NASA Internal Analysis Ready Data (ARD) Workshop Report
(Virtual Workshop held on November 20, 2020)

Status of this Memo

Suggested Citation

Change Explanation

This is the first version.

Copyright Notice

Abstract

On November 20, 2020, NASA ESDIS hosted an ARD Workshop with participants from across NASA's Earth Science Division (ESD). The goals of the workshop were to examine applications of the CEOS Analysis Ready Data for Land (CARD4L) [1] definitions and specifications as they applied to a select spectrum of NASA Earth Observing System (EOS) collections and come to agreement on NASA's definition of ARD. The workshop also strived to identify synergies, gaps, successes, and challenges in the ARD assessments that would guide a NASA process and policy to identify its data products (initially land) as ARD compliant. This is a report from that workshop.

Table of Contents

ESO-PUB-002
Category: ESO Publication
Updates/Obsoletes : none

ESO Staff
July 2021
ARD Workshop Report

1   Introduction

Analysis Ready Data (ARD) is a term increasingly used by scientists and information technologists to describe data products that are ready for immediate analysis with little or no data preconditioning or other modifications. Although it is a rather recent term, it can be argued that many popular NASA Earth Science remote sensing collections have always been ARD by virtue of their data formats and packaging, and the commensurate data documentation, including uncertainty and quality information. Some examples of these are the interdisciplinary NASA MEaSUREs collections.

In the past few years robust definitions of ARD have emerged, most prominently in the Committee on Earth Observation Systems (CEOS) community with the development of the CEOS Analysis Ready Data for Land (CARD4L) Product Family Specifications (PFSs). The PFSs serve as templates with specific criteria for assessing a data product and its remote sensing measurement (land surface temperature, land surface reflectance, radar etc.) as ARD. Only if a data product meets all the PFS "threshold" criteria (100% success) in the categories of General, Per Pixel, Radiometric and Atmospheric Corrections, and Geometric Corrections is it deemed ARD compliant (and also only after an expert review of the results). However, even with the pioneering work of CEOS, different disciplines and science communities (e.g., satellite data vs. Earth System Models; research vs. applications) may still have different interpretations of what ARD means.

On November 20, 2020, NASA ESDIS hosted an ARD Workshop with participants from across NASA's Earth Science Division (ESD). The goals of the workshop were to examine applications of the CARD4L definitions and specifications as they applied to a select spectrum of NASA Earth Observing System (EOS) collections and come to agreement on NASA's definition of ARD. The workshop also strived to identify synergies, gaps, successes and challenges in the

ESO-PUB-002   **ESO Staff**
Category: ESO Publication   **July 2021**
Updates/Obsoletes : none   **ARD Workshop Report**

ARD assessments that would guide a NASA process and policy to identify its data products (initially land) as ARD compliant.

2   Workshop Overview

Workshop attendees are listed in Table 1 in the Appendix and presentations are provided via the NASA Internal Workshop Confluence page [2]. After an introduction to the workshop, presenters provided brief assessments of select NASA EOS data products against the CARD4L specifications and other ARD criteria. Presentations were followed by discussions on challenges, recommendations, and next steps.

3   CARD4L Assessments of NASA EOS Data Products

This section provides summaries of each presentation and Appendix Table 2 provides an overview of the data product assessments against the CARD4L specifications.

3.1   MODIS Level-3 Land data products

The MODIS Land product diagnostic assessment includes Terra/Aqua/Combined data products on the sinusoidal grid: MOD09A1 (Surface Reflectance), MOD13Q1 (Vegetation Indices), MCD43A4 (BRDF-Adjusted Reflectance), MOD11A1 (Land Surface Temperature). These data products do not meet some critical requirements including water vapor corrections, but overall meet many other CARD4L threshold requirements.

3.2   Sentinel-1 On-demand Radiometrically Terrain-Corrected (RTC) data products

The Sentinel 1 On-demand RTC products are output in GeoTIFF format; their compatibility with ArcGIS makes these products GIS-ready. However, in order to comply with the SAR CARD4L specifications, we need to create machine-readable, standardized metadata and masks, and perform geolocation accuracy analysis.

3.3   MODIS Aqua/Terra Level-2 Sea Surface Temperature data product

The GHRSST MODIS Aqua/Terra L2P SST were assessed using the Surface Temperature CARD4L PFS. They were compliant in several important respects

- Complete pixel level quality and uncertainty information

- Well described coordinate systems and geolocation via CF metadata conventions and netCDF data model

- CF/ACDD standardized metadata improves machine readability and data usability

However, these products had some missing or incomplete assessment factors (extrapolating from the CARD4L criteria):

- Ungridded L2 pixels do not "line up". Must be transformed to L3 for easy time series analysis

- Algorithm and provenance information not completely documented

- Others factors not relevant or partial pass

It should also be noted that the foundational state of these datasets is suitable for downstream data manipulation services (e.g., data transformation), allowing for possible creation of on-demand ARD.

3.4   MODIS Level-3 Snow data product (MOD10A1)

The MODIS L3 Snow data were assessed against the Surface Reflectance CARD4L PFS. Several of the quality- and correction-type CARD4L requirements are not in this Level-3 data product, but likely in the input Level-1 and/or -2 data products. The data product also meets CARD4L geospatial requirements. In addition, some of the Surface Reflectance CARD4L PFS requirements are arguably irrelevant to the analysis-readiness of derived data products, like this snow product.

3.5   Challenges and Issues

The workshop enabled the identification of several key challenges to making NASA ARD compliant.

- Some NASA EOS Level-3 data products may not meet the CEOS ARD per-pixel quality and correction requirements, since these attributes are typically provided in the upstream Level-1 and/or -2 data products.

