LEVERAGING METADATA CONVENTIONS TO IMPROVE USABILITY OF AN *EASE-Grid 2.0* PASSIVE MICROWAVE DATA PRODUCT

Mary J. Brodzik
National Snow & Ice Data Center
CIRES/University of Colorado
449 UCB
Boulder, CO, 80309-0449 USA
Email: brodzik@nsidc.org

Molly A. Hardman
National Snow & Ice Data Center
CIRES/University of Colorado
449 UCB
Boulder, CO, 80309-0449 USA
Email: molly.hardman@nsidc.org

David G. Long
Department of Electrical and
Computer Engineering
Brigham Young University
Provo, UT 84602 USA
Email: long@ee.byu.edu

***Abstract*—Since 1978, a series of Earth-observing, satellite-borne passive microwave sensors has produced a rich record of microwave brightness temperatures. Passive microwave sensors can see through most clouds and collect measurements both day and night, which is especially useful in high latitudes during polar night. Passive microwave measurements are used to derive significant and meaningful climate records of many parameters, including the dramatic decline in Arctic sea ice. Earlier revisions of this significant climate record have been produced as flat, binary, gridded arrays with minimal or no file-level metadata and no machine-readable geolocation. Developed for many applications in polar regions, data were projected to polar azimuthal and global cylindrial projections that users found difficult to handle in standard mapping software packages. Funded by the NASA MEaSUREs program, we are using state-of-the-art image reconstruction techniques to produce a Calibrated Enhanced-Resolution Passive Microwave Equal-Area Scalable Earth Grid 2.0 Brightness Temperature (CETB) Earth System Data Record (ESDR) that leverages the improved *EASE-Grid 2.0* projection definitions and netCDF-CF metadata conventions to improve usability of the data products. We describe our approach to defining file-level metadata that is intelligible to standard software packages, including open source netCDF Operators (NCO) and Geospatial Data Abstraction Library (gdal), and the commercial ESRI ArcMap geospatial mapping tool.**

***Index Terms*—Passive microwave remote sensing, Software tools, Geophysical image processing, Geospatial analysis, Metadata.**

I. Introduction

Based on user feedback to the National Snow & Ice Data Center regarding the usability of historical gridded passive microwave data products, e.g. [1], [2], we began our project with clear goals for improving the transparency and usability of the Calibrated Enhanced-Resolution *EASE-Grid 2.0* Brightness Temperature (CETB) product [3], [4]. Users of historical versions of these data experienced difficult learning curves to properly configure geolocation parameters and were required to read long, human-readable, technical documents to determine simple parameters like array dimensions. Users were also frustrated by the loss of certain critical gridding and processing parameters that were never recorded in technical documentation.

This paper describes how we defined the CETB product to meet each of the following objectives: 1) Make the projected data understandable to software packages that assume the reference datum and projection ellipsoids are the same, 2) Define files with machine-readable metadata, 3) Use standard metadata conventions and best practices to incorporate file-level, machine- and human-readable contents, geolocation, processing and provenance metadata, 4) Include geolocation descriptions in multiple common description formats, and 5) Include sufficient standard geolocation information for GDAL to produce compliant GeoTIFF images that software packages like GDAL and ArcMap can easily import, understand and analyze.

When the metadata objectives are met, CETB files are much easier to use than has been the case with historical passive microwave data. In the past, a user wishing to compare gridded passive microwave data with other geolocated data was required to understand projection details and perform complicated ingest processing and reprojection transformations. With CETB files, geolocation and comparison with other data is now reduced to fast, simple standard operations (Fig. 1).

We present our experience as a case study, with the intention that our example may assist other data producers who wish to improve usability, interoperability and transparency of similar gridded data sets [6], [7].

