

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

[Meeting Takeaways Doc](#)

Community Participation Guidelines & Reporting: Click [HERE](#)

Analysis Ready Data in science and industry

Wednesday, January 27 • 11:00am - 12:30pm

More info & slides on Sched Session Page: <https://sched.co/g49Q>

Attendance & Check in

Add your Name / affiliation / pronouns / location / social media to a bullet

- Shannon Leslie / NSIDC / (she/her) / Boulder, CO
- Cindy Lin (Comm Fellow) / University of Michigan / She her / Ann Arbor, Michigan
- Ed Armstrong/ NASA JPL/, Los Angeles, CA
- Allan Doyle, NASA ESDIS Standards Office / (he/him) / Boston
- Aleksandar Jelenak / HDF Group / he/him / Northern Virginia
- Sara Lafia (Comm Fellow) / University of Michigan (ICPSR) / she/her / CA / @lafia_s
- Robert McGuinn / NOAA-NESDIS-NCEI / Northern Gulf Institute / Charleston, SC and Stennis, MS
- Jeff Siarto / NASA EED-2 Element 84 / Sunny East Lansing, MI / @jsiarto
- Clint Edrington / Northern Gulf Institute / NOAA NCEI
- Matthew Hanson / Element 84 / he/him / Dover, NH / @GeoSkeptic
- Gary Strand / NCAR-CGD
- Tony Castronova / CUAHSI / (he/him)
- Rob Crystal-Ornelas / Berkeley Lab / he/him/ Berkeley, CA / twitter: @rob_c_ornelas
- Brianna Key, Spatial Front Inc, NOAA/NOS Silver Springs MD
- Bidhya N Yadav, Ohio State University
- Dave Jones, StormCenter Communications | / GeoCollaborate /Annapolis, MD/he/him/ @stormcenter410 @Geocollaborate
- Marseille Bunk, ESDIS/GSFC
- Bob Chen / CIESIN-Columbia University & NASA SEDAC/ he-him / Falmouth MA
- Julien Chastang / UCAR-Undiata / Boulder / @julienchastang
- Stephan Klene/SSAI- NASA GSFC ESDIS/Northern Virginia
- Tracey Pilone/Element 84/Alexandria
- Sara Lubkin/NASA ESDS/MD
- Doug Schuster/NCAR
- Daniel Fuka/Virginia Tech/Albuquerque, NM
- Ryan McGranaghan / ASTRA LLC & NASA Goddard Space Flight Center / he-him / Mount Rainier, MD / @AeroSciengineer
- Karen Stocks / Scripps Institution of Oceanography / she, her / San Diego
- Rob Redmon / NOAA NCEI and NOAA Center for AI / he him / Boulder / @spwxrob
- Tyler Christensen, NOAA National Ocean Service, she/her, Silver Spring MD

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

-
- Angela Li, NASA/GES DISC, she/her, Greenbelt, MD
- Mei-Ling Freeman, NOAA NESDIS, she/her, Asheville, NC
- Jessica Hausman / NASA / she/her / Alexandria, VA
- Karen Moe, NASA Retired, she/her, Cheverly, MD
- Joe Lee, The HDF Group
- Ken Tanaka / University of Colorado-CIRES and NOAA-NCEI / he, him / Boulder, CO
- Diane Davies, NASA ESDIS
- Rich Signell USGS
- Sahara Ali / University of Maryland Baltimore County / she, her, hers / Baltimore, MD
- Ziheng Sun, George Mason University
- Crista Straub / USGS / she/her / Fort Collins, CO
- Doug Newman / NASA EED
- Ethan Davis, UCAR Unidata, he/him, Boulder, CO
- George Percivall, OGC
- H. K. "Rama" Ramapriyan, SSAI & NASA GSFC, remote worker (Camarillo, CA)
- Ge Peng, UAH/NASA MSFC IMPACT/Asheville, NC
- Jackie Kendall, SSAI, Lanham, MD
- Long Pham, NASA GES DISC, Rockville, MD
- Cyndy Parr, USDA NAL / she/her / Beltsville MD
- Heather Brown/NCEI/she-her/Asheville, NC
- James Gallagher, NASA/EED-2/OPeNDAP, Butte MT
- Patrick Quinn, NASA / EED-2 / Element 84, Ambler, PA
- Kim Valentine / NOAA / she, her / Charleston, SC
- Scott Henderson / UW eScience Institute / he,him / Seattle, WA
- Vitor Martins, MSU, East Lansing, MI
- Tim Ortiz, SAIC, Southern MD
- Jon Vandegriff, JHU Applied Physics Lab
- Steve Olding, NASA ESDSWG/ESO
- Felimon Gayanilo / TAMUCC, GCOOS / he,him / Corpus Christi, TX
- Leslie Hsu, USGS
- Byron Peters, NASA/SSAI, Greenbelt, MD
- Harry Singh, Ocean Networks Canada, Victoria BC

