CCI Data Engineering Working Group (DEWG) Meeting #1

8th CCI Collocation, 23rd March 2018

Alison Waterfall (STFC)
DEWG – Topics

1. Issues, Concerns & Needs – Updating the WG Catalogue.
2. CCI Data Standards.
3. CCI Data Standards In Action.
4. EU GDPR.

16 participants spanning new & existing ECVs.
DEWG – Why?

• CCI community **supporting you**.
• **Share** your data engineering challenges
• **Cultivate common** CCI benefits, opportunities & infra downstream -
  • E.g CCI Data Standards, CCI Open Data Portal, CCI Toolbox
• Also **ESA listening** to your concerns.
DEWG – Terms of Reference

OBJECTIVE

To ensure maximum usability of the datasets produced within CCI, and cultivate tools for their access, discovery and manipulation, through common CCI data standards.

MEMBERSHIP

The DEWG will include one representative from each CCI project, plus one representative from ESA who will be Chair of the working group, plus two members from ESA involved in ESA Climate Office Knowledge Exchange activities. Additional persons will be invited to attend meetings at the request of the Chair. The Project Manager of each CCI project will ensure participation of their project in the DEWG.
AIMS

1. To provide a **forum to manage data standards** across CCI projects
2. To ensure that suitable existing standards are adhered to within CCI data products
3. To agree common format and metadata requirements for CCI data products
4. To enable CCI projects to **feed back and iterate** on proposed data standards
5. To **share expertise** on data engineering, and to enhance coordination and communication to avoid duplication of data engineering effort.
6. To provide **ESA** with insight into the data engineering **needs and challenges** being experienced by CCI projects.

To achieve these aims, members of the DEWG will

- Address compliance of all ECV products to the CCI Data Standard.
- Provide details to the DEWG on cases where their CCI data products do not conform to CCI Data Standards. In all cases where the CCI Data Standards are not conformed, the respective CCI project team will propose solutions to the DEWG for discussion and implementation.
DEWG – Communications

- **Quarterly Meetings**; Six-monthly face-to-face (CCI Colocation & CMUG)
- **DEWG Chair** – Ed Pechorro (Chair) reaching out directly to all projects
- **CCI Data Standards Support**
  CCI Open Data Portal Project – Victoria Bennett & Alison Waterfall
- **Mailing list** - CCI.Data.Engineering.Working.Group@esa.int
- **Webpage** - cci.esa.int/working-groups
- **File repository**
  https://esabox.esa.int/owncloud/index.php/s/S0ZL0KN9K9pjMeX
- **Coming Soon** – Github repository. ESA Climate Office championing Open Source.
SEWG & DSWG report on data and system engineering support needs, challenges and nice-to-haves

• Patterns across ECVs

• New 2-year cycle starts – **2018 to 2020**.

• Updated for existing CCI ECVs

• New CCI+ ECV projects join in.

<table>
<thead>
<tr>
<th>Needs &amp; Challenges</th>
<th>Aerosol</th>
<th>Cloud</th>
<th>OMI</th>
<th>OMI+</th>
<th>Glaciers</th>
<th>AIS</th>
<th>Land Cover</th>
<th>Ocean</th>
<th>Sea Ice</th>
<th>Sea Level</th>
<th>SST</th>
<th>Soil moisture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited Confidence in Data Standard Maturity &amp; Process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Characterisation of Non-NetCDF Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Resource for Compliance Checking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Support with Climate Data Taxonomies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Uncertainties Nomenclature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Upstream Activity on Data Standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Support on High Grade Visualisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Coordination on Visualisation &amp; Web Presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited User Metrics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Web Update Support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Status Update</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Audience Scope of Material</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Best Practice on Standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Sharing of Design &amp; Code Resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Modes of Collaboration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Common Tooling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Document Preservation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Clarity on Operational Linkages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Information Desired Regarding ESA Climate Office</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More Information Desired Regarding GCOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSWG &amp; SEWG Should Remain Close or Integrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science Lead Contact Is Important</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCMU Contact Is Important</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meetings Focus Is Important</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wiki &amp; Mailing List Should Be Community Focused</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy Access To Resources Is Important</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Recognition of Decentralised Glaciers Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Obs4MIPs – Limited Status News
- Obs4MIPs – Limited Confidence
- Obs4MIPs – Limited Speed of Progress
- Obs4MIPs – Common Upstream Organisation Desired
- Obs4MIPs – Collaboration Desired
- Limited Solutions for Big Data Processing
- Limited Solutions for Big Data Access
- Limited Burden Sharing Solutions in Acquiring Data
- Limited Support with Cataloguing
- Further Masts Desired
- Limited Support with DOIs
- Limited Support with Quality Control
- Limited Support with Versioning
- Limited Commonality on Data Management Practice
- Limited Data Harmonisation
CCI Data Standards – Status

