraries exist, but are not yet adapted into specific tools for our applications.
- **Telemetry & Emerging Tech (LoRa):** Small-scale sensor users and geographically-dense applications are leveraging emerging low-power, long-distance radio technologies, such as LoRa. This radio technology and related network topologies are driving “open” networks for data transfer in near-real-time. The USGS is applying LoRa in the Next-Gen Water Network.
- **Sensor Metadata & Raw Data Archival:** Software tools to capture uniform metadata for sensor deployments don’t exist. People at all scales are doing this manually or not at all, challenging FAIR Data practices. "Raw" sensor data are useful for science questions that leverage noise or unintended behavior. Documenting deployments and archiving raw data will improve workflows and allow use of sensor data in novel, unanticipated ways.

Tags:
Environmental Sensor Networks, Data Management, Data Visualization, Metadata, Documentation

Want to learn more?
http://wiki.esipfed.org/index.php/EnviroSensing_Cluster

Multi-sensor data integration for cryosphere and hydrosphere monitoring

ABOUT: This session brought together data managers, practitioners and academics to discuss the fusion of different datasets and methods to monitor cryospheric change.

- **Scaling issues are crucial and widespread but are starting to be addressed by leveraging multi-resolution datasets**
- **Like many geosciences, machine learning is an emerging method for tackling these large datasets**
- **There continues to be a tradeoff between operational and research grade systems (e.g. SNOTEL network)**
- **Emerging technologies (e.g. drone-mounted sensors, airborne LiDAR/hyperspectral instruments) are filling gaps in our knowledge of the cryosphere**
- **We need to continue working on semantics and usability/accessibility of cryosphere data**

Tags:

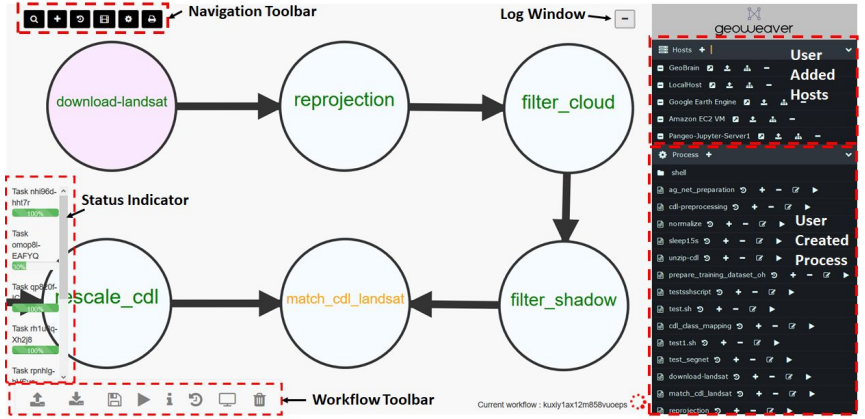
Cryosphere, drones, envirosensing, ESIPLab

Want to learn more?

Contact:
robertsb@oregonstate.edu

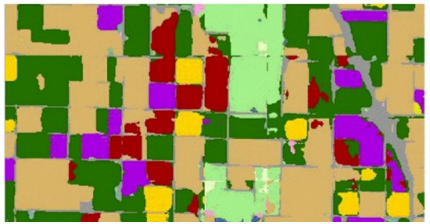
Advanced Geospatial Cyberinfrastructure for Deep Learning

We discussed the current landscape of cyberinfrastructure people use right now to do deep learning researches in geoscience.

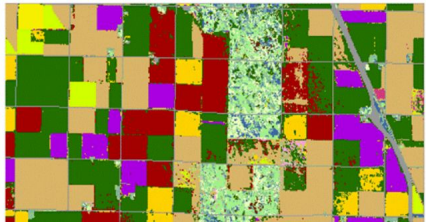


Takeaways:

- **Hybrid environment (Public & Private)**
- **Poor management is expensive**
- **Deep learning training data is key**
- **Replicability of deep learning in crisis**



Deep Learning Result



USDA Map

Tags:

Deep Learning; Cyberinfrastructure; Workflow Management System; Big Data; Cloud Computing.