- NASA EOS swath data products (i.e., many Level-1 and -2 products) don't typically meet CEOS ARD geospatial requirements due to irregular pixel size, spacing and irregular location over time.

- CARD4L Pass/Fail scoring is problematic for NASA EOS data products since data products may meet most criteria but may still fail the strict specification.

- Processing level terminology in the SAR community differs from that in the optical community

- CEOS ARD emphasizes time series analysis specifically, which only encompasses a portion of NASA EOS ARD use cases. (see https://wiki.earthdata.nasa.gov/display/ESDSWG/ARD+Use+Cases)

- Existing CEOS ARD PFSs may not be applicable for many derived NASA EOS data products and known NASA ARD use cases that involve usage across different types of instruments and parameters.

3.6   Recommendations and Next Steps

The key recommendations from the workshop are:

- Continue to engage with CEOS on the evolution of ARD specifications, including:
  - o   Non-land data product specifications
  - o   Additional SAR specifications

- o   Pass/Fail scoring alternatives

- o   Strictness of pixel colocation requirement

- For NASA EOS products that are in close alignment with CARD4L (e.g., MODIS Level-3 Land data products), make minor changes necessary for CARD4L compliance and work with CEOS on certification.

- Improve quantification of information loss in interpolation methods from Level-2 to Level-3/4 gridded data.

- Promote and improve data services (e.g, transformation/reformatting) for on-the-fly ARD creation.

- Promote NASA EOS Level-4 model and other assimilated data products as ARD and determine CARD4L compliance.

- Explore whether a more nuanced approach to labeling Level-1 and -2 data as Analysis-Ready (e.g., ready for certain kinds of analysis) might help NASA's user community.

3.7   Summary

The workshop elucidated the state of a select, but representative set of NASA EOS data products evaluated against the CARD4L specifications. Shortcomings were identified in assessments of the derived data products (ocean, snow, SAR), including the pixel co-location requirement in lower processing levels, and the overall scoring philosophy of the CARD4L specifications. To address these shortcomings, significant changes to the existing CARD4L PFSs or its assessment methodology would need to be implemented. NASA EOSDIS should endeavor to engage with CEOS on these new needs with the ideal outcome of eventually making their ARD specifications more NASA friendly, while at the same time pursuing the implementation of small adaptations to select Level-3 and -4 collections so that they completely adhere to the existing CARD4L specifications.

4   References

[1] CEOS Analysis Ready Data

[2] NASA Internal Analysis Ready Data (ARD) Workshop Confluence page

5   Authors' Addresses

Edward M Armstrong
NASA Jet Propulsion Laboratory
Physical Oceanography DAAC
MS 158-242
Pasadena, CA 91109

Shannon Leslie
National Snow and Ice Data Center

University of Colorado
Boulder, CO 80309-0449

Christopher Lynnes
NASA Goddard Space Flight Center
Code 586
8800 Greenbelt, MD 20771

Siri Jodha Singh Khalsa
National Snow and Ice Data Center
University of Colorado
Boulder, CO 80309-0449

ESO-PUB-002
Category: ESO Publication
Updates/Obsoletes : none

ESO Staff
July 2021
ARD Workshop Report

Appendix

Table 1. Workshop attendees (affiliation) and email addresses

Aaron Friesz (USGS)	afriesz@contractor.usgs.gov
Andy Mitchell (NASA ESDIS GSFC)	andrew.e.mitchell@nasa.gov
Brian Killough (CEOS SEO/NASA LARC)	brian.d.killough@nasa.gov
Chris Lynnes (NASA ESDIS GSFC)	christopher.s.lynnes@nasa.gov
Christine Bognar (NASA HQ)	Christine.McMahonBognar@nasa.gov
Dave Borges (NASA LARC)	david.borges@nasa.gov
Dave Meyer (NASA GSFC)	david.j.meyer@nasa.gov
Diane Davies (NASA ESDIS GSFC)	diane.k.davies@nasa.gov
Ed Armstrong (NASA JPL, ESO)	edward.m.armstrong@jpl.nasa.gov
Eric Fetzer (NASA JPL)	eric.j.fetzer@jpl.nasa.gov
Franz Meyer (NASA ASF)	fjmeyer@alaska.edu
Jorge Vazquez (NASA JPL)	jorge.vazquez@jpl.nasa.gov
Nancy Searby (NASA HQ)	nsearby@ndc.nasa.gov
Shannon Leslie (NASA NSIDC DAAC, ESO)	shannon.leslie@colorado.edu
Siri Jodha Khalsa (NASA NSIDC DAAC, ESO)	sjsk@nsidc.org
Stephen Olding (NASA ESDIS GSFC)	stephen.w.olding@nasa.gov
Tom Logan (NASA ASF)	talogan@alaska.edu
Wenying Su (NASA LARC)	wsu@ndc.nasa.gov
Vardis Tsontos (NASA JPL)	vtsontos@jpl.nasa.gov

Table 2. CARD4L Assessment Overview of NASA EOS data products

	CARD4L Requirements Categories			
Data product(s)	General metadata	Per-pixel metadata	Corrections	Geolocation
MODIS L3 Land	Pass	Pass[1]	Pass[2]	Pass
Sentinel-1 SAR	Unknown	Unknown	Unknown	Unknown

MODIS L2 SST	Pass	Incomplete on View Geometry	Pass and some N/A	Partial Pass
MODIS L3 Snow	Pass	Mostly pass; some N/A for derived snow product	Mostly N/A for derived snow product and/or in input L1/L2	Pass
Notes	[1] Saturation indicated by fill value + quality control bits (faulty L1B) [2] Corrections in ATBD, referenced in collection metadata in CMR			