• **v1.2**
  • Current baseline
  • Agreed by all *existing* CCI ECV projects

• **v2.0 proposed** Following study & feedback CCI projects.
  
  • **Existing projects** -
  • Proposed to *existing* CCI ECV projects at 8th CCI Colocation.
  • v2.0 is a final baseline across *existing* CCI ECV projects
  • Feedback please by **10th April 2018** from existing CCI ECV projects

  • **New CCI+ ECV projects** – Your feedback very welcome!
  • Once project starts, you enter Standards proposal cycle.

• **Quarterly update cycles.**
CCI Data Standards - Rationale

- Existing CCI data standards help provide consistency in formats, metadata and product names
  - Important for development of common tools (e.g. CCI Open Data Portal and CCI Toolbox)
  - Heterogeneous data and user communities: format can’t be too prescriptive
  - Takes account of existing initiatives (e.g. CF-conventions)
- Format has been updated for CCI+ KO
  - Takes account experience from the first phase
  - Will allow datasets to be fully integrated into the CCI Portal and Toolbox
CCI Data Standards - Overview

• **Data should be:**
  • Produced in netCDF-4 (classic) format
  • Conform with the CF (Climate and Forecasting) convention (currently version 1.7)
  • Use CF standard names for the main variables
  • Include the global attributes listed in document

• **Where users require other formats (e.g. GeoTiff, shape files)**
  • Should comply as far as possible (e.g. filenames, metadata)
  • Standardised netcdf products should also be produced
CCI Data Standards In Action

CCI Viewer
Climate data viewer of the ESA Climate Change Initiative
CCI Data Standards In Action

CCI Search
Climate data search interface for the ESA Climate Change Initiative

Processing Level (10)
Level 3 (10)
ECV: ocean colour
Frequency: month
Processing Level: Level 3

10 results

ESA Ocean Colour Climate Change Initiative (Ocean_Colour_cci): Global remote sensing reflectance gridded on a sinusoidal projection, Version 2.0

The ESA Ocean Colour CCI project has produced global level 3 binned multi-sensor time-series of satellite ocean-colour data with a particular focus for use in climate studies. This dataset contains the Version 2.0 Remote Sensing Reflectance product on a sinusoidal projection at approximately 4 km spatial resolution and at a number of time resolutions (daily, 5-day, 8-day and monthly composites). Values for remote sensing reflectance at the sea surface are provided for the standard SeaWiFS wavelengths (412, 443, 490, 510, 532, 670nm) with pixel-by-pixel uncertainty estimates for each wavelength. These are merged products based on SeaWiFS, MERIS and Aqua-MODIS data. Note, these data are also contained within the ‘All Products’ dataset. This data product is on a sinusoidal equal-area grid projection, matching the NASA standard level 3 binned projection. The default number of latitude rows is 4320, which results in a vertical bin cell size of approximately 4 km. The number of longitude columns varies according to the latitude, which permits the equal area property. Unlike the NASA format, where the bin cells that do not contain any data are omitted, the CCI format retains all cells and simply marks empty cells with a NetCDF fill value. (A separate dataset is also available for data on a geographic projection).

This data has been produced by the ESA Ocean Colour CCI project and provided to CEDA in the context of the ESA CCI Open Data Portal project.
DEWG Links & Actions

- **C3S CDS**
  - John Swinton (OC CCI)
- **Engineering Knowledge Sharing Resources**
  - Thomas Storm (Fire CCI)
- **Fiduceo**
  - Tom Block (SST CCI)
- **Uncertainties**
  - Caroline Poulsen (Cloud CCI)
Thank you

8th CCI Collocation, 22nd March 2018

Alison Waterfall (STFC)