

# PML

Plymouth Marine  
Laboratory

Listen to the ocean

## GloboLakes WP2: data processing and distribution

Steve Groom  
14<sup>th</sup> January 2014

[sbg@pml.ac.uk](mailto:sbg@pml.ac.uk)





PML lead, also UoReading and UoStirling

Aim: To retrieve functional indicators from archived and rolling satellite datasets & widely disseminate results

- 2.1 Develop highly automated data processing chain
- 2.2 Process the entire MERIS 300 m archive for the ~1000 target lakes + ARCLakes processing
- 2.3 Run a short time delay monitoring capability as a prime input to the Global Lake Observatory
- 2.4 Extend methods to ESA/ GMES Sentinel 3 OLCI & SLSTR & opportunistically to Sentinel 2 MSI.
- 2.4 Provide data via the web & ftp

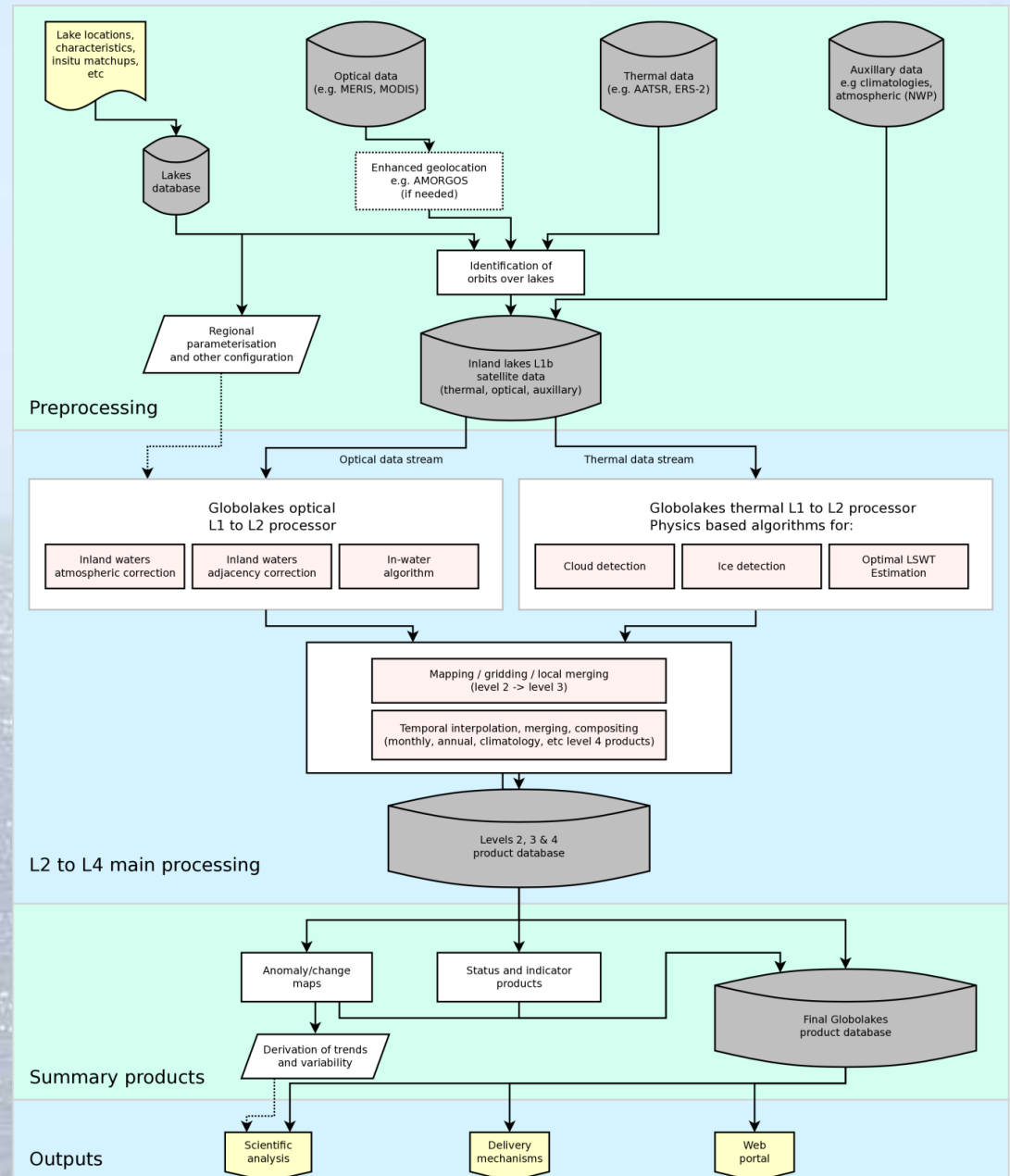




- Requirements (inputs)
  - From ESA/NASA/(BC?), complete global archive of MERIS level 1 FR data
  - From ESA NRT stream of Sentinel 2 and Sentinel 3 data
  - From WP1 and Diversity II atmospheric correction and in water algorithm ensemble processing methodology
  - Processing framework (possibly in BEAM)
  - From WP3 list of lakes to process (however, maybe easier to process everything – just then need worry about QC)
- Deliverables (outputs)
  - MERIS: time series of water quality values
  - Sentinel 3, OLCI: continuation of MERIS time series ~2015-2017; NRT lake monitoring
  - Sentinel 2, MSI: feasibility of MSI for lake monitoring