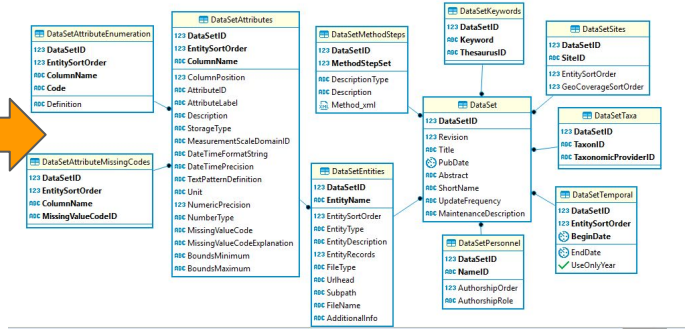
Want to learn more?

zsun@gmu.edu; annieburgess@esipfed.org

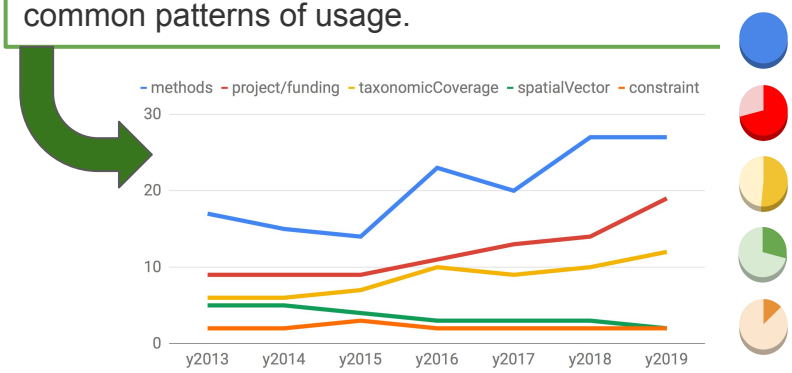
A Metadata Database Built on Usage Patterns in the LTER Network

ABOUT: Two topics, both related to Ecological Metadata Language

LTER-core-metabase is a relational database model based on the model developed at GCE LTER, with adaptations by three LTER sites, and now being advanced as a collaborative LTER project.



The second part of this session focused on a constrained profile of EML being developed by EDI to streamline script-development, but which is not tied to any specific back end storage system. Initial work is concerned with identifying common patterns of usage.



- **LTER-Core-Metabase is now 3rd generation, and designed for needs of groups creating EML**
  - https://github.com/lter/LTER-core-metabase
- **Participants agreed that an EML profile will make code development more practical**

Want to learn more?

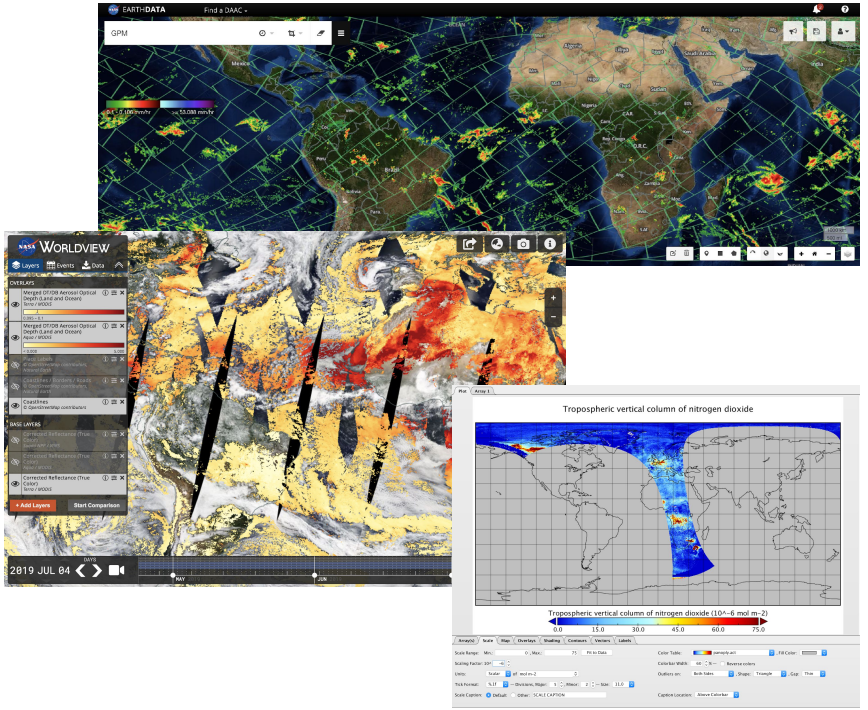
Metabase: https://github.com/lter/LTER-core-metabase
Profile of EML: info@environmentaldatainitiative.org

Tags: EML, Relational database, Metadata

Help Us Help You...

