

Splinter Session 11: Satellite Data File Formats and Tools for Easy Science Exploitation

Co-Chairs: Carsten Brockmann (Brockman Consult), Bryan Franz (NASA GSFC) and Simon Elliot (EUMETSAT)

The objective of this splinter session is to deliver consensus recommendations on data file content and format, meta-data, processing and analysis tools. These consensus recommendations shall be agreed by a larger community, as represented at the IOCS conference. Hence, progress compared to the current status shall be tangible progress and concrete statements. The splinter session was organised around three themes. This report presents a summary of the discussions and recommendations formulated for each of the three themes.

1. Data File Content and Formats

NASA OBPG (Ocean Biology Progressing Group) as well as ESA/Eumetsat with Sentinel 3 are currently moving towards a file format wherein data are stored in NetCDF4 format and utilize the Climate and Forecast metadata convention (NetCDF/CF). A primary reason for this transition is to improve interoperability between missions, reduce software maintenance requirements on existing data visualization tools, and improve compatibility with a host of third party tools and protocols (e.g., THREDDS/OpenDAP). Specifically, NASA OBPG is gradually migrating its Level 2 data from HDF4 into NetCDF/CF. The ESA Sentinel 3 format is a package where the image as well as ancillary data are stored in NetCDF4 files, with one file per variable, and an XML header file with all meta data. This follows the XDFU standard. KARI (Korea) is using HDF5 for GOCI data. NetCDF4 is a wrapper around HDF5, so transition from HDF5 to NetCDF4 may be feasible. Time series are also support by NetCDF4. Questions still need to be resolved as to how Level 3 data can be stored in an optimal way in NetCDF/CF and how the spatial and temporal binning needs to be adapted. Further, the CF convention does not fully cover the requirements for ocean colour and extension of the convention is required. This should be coordinated among the space agencies, an activity which could be led by IOCCG. Although there was an overall large agreement with NetCDF4/CF as common format, it was mentioned that some operational users requiring real time or near real time data require a format where satellite, in-situ and model data can be integrated.

Recommendation: The splinter session proposes to recommend to all Space Agencies to adopt netCDF4/CF for their ocean colour data.

2. Data processing, analysis and exploitation tools

Flexible and user –friendly data processing, analysis and exploitation tools have become critical to ocean color research with the increased availability of disparate ocean colour data; e.g, SeaWiFS, MODIS, MERIS, GOCI, and VIIRS. The splinter session has clearly expressed that the scope of required tools has to support visualization and validation, processing to higher-level products, modification and development of new processing algorithms, and

visualization must include all data product levels as well as analysis of single products up to long time series and full mission data sets. Most agencies operating ocean colour satellites today provide tools to its users which partly fulfil these requirements. The new releases of SeaDAS7 / BEAM 4.11 are a first step toward a common tool that is made available to users across agencies and works with (most) ocean colour products.

A key to success for the acceptance of ocean colour products is the full transparency of the processing algorithms through an open source release of the processing software. This was always the case for the SeaDAS software of NASA, and the exploitation of MERIS has been significantly improved with the availability of at least the Level 2 code in the ESA ODESA tool. The splinter group agreed that a multi-sensor approach as it is realised in SeaDAS is a significant plus and should be further pursued beyond the currently supported sensors. The demand for suitable, fast and robust data visualisation and analysis tools are very important for all users in order to quickly assess data quality and analyse single products as well as large data sets. The future development of the tools shall strive to support as many different sensors as possible and allow easy multi-sensor analysis with proper mapping and cross calibration functions. The IT technology of the tools shall be up-to-date in order to minimize processing times and thus allow for fast processing on very large datasets.

The splinter group emphasised that continuity of tools is important in order to insure user's developments and investments which build on current tools.

Finally it was discussed that the currently available tools provide a rich set of functionalities which are often not known by users. Data exploitation is thus not optimally done although proper tools are available. This has been discovered during training courses and forum discussions.

Recommendation: The splinter session recommends to the Space Agencies to further support the existing line of tools, namely SeaDAS, BEAM and ODESA, and to continue them for future sensors. This shall include further development as well as training of users.

3. Data distribution

Ocean colour research is often focused on global applications or long-term regional time-series analyses leading to large data volume requirements. Data users require the ability to access and download data through scriptable (batch download) mechanisms (e.g., through established ftp and http protocols). Data access methods should also support regional extraction to minimize the amount of data that must be transferred to satisfy research requirements. More targeted, direct access methods of data access such as THREDDS/OpenDAP should be exploited. The splinter session considers an efficient data distribution as critical for the success of any ocean colour mission.

EUMETSAT presented its plans for dissemination of Sentinel 3 data via the EUMETCast system. EUMETCast is a relay system where data are broadcasted via commercial telecommunication satellites. The system is in use since many years and provides high availability, maturity/stability, reliability, reproducible results, monitoring and control functionality, and maintainability. EUMETSAT currently distributes MODIS/Terra through EUMETCast and planning to include VIIRS, MERSI and OCM (both currently under negotiation), OLCI and potentially SGLI. The splinter sessions acknowledges the distribution via this medium and sees a great potential for efficient NRT distribution in it, but also stresses the point that continuation of traditional access means such as ftp is important. EUMETSAT is currently detailing the Sentinel 3 data dissemination through EUMETCast. Issues such as bandwidth need to be dimensioned properly. Concrete and justified requirements formulated by the international ocean colour community are required and shall be forwarded to EUMETSAT.

Recommendation: The space agencies should support large volume, batch data access and download (e.g., through established means such as ftp/http), as well as more targeted access through protocols such as THREDDS/OpenDAP.

Recommendation: Sentinel 3 data will be available through EUMETCAST. The ocean colour community shall provide concrete and justified requirements to EUMETSAT so that the distribution means can be properly dimensioned.