Agenda

- **Welcome**
 - Orient attendees to tech (Zoom, QiqoChat, Slido, this Google/Notes document)

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- Invite attendees to share their name/affiliation/location in the Zoom chat and in this document
- Invite others to take notes in this document (if desired)
- Slido poll question(s)
- **Community Participation Guidelines Reminder**
 - By participating in this session, you agree to adhere to the ESIP [Community Participation Guidelines](#)
 - Report an Issue: <https://www.integritycounts.ca/org/esip>
- **Introduction, logistics, attendee poll**
- **Speaker presentations, Q&A**
 - Ignacio Zuleta (ARD Zone): *Virtual constellations, ARD and sensor fusion: the future of earth observation* [Missed presentation, Technical Issues]
 - Steven Labahn (USGS, CEOS Land Surface Imaging Virtual Constellation (LSI-VC) Co-Lead): *CEOS Analysis Ready Data (ARD)*
 - Chris Lynnes (NASA): *Analysis Ready Satellite Data in NASA's EOSDIS*
- **Breakout room discussions**
 - P
- **Summary reports, session takeaways**

Resources for this session

- **COAST:** <https://ceos.org/ourwork/ad-hoc-teams/ceos-coast/>
- OGC now has a specification for Discrete Global Grid Systems: <https://www.ogc.org/projects/groups/dggsswg>
- Cloud Computing Cluster knowledge base and artifacts on data optimization (additional resources on first page of doc)
 - https://docs.google.com/document/d/1yQE4r1XuT_AP5yVcHxx5P-y9kZ_NDtPP8Ess5ZISUU/edit#heading=h.s547ccq8v1m9
 - <https://esipfed.github.io/cloud-data-optimization/>
- Pangeo COG best practices
 - <https://github.com/pangeo-data/cog-best-practices>
 - <https://discourse.pangeo.io/t/cloud-optimized-geotiffs-pangeo-best-practices/999>

Session notes

Speaker Q&A

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- Steven Labahn (USGS, CEOS Land Surface Imaging Virtual Constellation (LSI-VC) Co-Lead): *CEOS Analysis Ready Data (ARD)*
 - Ryan McGranaghan: Does CEOS provide more granularity on what 'use with ease and confidence' means for a new user?
 - Steven replied: Ease of use - guaranteed that product is met with a certain kind of quality and standard. Showed scorecard summary tables - level of assurance around what a data product contains
 - Are all data transformations (data and code), from original to ARD, saved? What system is used to save this history?
 - ARD is a powerful "holy grail" and CARD4L is so well-defined, is there a danger that people might think ARD is a solved problem
 - Why did you decide not to specify the metadata format & data packaging? That seems like a key usability requirement.
 - In STAC, we are currently working on creating CARD4L extensions, and there are two open PRs for SAR and Optical.
 - Can you say a few more words about COAST?
- Chris Lynnes (NASA): *Analysis Ready Satellite Data in NASA's EOSDIS*
 - *pixels expand near the pole simply because of the projection used. the FOVs are function of off-nadir angle only*
 - *One of the challenges is that the "processing level" definitions appear to be different between land and atmosphere... where in land Level-2 is gridded/ground referenced whereas that's Level-3 in the atmosphere domain*
 - *How might the ESIP endorsed Analytics definition be updated with ARD concepts?* <https://datascience.codata.org/articles/10.5334/dsj-2017-006/>
 - *Confounding factors?*
 - *In terms of "easy to used spatial characteristics", have you considered using a standardized hierarchical spatial indexing system like H3? see H3geo.org.*
 - *How did you create AIRS L2G time-aggregated data before you run PANDAS on it? Did you store everything in memory or save it in an intermediate format?*

Ed Armstrong - Breakout room 1 notes:

- **For ARD producers:** What is your (organization's) approach to producing ARD?
 - No real interaction on this topic
- **For ARD users:** What use cases does ARD fulfill for you?
 - Ed: Interdisciplinary use cases important
 - Douglas: Toolkit and services important for on the fly generation
 - Rob R (Space Scientist): Uncertainty documentation important ! Machine consumable. Ready for data assimilation.