- Existing ocean colour processing systems are being designed to make changing processing chain easy
- Processing done in stages
  - Generate products, map products, annotate, archive...
- Batch-processing oriented structure
  - Many small jobs (~10,000 / day)





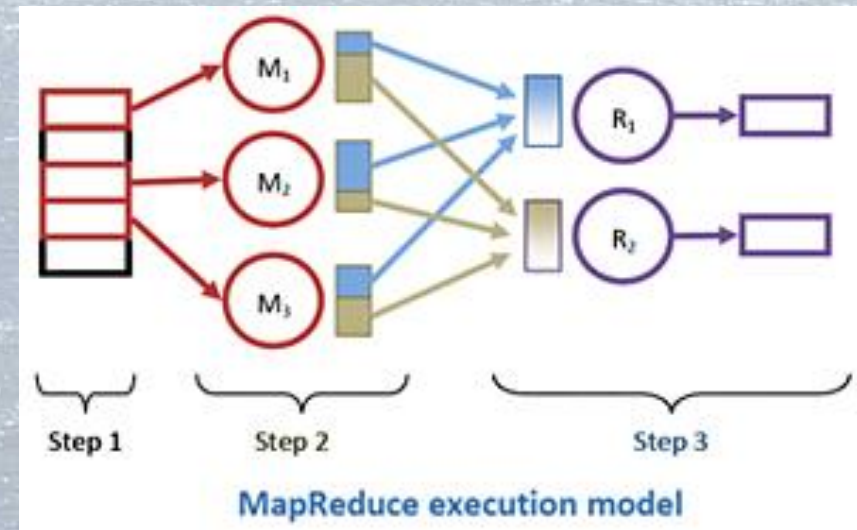
- The GloboLakes workflow will likely be based on the ESA Ocean Colour Climate Change Initiative (OC-CCI) processing chain
  - Currently 4-km
  - 1-km in trial
  - 0.3 km ultimate goal
- Algorithm in the processing chain will be based on WP1 research and also hopefully experience in Diversity II
- A version 1 could be to implement the Diversity II system and then run over 1000 lakes

- PML Commodity Linux cluster
  - PML's existing systems:
    - 48 x Dell 1U compute node  
(16GB RAM, 4 core, small disk, 1Gb NIC)
    - + 36 x Dell R200 class machines  
(8GB RAM, 2-4 core, small disk, 1Gb NIC)
- PML disk storage nearly 2 petabytes of on-line RAID6 storage
- GloboLakes will pay for an additional:
  - 16 grid nodes
  - ~100TB + ???TB 2015



The pretty half of the server room

- OC CCI processing will be undertaken on a Brockmann Consult Calvalus system
  - 23 x supermicro 1U compute nodes 16GB, 4 core, 10TB local storage (organised as HADOOP HDFS)
  - Processing software in BEAM framework will be optimised for use with Calvalus processor
- Note GloboLakes wont use the OC CCI processor; instead the PML cluster could be configured as a Calvalus system



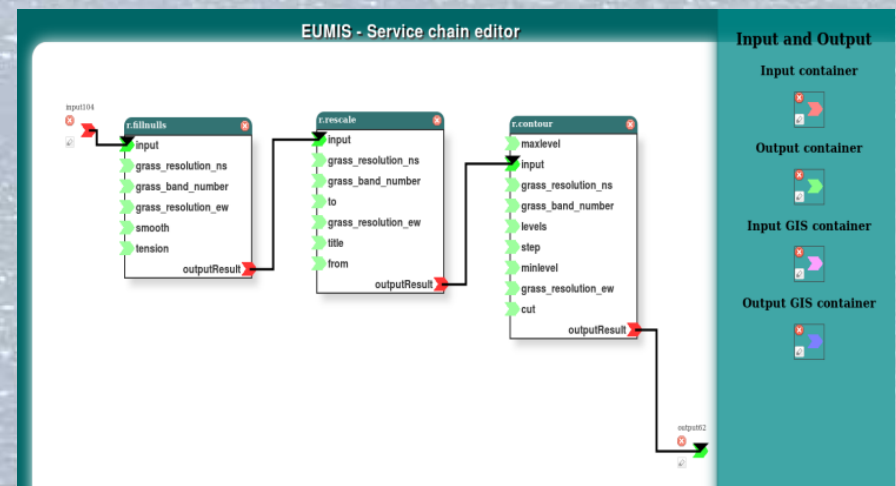
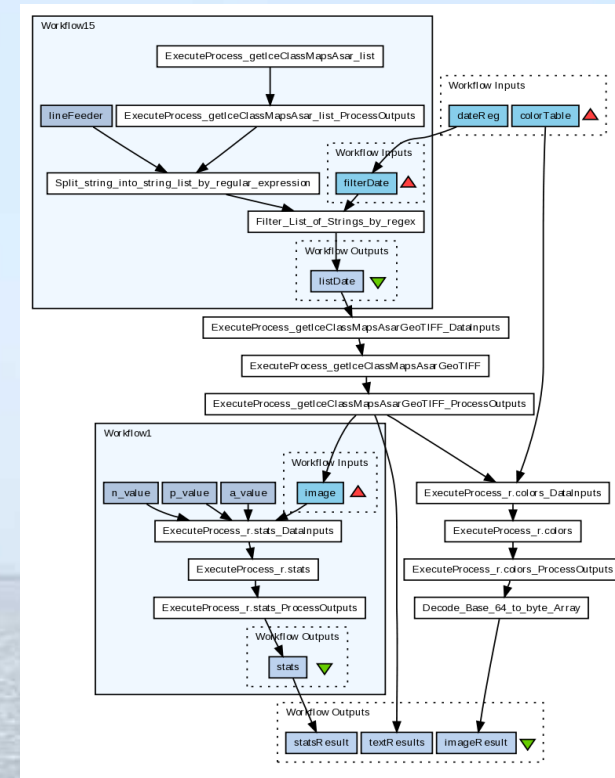


