

PML

Plymouth Marine
Laboratory

Marine Matters

GloboLakes WP2: data processing and distribution

Steve Groom
10th December 2012

sbg@pml.ac.uk





PML lead, also UoE and UoS

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

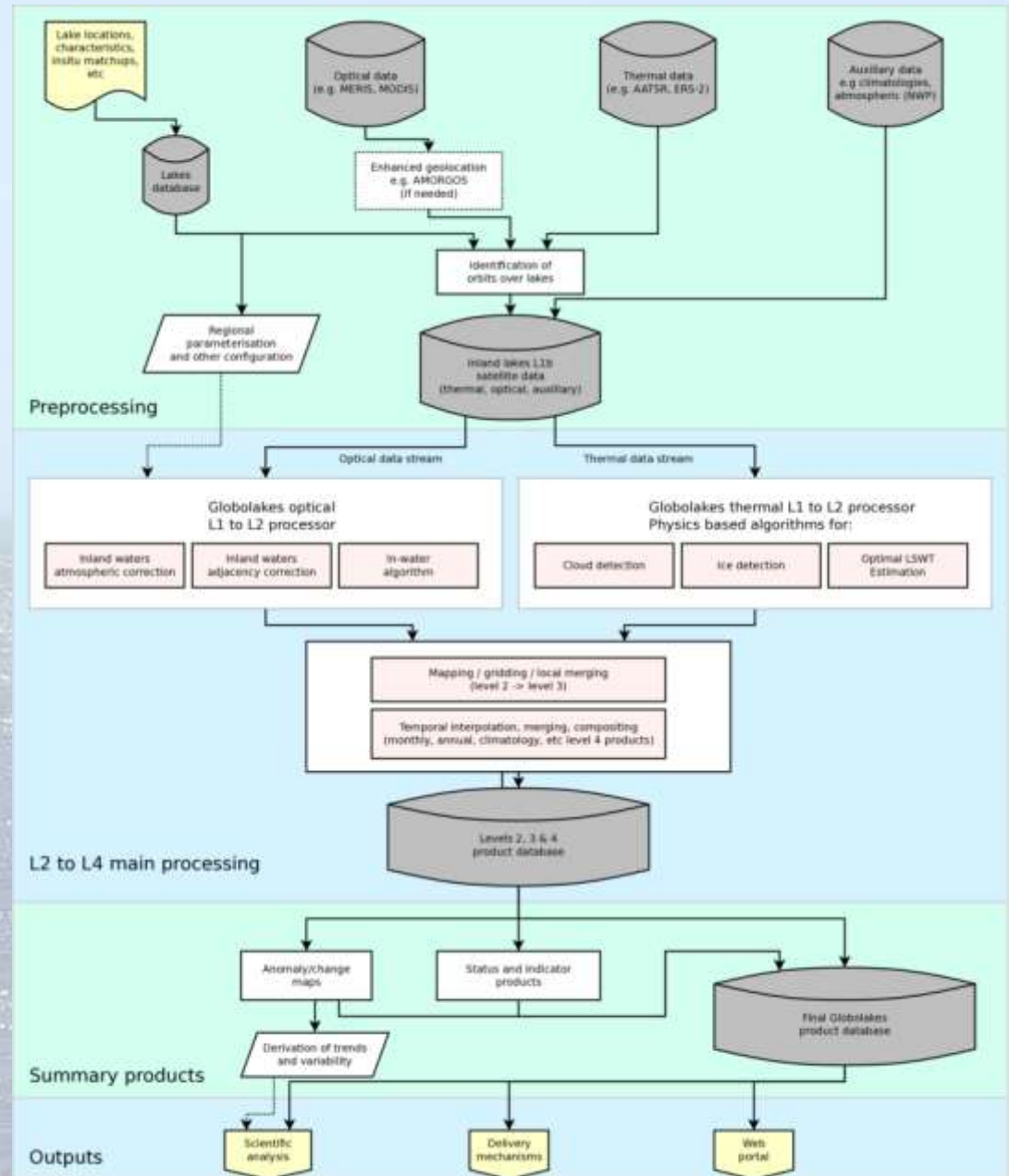
- 2.1 Develop highly automated data processing chain
- 2.2 Process the entire MERIS 300 m archive for the ~1000 target lakes + ARCLakes processing
- 2.3 Run a short time delay monitoring capability as a prime input to the Global Lake Observatory
- 2.4 Extend methods to ESA/ GMES Sentinel 3 OLCI & SLSTR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



- Requirements (inputs)
 - From ESA, complete global archive of MERIS level 1 FR data
 - From ESA NRT stream of Sentinel 2 and Sentinel 3 data
 - From WP1 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 ~2014-2017; NRT lake monitoring
 - Sentinel 2, MSI: feasibility of MSI for lake monitoring