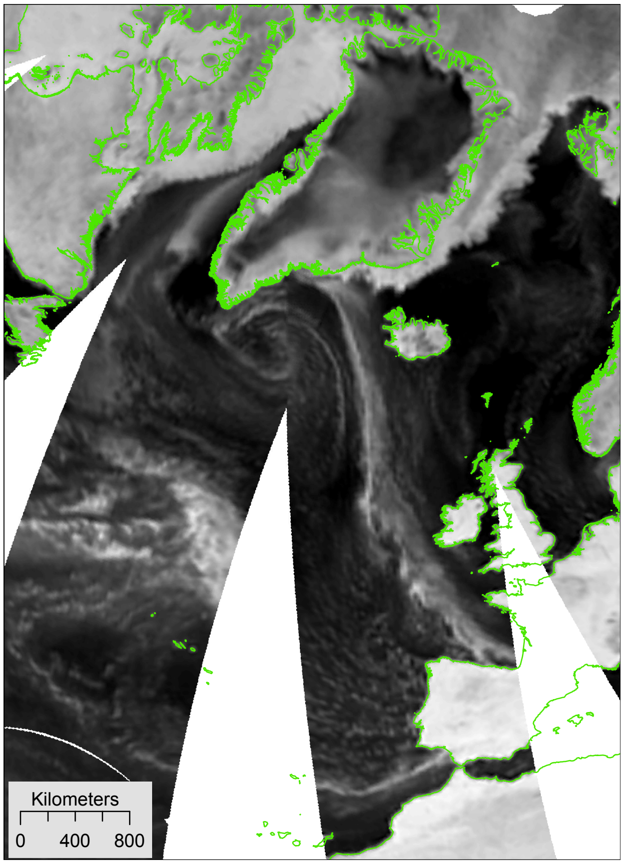

Fig. 1. CETB 3.125 km SSM/I 37 GHz horizonatlly-polarized evening overpass brightness temperature GeoTIFF, Jan. 5, 2003, zoomed to North Atlantic, easily overlaid in ArcMap with coastlines [5] with no special steps required.

II. Methods

A. Make the projected data understandable to software packages that assume the reference datum and projection ellipsoids are the same

In recent years, the GeoTIFF metadata standard [8] has emerged as a popular format for embedding geolocation information into image files. However, none of the historical gridded passive microwave data products could be formatted as GeoTIFFs without reprojection, because the cartographic projection ellipsoids did not match the WGS84 reference datum used for the source data geolocation [9].

To eliminate this problem in the CETB product, we use the *EASE-Grid 2.0* projection definitions. In the *EASE-Grid 2.0* definition papers, [10], [11], appendices included implementations of the forward and reverse map projection transformations and the corresponding reference Open Source Geospatial Foundation (OSGeo) PROJ.4 string definitions. By taking these steps to formally define *EASE-Grid 2.0*, we accepted accountability for this set of critical technical details upon which the CETB product depends.

Further, we have worked with the European Petroleum Survey Group (EPSG) to authoritatively define in the EPSG Registry (www.epsg-registry.org) each of the three equal-area projections that we use for CETB (*EPSG:6931-3*). Users of any software that understands proj.4 strings and/or EPSG ProjectedCRS codes can now correctly geolocate the CETB product images, without fussy and potentially error-prone transformations or reprojections needed for comparisons with other geographic information.

B. Define files with machine-readable metadata

Machine-readable file formats enable usability with common software tools and multiple languages. Users can read files immediately using the language and tools with which they are most comfortable, rather than spending a long time, sometimes days or weeks, understanding a binary file format and writing customized, potentially error-prone, software to read and interpret unfamiliar projected data arrays.

For the CETB data files, we considered HDF5 and netCDF4 formats. We selected netCDF4 because we were more comfortable with the conventions and terminology used in the netCDF4 application programming interface (API). Since netCDF4 is implemented with HDF5, we were also able to take advantage of powerful internal file compression capabilities of HDF5.

C. Use standard metadata conventions and best practices to incorporate file-level, machine- and human-readable contents, geolocation, processing and provenance metadata

We employed the Climate-Forecast (CF) metadata convention for encoding the geolocation information required to describe our projected data. The CF convention was flexible and adaptable for including customized fields to describe our processing parameters and algorithm settings.

As an ESDR, the CETB product uses as input a new, Level 2 (swath format) Fundamental Climate Data Record (FCDR), which includes a completely reprocessed historical SSM/I and SSMIS record, including new efforts to improve intersensor calibrations [12]. To ensure transparancy and reproducibility, we employ provenance metadata in file-level attributes in each CETB file, to record the specific Level 2 FCDR files used to derive the CETB image reconstruction.

The NASA Earth System Data Systems Dataset Interoperability Working Group (DIWG) Dataset Interoperability Recommendations for Earth Science [13] were recently approved by NASA and released as general best practices for data providers. We incorporated several of these recommendations in CETB files, including 1) choosing a minimum set of CF variable attributes that included variable bounds, fill values, packing convention details, and units, 2) specifying spatio-temporal bounds attributes, and 3) including a degenerate time dimension, to enable NCO concatenation operators to easily aggregate gridded data arrays into space-time "cubes".

While the CF conventions are powerful and flexible, the sheer number of attributes allowed by the convention can be extremely confusing to a data producer. The JPL Metadata Compliance Checker (MCC) (http://podaac.uat.jpl.nasa.gov/mcc) was an invaluable tool in verifying CF compliance. We enthusiastically recommend this web-based interface because it was simple to use and easy to understand violations and remedies, as we finalized the metadata contents of CETB data files.