- CEMS hardware at Harwell
  - Elastic cloud of virtual machines running on JASMIN-CEMS infrastructure (undergoing expansion)
  - 6 petabytes local storage
  - Particularly relevant since S2 and S3 data available locally



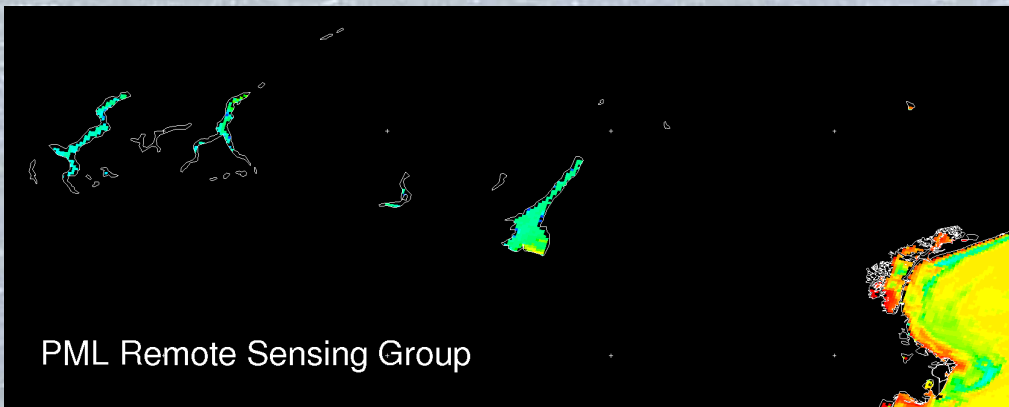
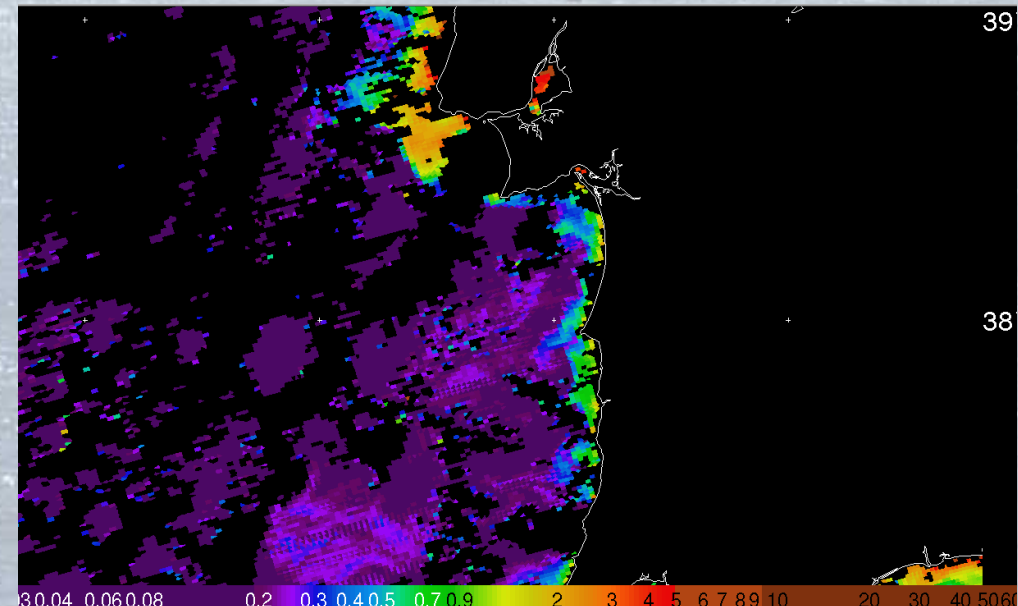


- Investigating web-based processing in various FP7 projects
  - Processing components made available via WPS standard
  - Users assemble components into a chain of processing
  - Semantically aware, so one can't accidentally plug a chlorophyll data source into a temperature conversion WPS component
  - Aiming for arbitrary web-based "GIS" functionality
  - Possibility of users doing their own lake data processing





- Essentially same system as 2.2
- Focus on operational monitoring for water quality monitoring / comparison with in situ etc.
- Scope i.e. UK only or, monitored lakes or all 1000 lakes?
- Was to be based initially on MERIS but contact with Envisat lost on 8 April 2012
- Sentinel 3 OLCI due for launch in 2015?
  - Data available in 2016?