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



- Commodity Linux cluster
 - PML's existing systems: 2 x 36 x Dell R200 class machines (8GB RAM, 1-4 core, small disk, 1Gb NIC)
- GloboLakes will pay for an additional:
 - 16 grid nodes in 2013-14*
 - ~100TB in 2012-13 & ???TB 2015*
- Also looking at
 - CEMS hardware at Harwell
 - Brockman Consult Calvalus system

* To be changed depending on S2 and S3 launches

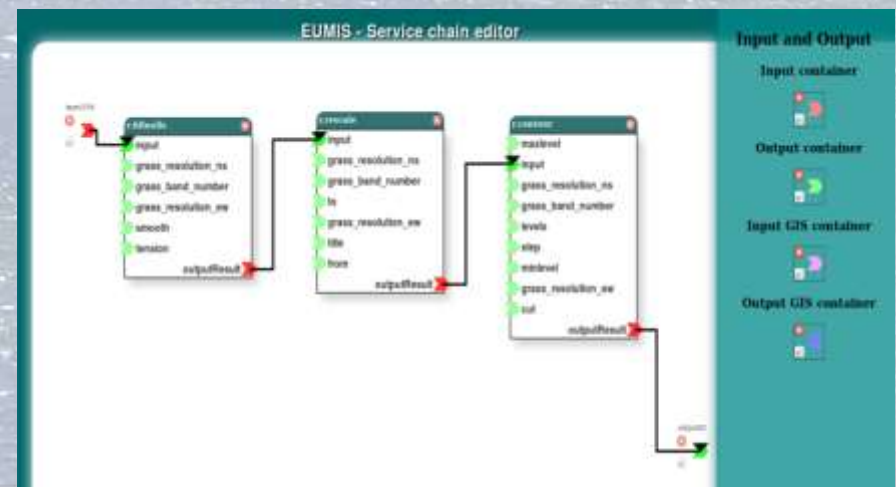
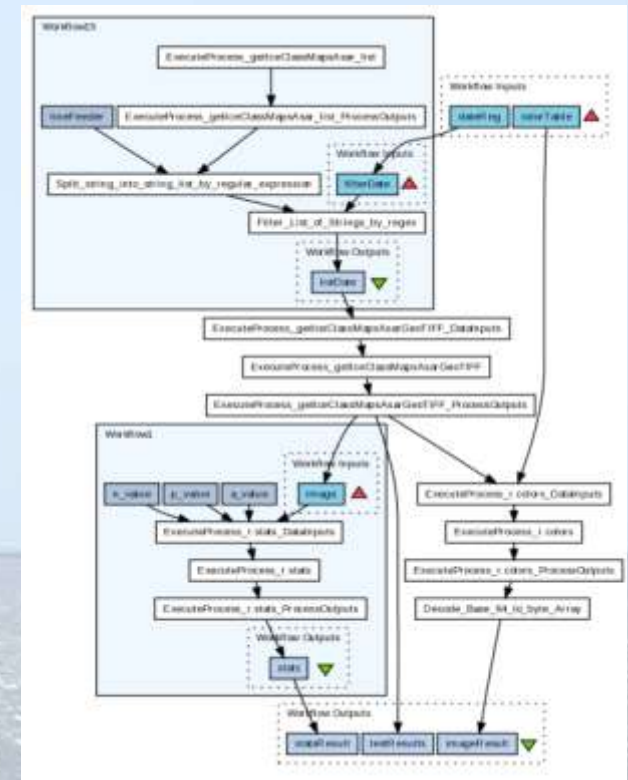


The pretty half of the server room



Investigating web-based processing in FP7 NETMAR

- 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
- Alternative is MODIS that has limited capability at 500m: 2 bands that can be used for chl-a retrieval in large lakes
- Sentinel 3 OLCI due for launch in 2013/4



MERIS algal-2 300m, 21st Feb. 2012, 0939



MODIS OC3 500m, 21st Feb. 2012, 1208



- Research activity
- Will depend on data availability, processing levels etc.





- Web site based on NEODAAS browser and ChloroGIN Lakes
 - 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



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):

AVAILABLE AREAS

--



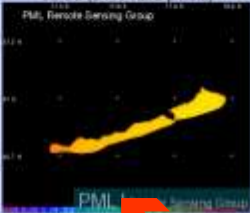
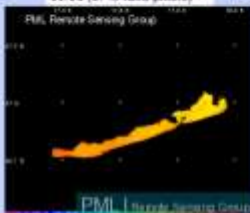
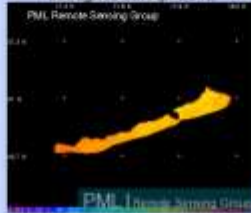
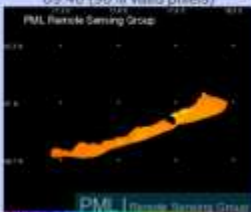
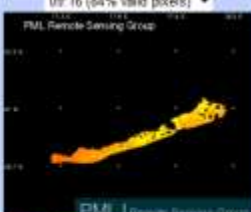

