Dataset Interoperability 2017-2018

Overview

The Dataset Interoperability Working Group (DIWG) strives to improve the interoperability of NASA-stewarded datasets by identifying best practices to structure datasets, how to check compliance of such datasets to metadata conventions, and how to adopt and, where necessary, extend those conventions. In 2013/14 we focused on dataset type interoperability, compliance checking, and extending CF conventions. In 2014/15 we launched a long-term focus on best practices for storing gridded and swath datasets. Our initial recommendations encompass both gridded (typically L3-L4) and swath (L2) data. In 2015/16 we laid the groundwork for more thorough consideration of swath datasets by setting up a GitHub repository to collect samples of NASA swath datasets and use them as examples when discussing swath-like data issues. In 2016/17 we examined several aspects of NASA swath dataset structures related to time and other coordinate types; how to unambiguously provide units for dimensionless quantities, preferred file name extensions for HDF5, netCDF-4, and HDF-EOS5 files; and provided invaluable input on the current NSF EarthCube proposal for the CF swath data encodings. In 2016/17, the ESDS Standards Office published RFC-028 with the dataset interoperability recommendations resulting from the previous years of our activities.

Proposer(s)

Aleksandar Jelenak
Peter Leonard

Draft Action Plan

Mission Statement

Improve interoperability of NASA-stewarded datasets by:

- recommending best practices for their structure and content;
- specifying how to check compliance with various conventions and software tools;
- modifying or extending relevant conventions.

Stakeholders

- Data producers: science teams, modelers.
- Library developers: HDF, netCDF.
- Vetting conventions: CF, HDF-EOS.
- Software tool developers.
- Other Working Groups:
  - Vertical Profile Data Visualization
  - Geospatial Web Services
  - Recommendations on Map Projections for Earth Observation Data
  - Documenting Data Gaps

Approach

- Regular monthly meetings.
- Mailing list for engaging the wider DIWG community not regularly participating in monthly meetings.
- Maintain the DIWG GitHub repository (https://github.com/diwg/diwg) to collect samples of NASA Earth Science datasets.
- Liaise with the NSF EarthCube project developing CF convention extensions for hierarchies (groups), group-aware metadata, and swath data encodings.
- Proactively seek common areas of interest with the stakeholder Working Groups.
- Establish communication with NASA representatives of commercial software vendors to keep them abreast of the outcomes (preferably in the form of RFCs) and seek feedback about data interoperability issues encountered with NASA Earth Science data products.

Outcomes, Deliverables

- Detailed specification of different swath feature types and their CF convention encodings.
- Recommendations how to achieve interoperability between comparable swath feature types in CF and HDF-EOS conventions.
- Another ESDS-endorsed RFC for adopted recommendations.
- Recommendations for CF Discrete Sampling Geometries data.
Proposal Discussion

- Use the comments area below to provide additional input, opinions, suggestions, etc.
- Edit this page to add relevant content, e.g. start adding stakeholders, activities, and deliverables to the "Draft Action Plan" section.
- Use the page comments to show your support for this proposal.
- Feel free to upload supporting documentation to the proposal page.
- "Watch" this page to be notified of updates.