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- Byron P: GIS applications are important. Issue quality control for data preparation and data understanding.
- Discussion about Pangeo.....
- **Should ESIP get involved in ARD? How?**
 - ESIP cloud computing cluster is one possibility: P Quinn working on data chunking.....
 - Annie: How can ESIP facilitate? What are the resources needed?. Define the problem space.
 - Rich: Training on resources and tools important. Its bifurcated, different material at different institutions.
 - Alek: Follow OGC Testbed model. Sponsered orgs provide common data. Like a test bed that people could evaluate.
 - Chris L: OGC is time consuming. Better HDF “Zoo”, where data can be collocated with tools

Participants: SJK, Rick Signell, Ignacio Zuleta, Byron Peters, Annie B, Jeff Siarto, jhrg5901, Madison L, M Freeman, Monica Youngman, Rob Redmond, Douglas Rao

Shannon Leslie Breakout room 3 notes:

- **For ARD producers:** What is your (organization’s) approach to producing ARD?
- **For ARD users:** What use cases does ARD fulfill for you?
- Should ESIP get involved in ARD? How?
 - If ESIP had a group that helped with terminology, collecting info. How should people understand the different terms?
- Cloud-focused discussion:
 - ESIP cloud cluster has talked about ARD; there is a document with this info
 - Pangeo project has a draft standard for metadata that includes chunk locations in files you are pointing to; provides universal way of making any data format cloud-optimized
 - Metadata becomes very important (CF conventions) with cloud optimized data
 - Can’t hide metadata through some layer of service
 - Expose grid in same way; need to interpret metadata in there and present same way
 - Question: Are cloud-ready and ARD synonymous?
 - Not necessarily, but all (cloud) workflows include the re-chunking step.
- Other
 - Need to expand ARD definition beyond satellite
 - ARD becomes more important as access methods expand
 - Seems to be emphasis on time series (currently)

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- People generating time series are maybe going away from COG, going to zarr...
 - Want to expand ARD to climate data, other types of data
 - Forecasters traditionally look at 2D maps, but now using 3D
 - Broaden definition of ARD in future; ARDs depend on users - e.g., Netflix users want Earthdata in video. Alexa users want Earthdata in voice. Minecraft users want Earthdata in 3D blocks.
 - ARD is a format that users can understand and put into their tools
 - Liked the idea of improving the tooling and helping people learn how to do
 - ERDDAP: <https://coastwatch.pfeg.noaa.gov/erddap/index.html>
-

Steve Olding / Chris Lynnes - Breakout room 2 notes:

- **For ARD producers:** What is your (organization's) approach to producing ARD?
- **For ARD users:** What use cases does ARD fulfill for you?
- Should ESIP get involved in ARD? How?

- Many ways to make data ready for analysis - e.g. make tools work better, put out more information with the data, providing training or how-to guides.
- Analysis ready may mean different things to different users / domains.
- Useful to look at extreme cases? May have some data that are useless for analysis. First step identify data that can be made analysis ready - ease of upgrading with automated tools. Maybe able to design thresholds - from unusable > usable with manual work > usable automatically. Non-binary forms of analysis ready data e.g. ready for certain types of analysis.
- On-the-fly ARD provides an option to defer some decisions e.g. what projection to use.
- Pre-process steps to get to what a user would consider AR. Difference seems to be whether those steps can be automated or manual. Seems like the same destination is desired but the approach is different.
- ARD need to be collated and formatted in a similar way.
- Tiling needs to be addressed in ARD.
- Data formats and metadata also important.
- Someone has to pay for the processing to get to ARD.
- Users want predictability and repeatability.
- Performance issues with on-the-fly ARD. Will diminish with move to cloud and less expensive compute power.
- Different needs. NASA primarily serves the research community vs. application or commercial community. Would it be useful to Survey users?