MERIS algal-2 300m, 21<sup>st</sup> Feb. 2012, 09:39MODIS 500m, 13<sup>th</sup> Jan. 2014, 13:15

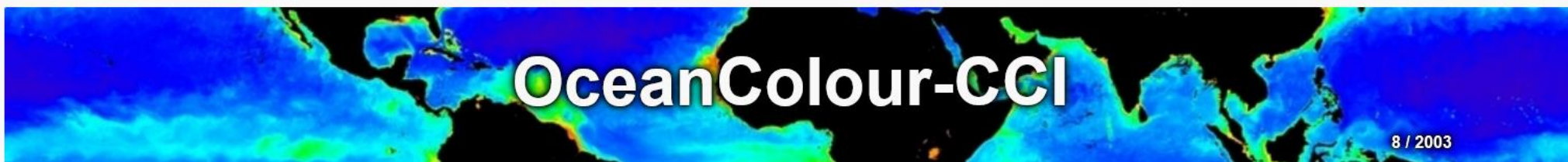


- Main focus on Sentinel 3 OLCI
  - Will provide main sensors for observation of lakes towards end of GloboLakes
  - MERIS-like: 300m resolution;
  - Global observation (MERIS was more regional); extra bands; tilt to west
  - Two sensors due for launch: observation every 1-2 days
  - Launch in 2015? Over one year later than originally planned in GloboLakes
  - Data availability, processing levels are issues.
- Also interest in Sentinel 2 MSI
  - Higher resolution; primarily a land instrument but great potential for inland waters; 10-60m resolution; 5 day repeat with two satellites
  - First instrument launch due 2014?
- Interaction with ESA OC CCI, EC FP7 GLaSS (?) or HYROC (?)



- Web site based on NEODAAS browser, ChloroGIN Lakes and Ocean Colour CCI portal
  - Best for looking at NRT data
  - Browsing images
- FTP service
  - Most frequently requested in CCI user requirements survey
  - Best for long time series
- OPeNDAP NetCDF subsetting service
- OGC based solutions

- Aim to provide a portal that has different access options



Composite Browser



Web GIS Portal



OPeNDAP



FTP

## About

This site provides satellite observations of ocean colour, focusing on the Ocean Colour Climate Change Initiative project ([see more about this project](#)).

## Contact us

Contact us via the [OC CCI processing email](#).

## Useful links

- [NASA OceanColor](#)
- [International Ocean-Colour Coordinating Group](#)

## Documentation



- Individual lakes (or groups of lakes)

## ChloroGIN Lakes

Earth Africa Europe Lakes Latin America Contact Partners Documents In Situ Related project

This web page is a demonstrator of the end-to-end application proposed in the Draft Task Sheet for the "GEO 2012-2015 Work Plan: Global Inland and Near-Coastal Water Quality Information System". It provides access to pre-operational monitoring of Lake Balaton, Hungary and a few lakes in northern England, Scotland and Northern Ireland, UK. Other lakes will be added in time.

Select your area of interest by clicking on the map. Available datasets will be highlighted and displayed in the list below. View the dataset by clicking on the colour coded "View" link (the dataset will be opened in the provider's own portal).

NOTE: Due to problems with ENVISAT images from the MERIS sensor will not be available.

Selection Date (dd-mm-yyyy):

Today

AVAILABLE AREAS



## NEODAAS

Logout Hello sbg  
System status: ● ● ●  
Introduction [Area: my ]  
Balaton lake larger geographi  
Getting started  
Demonstrations  
Current summary  
Help/FAQ  
Contact us  
+/- Navigate data  
Select area  
Select date  
+/- Current data  
FlexiView multi-products  
+/- Sensor  
+/- MERIS 300m  
Chlorophyll (algal\_1)  
Chlorophyll (algal\_2)  
RGB  
Radiance (nLw 560)  
Aerosol opt. thick  
Yellow substance  
SPM  
Vegetation (toa)  
Vegetation (boa)  
Chlorophyll (oc3)  
Kd 490 (SeaDAS)  
Aerosol opt. thick 86  
Reflectance 413nm  
Reflectance 443nm  
Reflectance 490nm  
Reflectance 510nm  
Reflectance 560nm  
Reflectance 620nm  
Reflectance 665nm  
Reflectance 681nm  
+/- Comparisons  
User preferences