Exploring Data Pathfinders at Earthdata.nasa.gov

- NASA's Earth Science Data Systems has a wealth of information and data that can be used in a variety of projects - from research to decision making.
- The data available from the Earthdata site are not readily findable nor accessible - some beginner-oriented material is needed.
- Data pathfinders from NASA ESDS can serve as a tool to provide background on the different types of data available for a specific theme, access to commonly used datasets for a specific measurement and then resources and tools for exploring the data.



Earth science, Data Accessibility

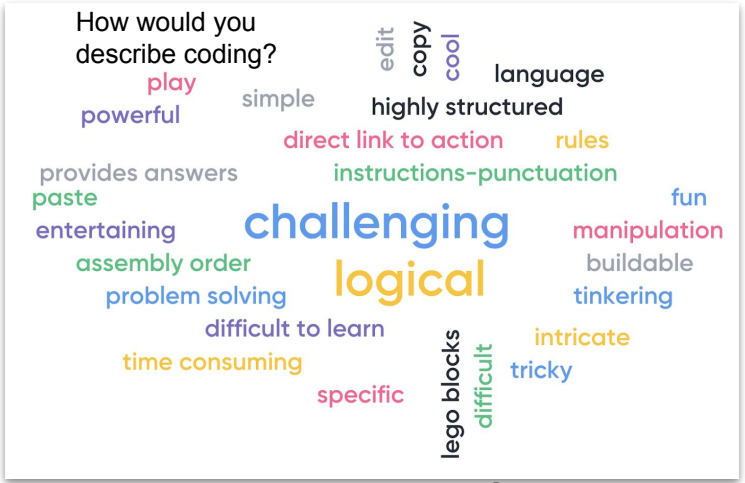
Want to learn more?

Contact: cynthia.r.hall@nasa.gov

Data to Action Educator Workshop: Jupyter Notebooks and other Earth Science Tools

ABOUT: This session brought together educators in the Tacoma and Seattle area to explore a Jupyter Notebook activity and other data tools available through ESIP community members

- GOES-R Series WebApps - participants learned how combining different GOES 16/17 channels into RGB highlights different aspects of the image.
- The Explore Atlantic Storms Jupyter Notebook provided secondary educators hands on experience (some for the first time) with viewing, modifying, playing, and running (python) code to analyze data.
- SuAVE Survey Analysis via Visual Exploration is an incredibly powerful and fun tool to explore and use in the classroom.



Tags:

Education, Jupyter Notebook, Data tools

Want to learn more?

Contact: olds@unavco.org

EPIC Fails in Earth Science Informatics

An unexpectedly large group of people attended

Easiest session to manage ever - the audience would not stop talking and offering stories of their failures

The failures were of many types:

a. Human (e.g., underestimates of how long the project would take, people competing,)
b. Technology (e.g., not keeping up with technology changes, not understanding the technology)
c. Organisational (e.g., changing organisational dynamics, changing priorities)
d. Too many more to list…..

But the most epic failure of all was that people have no where to report failures even if they wanted to

Suggestion was made to

a. Develop a blog/special journal issue/??? where people can report their failures without fear
b. Rerun the session again, but try and structure it more around types of failures!