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

Allan Doyle - Breakout room 4 notes:

- **For ARD producers:** What is your (organization's) approach to producing ARD?
 - Pulling data out of HDF files, providing it in other formats (e.g., csv) for ML pipelines, etc
 - Search by variables (e.g., NO₂), find the data, and the CSV data is displayed (Lingua Logica): <https://nasamadesimple.net>
 - Developed a way of reading metadata
 - Created a basic streaming interface (as csv, json, binary) w/applications in heliophysics; not yet migrated to the cloud
 - Very basic (programmer focused) web site: hapi-server.org
 - There are clients and server codes also in github
 - The specification is available here:
 - <https://github.com/hapi-server/data-specification>
- **For ARD users:** What use cases does ARD fulfill for you?
 - Data format (how to read conventions in a standardized manner)
 - Need it for ML and Deep Learning
 - Which data structure do we prefer to read our data? NumPy
 - Challenge: no standardization, VIIRS vs Helio, cmip6 repository - registered themselves with that particular organization - ESFG
 - Provided filters and which organizations you want data from (NASA...)
 - Data come from models, not real observations
 - Has anyone worked with CMIP6 data repositories? Though it comprises model/simulated data, a similar repo is needed for observational data
- Should ESIP get involved in ARD? How?
 - Update the [endorsed definition](#) of *analytics* (*versus analysis*)?
 - *Analytics: already done analysis*
 - Not all the same (for some, gridded data)
- Other:
 - Distinction between analysis ready data and tools that help users (e.g., for viz): not by preparing the data
 - Dataset interoperability
 - Standardization - data volumes are much larger in earth science.
- High-level takeaways
 - ARD on-the-fly can help tailor data for specific analysis needs
 - ARD Services
 - ARD tools
 - ARD can feed AI/ML

- See AI-themed sessions during rest of ESIP meeting
- It's hard for users to talk about ARD without talking about "format", i.e. data format or format of data stream from a service

Before session close

- **Capture 3 Takeaways & Shared insights**

- What did we discover?
 - i. Analysis means different things for different people. Research vs applications for example. Can be different for different disciplines (Space vs earth)
 - ii. On-the-fly analysis data - as a way of providing "ARD" like access to L1/2 data
 - 1. L2 could be treated as point cloud data
 - iii. Importance of metadata conventions for cloud-optimized formats
 - 1. Recognition: Time-series, expanding other kinds of model data and non-satellite data really
 - iv. ARD services and tools, not just data; ARD for machine learning pipelines - hard to talk about ARD without talking about format and to a lesser extent services
- How can we help each other?
 - i. ESIP Cloud Computing clusters
 - ii. Interactive workshop in future: On the fly and data conditioning and prep for ARD
 - 1. How can we leverage HDF "Zoo"
- What knowledge and resources do we have to share?
 - i. Knowledge and experience in ARD is bifurcated. Bringing together experiences and approaches is challenging
- Important remaining questions?

- **Closing circle**

- Possible next steps
 - "Test bed" approach with datasets (OGC model - data + recipes) sponsored by ESIP providing data in a standard way (but labor intensive)
 - Zoo of analysis ready data - how well is this tool working with ARD?
 - Related discussions ongoing in Cloud Computing Cluster (not so much metadata content, but with cloud-optimized formats- ARCO!)

2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

- Shout-outs & thanks

5 ways to stay involved with ESIP

- Join the conversations on our [ESIPFed Slack Channel](#) and say hello on #general.
- Keep up with the latest from in and around ESIP by signing up for the [ESIP Monday Update](#) or the mailing list of this group.
- Share your feedback, questions, and suggestions with our Twitter community [@esipfed](#) and [#esipfed](#).
- Hop on an [ESIP Telecon](#) that interests you - no need to RSVP!
- At this meeting, check out the Research Showcase Poster and Demo Gallery.

Feedback

Please take a moment to share your thoughts on this session in Sched by going to the Sched session page linked at the top of this document.

Paste in Zoom Chat notes that you would like to save here

Slido Poll Results

When it comes to Analysis Ready Data, I am ...

0 3 4

enthusiastic



cautious



confused



2021 ESIP Winter Meeting (Jan. 26th-29th, 2021)

<https://2021esipwintermeeting.sched.com/>

Community Participation Guidelines & Reporting: Click [HERE](#)

In a word or phrase, what does Analysis Ready Data mean to you? (You can enter multiple responses.)

0 2 7