⏪ Back one week    ⏩ Back one day

**MERIS 300m**  
Thursday, 5 April 2012

⏩ Forward one day    ⏪ Forward one week    🕒 Go to Today

Chlorophyll (algal_2) image nearest to 5th Apr 2012	This week's Chlorophyll (algal_2)		Chlorophyll (algal_2) Composite
<p>28 Days Ago (5th Apr 2012) 09:30 (91% valid pixels)</p> <p style="font-size: 0.8em;">PML Remote Sensing Group</p> <p style="text-align: center; font-weight: bold;">You are viewing scenes:</p> <p style="text-align: center; background-color: #0056b3; color: white; padding: 2px;">Switch to Composites</p>	<p>29 Days Ago (4th Apr 2012) No images</p> <div style="background-color: #ccc; padding: 20px; text-align: center; font-size: 1.5em;">no data available</div>	<p>30 Days Ago (3rd Apr 2012) 09:06 (87% valid pixels)</p> <p style="font-size: 0.8em;">PML Remote Sensing Group</p> <p style="text-align: center;">1st Apr 2012 No images</p> <div style="background-color: #ccc; padding: 20px; text-align: center; font-size: 1.5em;">no data available</div>	<p>From 30th Mar 2012 through to 5th Apr 2012</p> <p style="font-size: 0.8em;">PML Remote Sensing Group</p>
<p>31st Mar 2012 09:16 (84% valid pixels)</p> <p style="font-size: 0.8em;">PML Remote Sensing Group</p>	<p>30th Mar 2012 09:50 (33% valid pixels)</p> <p style="font-size: 0.8em;">PML Remote Sensing Group</p>		

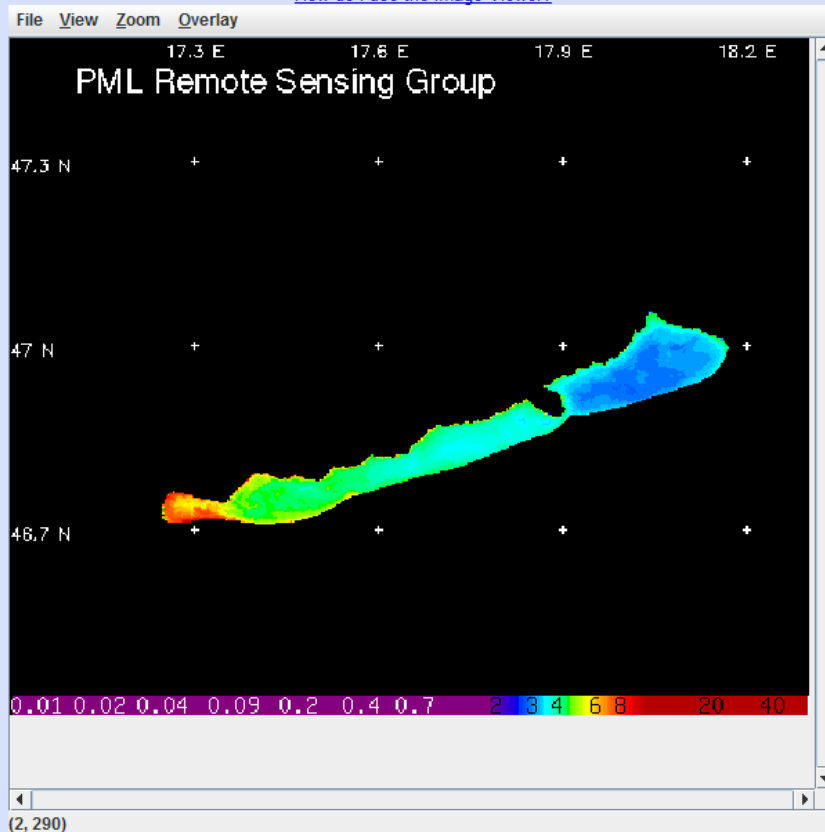
Units: mg m<sup>-3</sup>  
For additional information on this product, please check the [help/FAQ](#).  
[Permalink](#)





## MERIS algal\_2 (2012-04-05 09:30)

[How do I use the Image Viewer?](#)



Interactive Str... [X]

Reset

Values [X]

Lat/Lon Units: Decimal Degrees [v]

X: 0001

Y: 0241

Latitude: 046.8517

Longitude: 017.0027

Digital Number: 000

Real World Val: 000.0100

Selected Pixel: [black bar]

## MERIS algal\_2 (2012-04-05 09:30)

[Download data file](#)  
[View without java](#)  
[Download netCDF](#)

Custom Zoom [X]