Notes are on: https://docs.google.com/document/d/1SeRTywKNJ6FdTX0qKXOxmsube_cpo0Gn4kIBcE5IGlw/edit

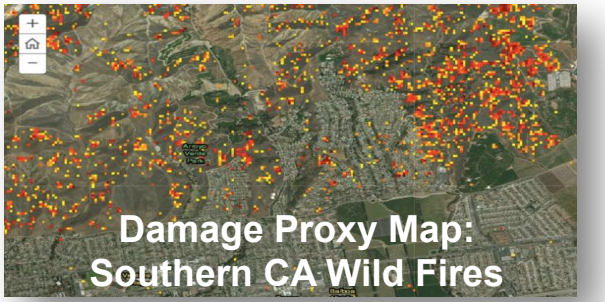
Identifying Trusted Data Sources for Operational Decision Making & ORL Criteria

ABOUT: *Exploring Federal data services to improve discovery and access. Collaborating with emergency managers to improve data-driven decision making in their environment, based on their use cases and data needs.*

- Clarifying local EOC approach to request federal resources for response and assessment (data and aircraft); More is needed to educate local and regional personnel.
- Trust issues as handled by the All Hazards Consortium using ORLs; Determining Operational Readiness Levels for federal data sets that powers 30-second decision making.
- Federal Data offerings for Disaster Response - Which agency has what information? There are at least 4 federal portals for emergency information. FEMA may be seeking a resolution, potentially an application that could tell what data has been collected and, if so, where it can be found.

Tags: Disaster Lifecycle Cluster, Trusted Data, ORL - Operational Readiness Levels, Collaboration, Data Sharing

http://maps.disasters.nasa.gov



ORL Definitions

ORL 1	Data available NOW 24/7 & Secure (SSL / HTTPS) Immediate SA & Decision Making DM [30 sec decisions] No Down Time – Operational Data is Critical to Decision Making Person available to contact (Fix link or service, report issues for open ticket)
ORL 2	Data available regularly, SA & DM [30 sec decisions] Event-driven, may be delayed due to acquisition and processing time required Likely very useful for Situational Awareness (SA) & Decision Making (DM) Person available to contact
ORL 3	Emerging operational data and/or mature testing phase Data not guaranteed Potential to improve SA and DM Target operations in 6-12 months
ORL 4	'New' emerging datasets, applications testing phase, training available Being evaluated for accuracy, validated, usefulness Target for operations 12+ months Not likely to be immediately useful for operations but could be

Want to learn more? Co-Chairs

Contact: dave@stormcenter.com
karen.moe@earthlink.net

Meet the Maintainers

Information Maintainers support the maintenance of information and those who manage, maintain, and preserve information systems.

1. ESIP scientists and data scientists and project managers are only part of a larger team that keeps Earth information active and durable. Maintenance is about fostering and caring for relationships: Who decides to maintain is all of us. This requires awareness and kindness.
2. Maintenance is more complex than it seems: Value is built over centuries, the US has no energy policy; predicting one climate variable requires dozens of inputs and complex interactions; software and data must be maintained together.
3. Infrastructure (such as our energy infrastructure) can get to the point where the trillions of dollars needed to update it might be better spent replacing it with something highly distributed. Maintenance does not mean the (pernicious) status quo.

Tags:

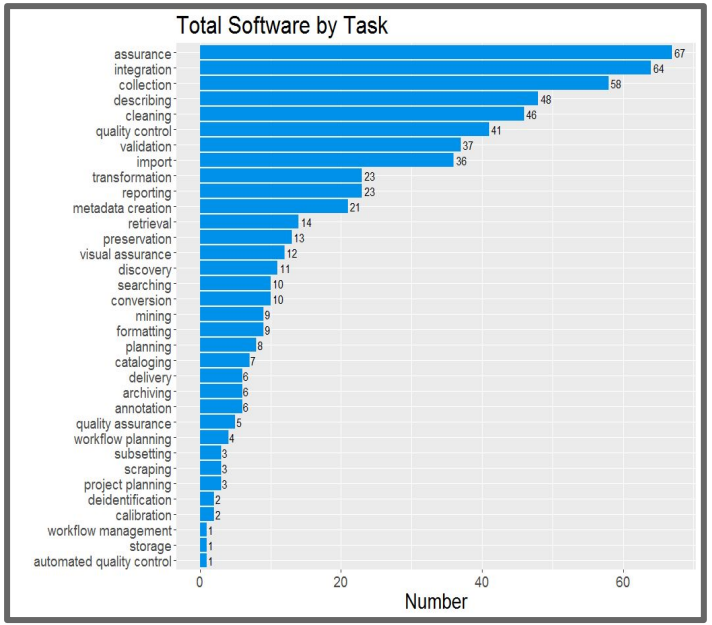
Care, Maintenance, Stewardship, Ethics

Want to learn more?