D. Include geolocation descriptions in multiple common description formats

Feedback to NSIDC repeatedly underscores the difficulties that users encounter in identifying projected data coordinates. For the CETB file definitions, we generally adhered to the DRY (Don't Repeat Yourself) principle of software engineering, except for this important content. We deliberately violated the DRY tenet and included projection description information in the CF coordinate reference system (crs) variable attributes, as: 1) CF projection attributes, 2) PROJ.4 string, 3) EPSG ProjectedCRS code and EPSG Well-Known-Text (WKT) encoding, and 4) the "grid parameter definition" (.gpd) filename used by NSIDC mapx software, and 5) references to peer-reviewed papers describing EASE-Grid 2.0. For example, in a 25 km Northern Hemisphere CETB file, the crs variable contains the following, highly redundant, information:

```
char crs ;
  crs:grid_mapping_name =
    "lambert_azimuthal_equal_area" ;
  crs:longitude_of_projection_origin =
    0. ;
  crs:latitude_of_projection_origin =
    90. ;
  crs:false_easting = 0. ;
  crs:false_northing = 0. ;
  crs:semi_major_axis = 6378137. ;
  crs:inverse_flattening =
    298.257223563 ;
  crs:proj4text = "+proj=laea +lat_0=90
    +lon_0=0 +x_0=0 +y_0=0 +ellps=WGS84
    +datum=WGS84 +units=m" ;
  crs:srid =
    "urn:ogc:def:crs:EPSG::6931" ;
  crs:references = [links omitted]
  crs:crs_wkt =
    "PROJCRS[\"WGS 84 /
    NSIDC EASE-Grid 2.0 North\",..."
  crs:long_name = "EASE2_N25km" ;
```

While we take a risk in violating DRY that we may erroneously include inconsistent projection parameters in various interpretations, we believe this to be outweighed by improved comprehensibility for a potential user who may only understand one way for defining the projection.

E. Include sufficient standard geolocation information for GDAL to produce compliant GeoTIFF images that software packages like GDAL and ArcMap can easily import, understand and analyze

While adherence to conventions at a theoreical level is a necessary first step in improving usability of geolocated data files, the *de facto* demonstration of actual usability requires testing with current versions of target software.

In the following example, gdalinfo correctly interprets the CETB CF-compliant metadata coordinate system:

```
$ gdalinfo \
  NETCDF:"cetb_file.nc":TB_num_samples
Driver: Network Common Data Format
Files: cetb_file.nc
Size is 720, 720
Coordinate System is:
PROJCS["LAEA (WGS84) ",
  GEOGCS["WGS 84",
    DATUM["WGS_1984",
      SPHEROID["WGS 84",
        6378137,298.257223563,...
  AUTHORITY["EPSG","6931"]]...
```

The command-line GDAL utility *gdal_translate* can then be used to extract the variable `TB_num_samples` from the CETB .nc file as a legal GeoTIFF. Since we have used the *EASE-Grid 2.0* projection definitions, no transformations for reprojection or datum transform are necessary to overcome the problems that users have encountered in the past [9], [14]. Since we used netCDF and CF conventions to encode the projection definition and variable attributes, we simply instruct *gdal_translate* to interpret the file using NETCDF formatting:

```
$ gdal_translate -of GTiff -b 1 \
NETCDF:"cetb_file.nc":TB_num_samples \
TB_num_samples.tif
Input file size is 720, 720 \
0...10...20...30...40...50...\
60...70...80...90...100 - done.
```

The resulting GeoTIFF can be read and correctly reprojected without any further special instructions by any geospatial software that understands GeoTIFF, including GDAL utilities and the ESRI ArcMap geospatial analysis tool.

During our testing process, we have identified several issues reading CETB data files with various CF-enabled packages, including GDAL, Panoply and ESRI ArcMap, that required technical issues to be documented and reported to software developers. While the reporting process can be time-consuming and tedious, in each case it has increased our understanding of how the tools make use of the metadata we are including. We found the effort to be worthwhile, because all of the software vendors have been willing to work to remedy the problems we identified.

III. Conclusion

We have presented a case study of metadata design decisions used to produce the NASA MEaSUREs Calibrated Enhanced-Resolution Passive Microwave Equal-Area Scalable Earth Grid 2.0 Brightness Temperature (CETB) Earth System Data Record (ESDR). We explained our use of metadata objectives that leverage selected data formats and conventions to significantly improve the usability and interoperability of the CETB product, compared to similar earlier versions of gridded microwave data. Our users will benefit from these data design choices because the metadata significantly reduce the steps and technical knowledge required to correctly import, geolocate, analyze and reproject the CETB data product.