120% Reset Fit

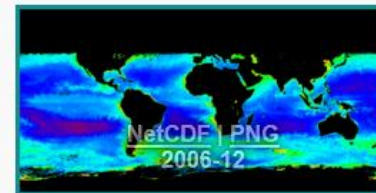
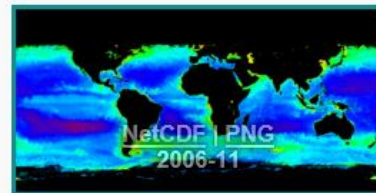
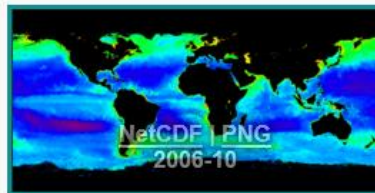
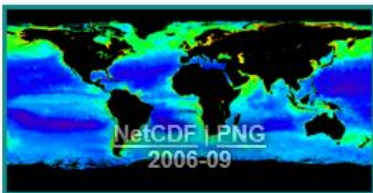
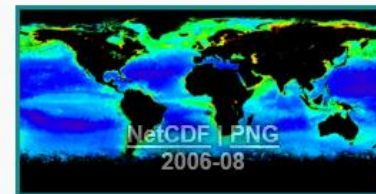
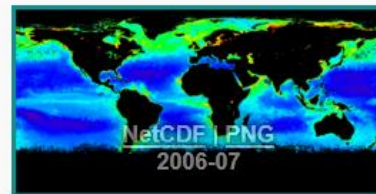
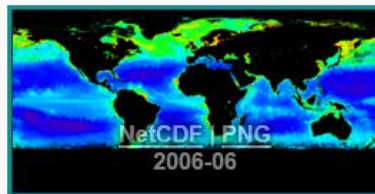
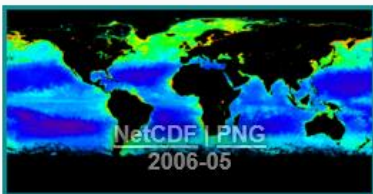
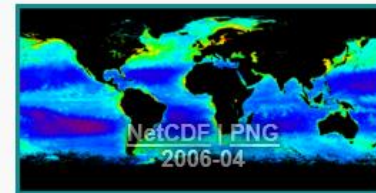
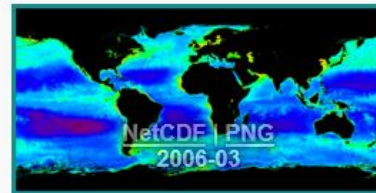
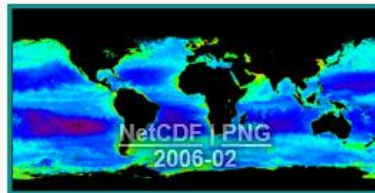
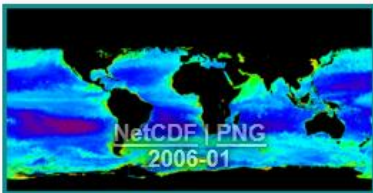
- Global coverage including ALL lakes?

## OceanColour-CCI composites browser

Viewing: (OC\_PRODUCTS) All Ocean Colour products

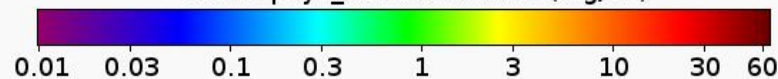
As  composites, from the date  to the date  with a maximum of  results per page.

[Newer >](#)



[Palette values \(CSV\)](#)

Chlorophyll\_a concentration (mg/m<sup>3</sup>)





- Fundamental data service behind all modern mapping web portals
  - Increasingly widely adopted
  - Format popularised by similar systems, such as Google Maps
- Open Geospatial Consortium Web Map, Coverage and Feature services
  - Modern version of OPeNDAP
  - Geographic addressing
  - Supply map tiles in image (WMS) or data (WCS) format, or point data (WFS)



Browser window: <http://www.dev.oceancolour.org/portal/#> | ESA-OC\_CCI Visualisation P... x

File Edit View Favorites Tools Help

Suggested Sites Project Balances Web Slice Gallery Google Maps

Layers **Data Layers**

Models Reference Options

chlor a

View Date: 01-04-2013 | Quick Region: Choose a Region

**Data Analysis** Analysis

Data Selection Tools

Selection Analysis Data Export

Graphing

Gallery:

Data

Make sure to select the correct layer:  
chlor a

Area and Time of Interest

Time Range: Select a Range

New Range Hide Range Rename Range

Delete Range

Date: yyyy-mm-dd to yyyy-mm-dd

Bbox: -82.48213,47.57402,-82

Generate Graph

**Hovmoller On Of Chlor A**

Time

1998

2000

2002

2004

2006

2008

2010

2012

Longitude

-87.4375 -86.6875

2.41

1.35

0.75

0.41

0.23

Variable Selecti... ?

chlor a

997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 20

100%



- “Global Earth Observation for integrated water resource assessment”
- New EC FP7 project on the Water Cycle (all aspects; water quality is a small part)
- 4 year, started 1 Jan 2014
- Main PMLA (PML’s SME) focus is developing a Water Cycle Integrator
- But some WQ work focussing on Estonian lakes with Tiit Kutser and Estonian national authorities



- Processing framework under development and based on OC CCI project
- Through collaboration with Diversity II hope to implement results from that project as a starting point







- **Cesium – 3D viewer**
- Library for creating 3D globes and 2D maps in a web browser
- Open source under the Apache 2.0 license free for commercial or non-commercial use
- Google Earth look and feel but within a browser with no requirement to install plugins or install any additional software.
- Support for OGC WMS with WFS pencilled in for support later this year; KML is also supported.
- Support for a variety of vector layers and formats.
- Under rapid, active development with a thriving developer community and excellent documentation