Contact: themaintainers.org

The Information Management Code Registry: Software Solutions For Information Management Needs

We presented the beta-release to gather feedback before moving to production.



Takeaways:

- **Constrain scope further**
- **Align vocabularies with GCMD and SWEET**
- **Add detail to metadata for machines**
- **Mine the web**

Tags:
Information management, Software

Want to learn more?
Contact: esip-imcoderegistry@lists.esipfed.org

Conveying Information Quality - Recent Progress

ABOUT: The Information Quality Cluster (IQC) has been active since 2014 improving understanding of various aspects of information quality and fostering collaborations nationally and internationally. Much progress has been made in several areas. Purpose of session was to share status and accomplishments and discuss future directions

- Progress was reported from NASA's Data Quality Working Group, NOAA's Data Stewardship Maturity Matrices, OGC's Data Quality Domain Working Group
- Collaborations have developed during the last two years with connections established with other clusters and non-US/international groups including E2SIP
- Community engagement efforts to seek outside collaboration at IQC-organized Fall AGU sessions have been quite successful
- Progress is being made on development of a white paper on Earth science data uncertainty.
- IQC should consider airborne and in situ data as well - some new use cases are needed

Stewardship Maturity Matrix for Climate Data (SMM-CD) of WMO Commission for Climate

SMM-CD Category

Aspect	Data Access	Usability & Usage	Quality Management	Data Management
	Discoverability	Data Portability	Quality Assurance & Control	Preservation
	Accessibility	Documentation	Quality Assessment	Metadata
		Usage	Uncertainty Analysis	Governance
			Data Integrity	

(Guidance Booklet: bit.ly/SMM-CD-Manual)

Want to learn more?
http://wiki.esipfed.org/index.php/Information_Quality

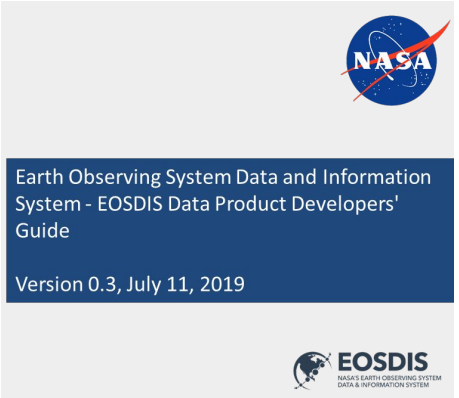
Tags:
Info Quality, Collaborations, DSMM, Uncertainty

Data Product Developers' Guide (DPDG) Workshop

ABOUT: The DPDG Working Group within NASA's Earth Science Data Systems Working Groups has been developing a guide to assist science data product developers in designing and producing products that are interoperable and conveniently usable by the community. Purpose of the workshop was to get attendees to review and comment on DPDG v0.3.

- Most attendees felt that the DPDG would be a useful contribution to help data producers make usable data products
- > 80 comments were collected during session from four subgroups reviewing different sections of document
- **Key Comments**

- Retitle the document to include "Data Producers" in title
- Provide upfront paragraph to tie sections to particular goals of particular readers
- Provide guidance on facilitating user feedback on data product usability and quality
- Rework discussion on chunk size for data
- Note how the Common Metadata Repository can make good use of certain types of metadata to make data more Findable



Want to learn more? Write to:

- Hampapuram.Ramapriyan@ssaihq.com
- pleonard@sesda.com
- Christopher.Lynnes@nasa.gov

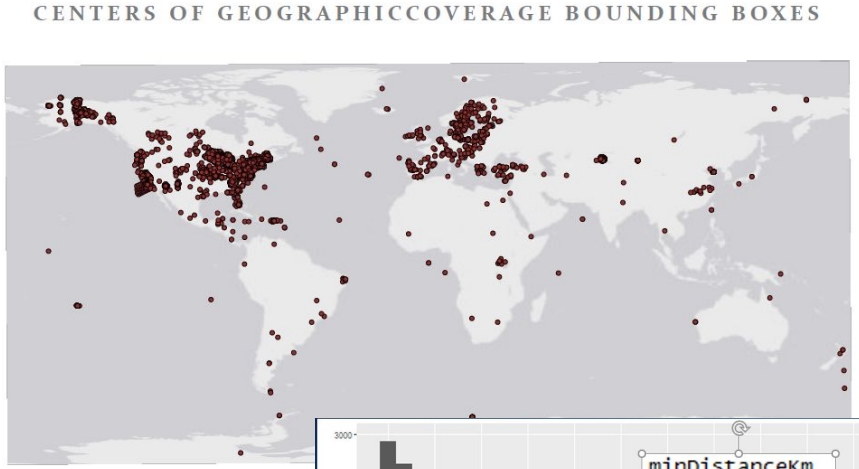
Tags:

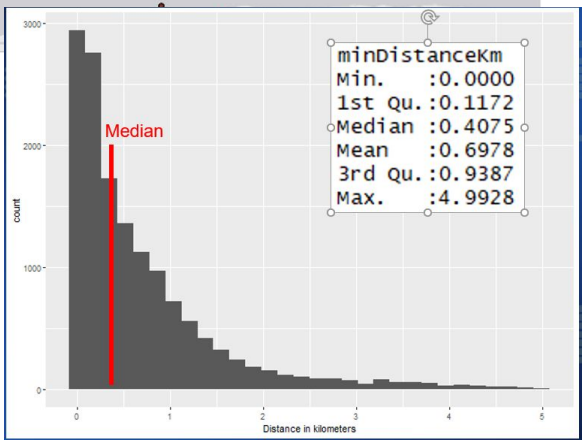
Data Products, Guide, Reusability, Interoperability

Location, Location, Location: Enabling Data Discovery By Place

ABOUT: Discussion of the use of Gazetteers to identify standardized names of places to include in metadata

- **It would be helpful to be able to search data with polygons that reflect the shape of the area being searched and are not simply a bounding box.**
- **A tool that accepts coordinates and returns the closest named place from a gazetteer would be useful to the community.**
- **Best practices for including place names in dataset documentation (e.g., always include State and County), would be welcome.**
- **Next step: Contact ESIP Documentation Cluster**





Tags:

Data Discovery, Metadata

Want to learn more?

Contact: krvander@fiu.edu; jhp7e@virginia.edu

Semantic Technologies (SemTech) Committee:

Promoting development of SemTech in support of Earth science data discovery, dissemination, and analysis

SemTech featured ***heavily*** !!!

- Full day GeoSemantics Symposium with 77 registered in cvent & 15 online
- 10 sessions directly related to SemTech - led by community members
- Creation and rollout of yet another new cluster (science-on-schema.org) this marks 3 new grassroots clusters in the last year.

Takeaways:

- **Tremendous potential exists for Machine Learning (ML) and SemTech to collaborate on prominent issues facing the ML community.**
- **SemTech committee are working on a SWEET publication which will follow-up the seminal Raskin and Pan, 2005 publication**
- **GeoSemantics Symposium will now happen twice a year**

Tags:

FAIR Data, Knowledge, Domain Applications

Want to learn more?

Contact: esip-semanticweb@lists.esipfed.org

Thank You, Meeting Sponsors!

Research as Art

esri
THE SCIENCE OF WHERE™

Bronze

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Lanyards

Element 84

Friends

AUI

In-Kind

figshare

amazon web services

Many Ways to Stay Connected

DISCOVER
Find people and tools to make your data findable, accessible, interoperable, and reusable.

COLLABORATE
Join-in or create a new collaboration area around your Earth science data challenges.

INNOVATE
Utilize small-grant funding to build or expand Earth data technologies.

NETWORK
Extend your network. Build connections across federal agencies, the private setor, and academia.

JOIN
Encourage your organization to join ESIP's 110+ member organizations. Unlock membership benefits: start new collaborations, apply for funding, and more.

Join the Data to Action Webinar Series: esipfed.org/webinars.

Sign up to receive Monday Updates: eepurl.com/rJQYn.

View & share the ESIP 1-pager: esipfed.org/onepager.

Thank you for attending the 2019 ESIP Summer Meeting Highlights Webinar!

See you at our next in-person meeting: January 7th-9th, 2020 in Bethesda, MD!
More info at https://2020esipwintermeeting.sched.com.

Questions? Contact staff@esipfed.org.