Acknowledgment

The authors would like to acknowledge Julia Collins at NSIDC for detailed traceability to CF and ACDD standards and initial work to define the CETB metadata template. This work has been funded by NASA MEaSUREs 2012. We use the Janus supercomputer, which is supported by the National Science Foundation (award number CNS-0821794) and the University of Colorado Boulder. The Janus supercomputer is a joint effort of the University of Colorado Boulder, the University of Colorado Denver and the National Center for Atmospheric Research.

References

[1] R. L. Armstrong, K. W. Knowles, M. J. Brodzik, and M. A. Hardman, "DMSP SSM/I–SSMIS Pathfinder Daily EASE-Grid Brightness Temperatures," National Snow and Ice Data Center, Boulder, CO USA, 1994, updated 2016, Digital Media.

[2] K. W. Knowles, M. H. Savoie, R. L. Armstrong, and M. J. Brodzik, "AMSR-E/Aqua Daily EASE-Grid Brightness Temperatures," National Snow and Ice Data Center, Boulder, CO USA, 2006, updated 2011, Digital Media.

[3] D. G. Long and M. J. Brodzik, "Optimum Image Formation for Spaceborne Microwave Radiometer Products," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2763–2779, May 2016.

[4] M. J. Brodzik, D. G. Long, M. A. Hardman, A. Paget, and R. L. Armstrong, "MEaSUREs Calibrated Ehanced-Resolution Passive Microwave Daily EASE-Grid 2.0 Brightness Temperature ESDR," National Snow and Ice Data Center, Boulder, CO USA, 2016, Digital Media.

[5] P. Wessel and W. H. F. Smith, "A Global, Self-Consistent, Hierarchical High–Resolution Geography Database," 2015, V2.3.4, http://www.soest.hawaii.edu/pwessel/gshhg, Digital Media.

[6] J. Gerring, "What Is a Case Study and What Is It Good For?" *Amer. Pol. Sci. Rev.*, vol. 98, no. 2, pp. 341–353, 2004.

[7] B. Flyvbjerg, "Five Misunderstandings About Case–Study Research," *Qualitative Inquiry*, vol. 12, no. 2, pp. 219–245, 2006.

[8] N. Ritter and M. Ruth, "Geotiff Format Specification, Geotiff Revision 1.0, Spec. Version 1.8.2," 1995, accessed 2017–01–07 at http://mac.mf3x3.com/GIS/GEOTIFF/geotiff_spec.pdf.

[9] T. Haran, "HDF-EOS vs. Geotiff: GIS Problems When Projection and Datum Spheroids Are Different," 2008, presentation at MODIS Science Team Meeting, May, 2008, Accessed 2011–11–27 at ftp://sidads.colorado.edu/pub/ppp/conf_ppp/Haran/HDF-EOS_vs._GeoTIFF__GIS_Problems_when_Projection_and_Datum_Spheroids_are_Different.pdf.

[10] M. J. Brodzik, B. Billingsley, T. Haran, B. Raup, and M. H. Savoie, "EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets," *ISPRS Int. J. Geo-Inf.*, vol. 1, pp. 32–45, 2012.

[11] ——, "Correction: Brodzik, M.J., *et al.* EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets, *ISPRS Int. J. Geo-Inf.* 2012, 1, 32-45," *ISPRS Int. J. Geo-Inf.*, vol. 3, pp. 1154–1156, 2014.

[12] H. Semunegus, W. Berg, J. J. Bates, K. R. Knapp, and C. Kummerow, "An Extended and Improved Special Sensor Microwave Imager (SSM/I) Period of Record," *Journal of Applied Met. and Climatology*, vol. 49, pp. 424–436, 2010.

[13] C. S. Zender, P. J. T. Leonard, and *et al.*, "Dataset Interoperability Recommendations for Earth Science," International Journal of Remote Sensing, Technical Note ESDS–RFC–028v1.1, July 2016.

[14] B. Billingsley and T. Haran, "Using GeoTIFFs for Data Sharing: Limitations and Solutions," 2008, Eos Trans. AGU, 89(53), Fall Meeting Suppl., Abstract IN23B–1078, Accessed 2011–11–27 at ftp://sidads.colorado.edu/pub/ppp/conf_ppp/Billingsley/Using_GeoTIFFs_for_Data_Sharing_Limitations_and_Solutions.pdf.