Layers

View Date: 05-11-2012

**Data Layers**

- Operational Layers
  - NOAA\_Coral\_Reef\_Watch\_Operational\_Twice-Weekly\_Near-Real-Time\_Global\_50km\_Satellite\_Coral\_Bleaching
  - v wind @ isobaric
  - sea surface temperature
  - MRCs\_Ecosystem\_Variables
    - Chlorophyll a
    - Nitrate concentration
- Reference Layers
  - The Black Sea (KML)
  - AMT19 Cruise Track
  - AMT17 Cruise Track
  - AMT16 Cruise Track
  - AMT15 Cruise Track
  - AMT14 Cruise Track
  - AMT13 Cruise Track
  - AMT12 Cruise Track

Scalebar - sea surface temperature ? - x

34  
25  
16  
7  
-2

sea\_surface\_temperature (Celsius)

Maximum Value: 34

Recalculate Scale

Logarithmic Scale

Reset Scale

Minimum Value: -2

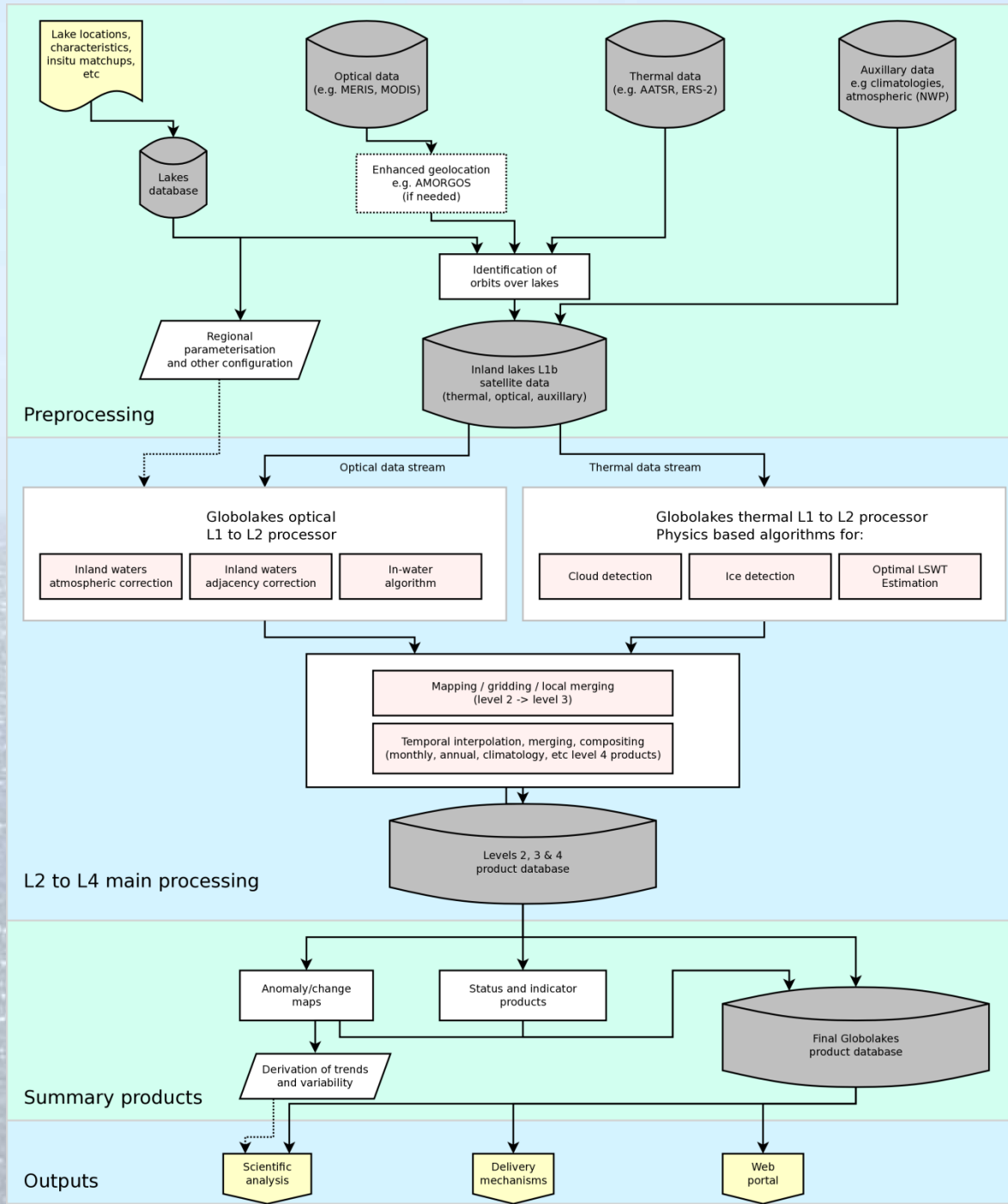
Base Layer: GEBCO

Quick Region: Choose a Region

Metadata - sea surface temperature - x

Source:  
Name: sea surface temperature  
BoundingBox: 85.0N, 179.5E, -80.0S, -180.0W  
Date Range: 28-11-2000 to 05-11-2012  
Abstract: Twice-weekly global 50km satellite nighttime sea surface temperature





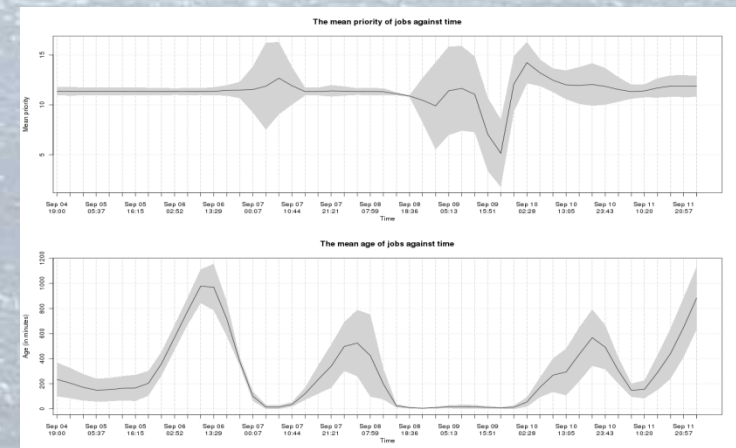
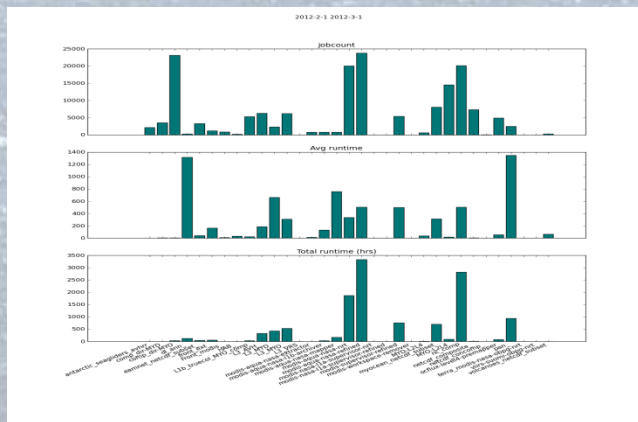
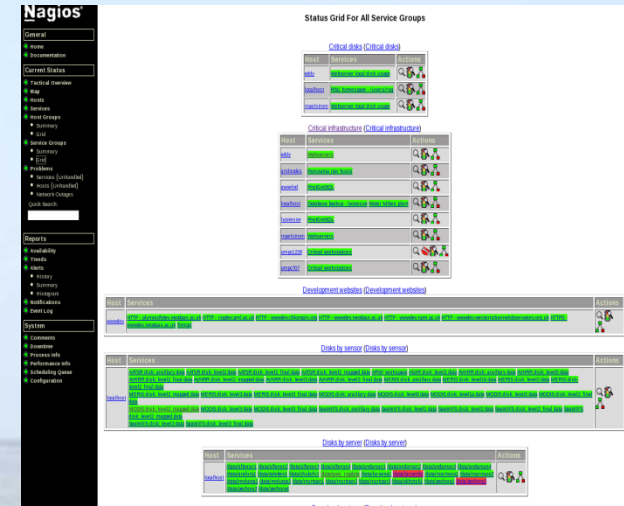
GloboLakes Processing Flow Chart



- **Modularity** of processing is essential to this style of system
  - need to use third-party software (e.g.. BEAM), so must be **flexible** and able to wrap external modules into processing chain
- **Supervisor framework** to encapsulate, control and monitor in-house or external software



- Monitoring a large system is a challenge
  - Should be responsive to errors
  - Should abstract detail to reduce “swamping”
  - Must be actively policed (things will break!)
  
- PML monitors on many levels
  - Detailed:** debug logs, etc.
  - Interrupt:** email to processing alias
  - System:** nagios
  - Abstract:** graphing, analysis tools





- “CEMS is a Cloud Computing infrastructure hosting EO and climate datasets (including some in-situ datasets).
- CEMS will provide the climate research and exploitation community with a single point of access to the required data as well as a hosted processing facility. This will support algorithm and product development, improvements in data quality and the timeliness of data and product dissemination
- CEMS is sized to be able to store 1.7PBytes of EO and climate data and features 480 cores for data processing”
- CEMS is included in the UK Collaborative Ground Segment proposal and so will have Sentinel 2 and 3 data available
- May be politically expedient to do processing on CEMS (or have no choice) but
  - What is price?



- CCI experimenting with “Calvalus” cluster
  - Similar hardware, but larger local disks
  - Runs Apache Hadoop framework
  - Key benefit: **data-local parallelisation** [MapReduce]
  - Processing happens where the data are, rather than moving the data over the network
  - Experiments with MERIS L1-L2 using Polymer atmos. correction
    - = 1 year of global RR data
    - = 1 day on the hardware – we expect greater speed up when using the SW infrastructure