NEODAAS

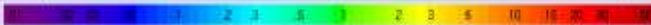
Logout Hello sbg
System status: ● ● ●
Introduction (Area: my)
Relatem lake larger geographic
Getting started
Demonstrations
Current summary
HelpFAQ
Contact us
+ Navigate data
- Select area
- Select date
+ Current data
Flex/View multi-products
+ Sensor
+ MERIS 300m
Chlorophyll (algal_1)
Chlorophyll (algal_2)
RGI
Radiance (NLW 500)
Aerosol opt. thick
Yellow substance
SPM
Vegetation (toa)
Vegetation (toa)
Chlorophyll (oc3)
Kil 490 (SeaWiFS)
Aerosol opt. thick 88
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: small;">PML Remote Sensing Group</p> <p>You are viewing scenes: Switch to Composites</p>	<p>29 Days Ago (4th Apr 2012) No images</p> <div style="background-color: #ccc; padding: 20px; text-align: center;">no data available</div>	<p>30 Days Ago (3rd Apr 2012) 09:06 (87% valid pixels)</p>  <p style="font-size: small;">PML Remote Sensing Group</p>	<p>From 30th Mar 2012 through to 5th Apr 2012</p>  <p style="font-size: small;">PML Remote Sensing Group</p>
	<p>2nd Apr 2012 09:40 (90% valid pixels)</p>  <p style="font-size: small;">PML Remote Sensing Group</p>	<p>1st Apr 2012 No images</p> <div style="background-color: #ccc; padding: 20px; text-align: center;">no data available</div>	
	<p>31st Mar 2012 09:16 (84% valid pixels)</p>  <p style="font-size: small;">PML Remote Sensing Group</p>	<p>30th Mar 2012 09:50 (33% valid pixels)</p>  <p style="font-size: small;">PML Remote Sensing Group</p>	

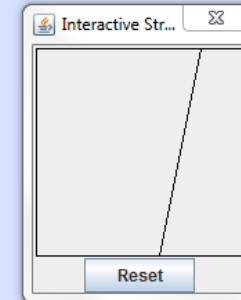
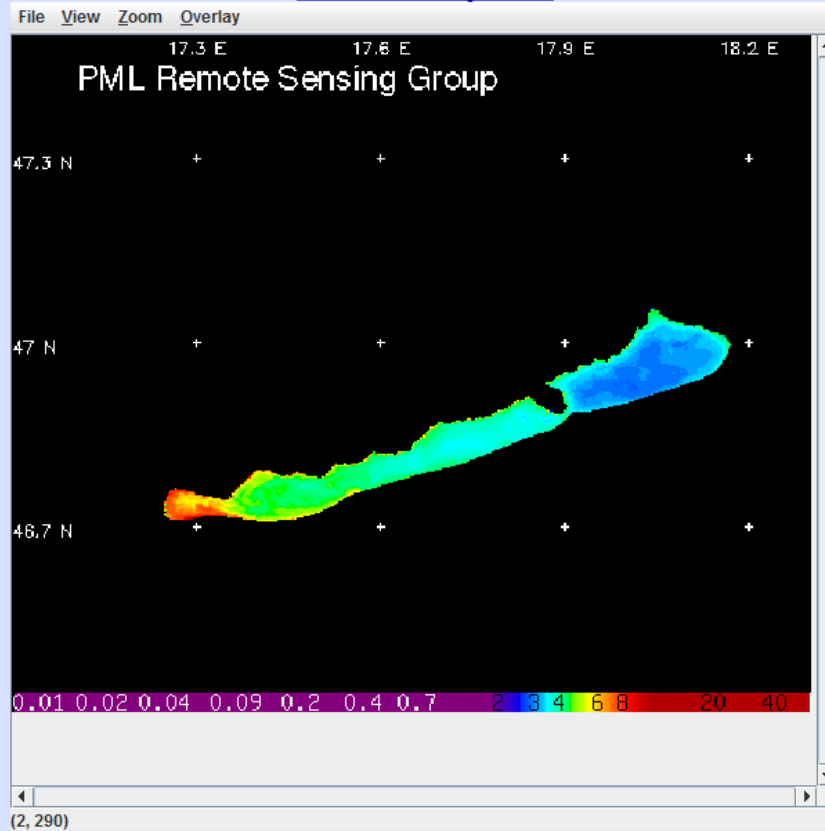


Units: mg m⁻³
For additional information on this product, please check the [help/FAQ](#)
[Feedback](#)



MERIS algal_2 (2012-04-05 09:30)

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



Values

Lat/Lon Units: Decimal Degrees

X: 0001

Y: 0241

Latitude: 046.8517

Longitude: 017.0027

Digital Number: 000

Real World Val: 000.0100

Selected Pixel: [black box]

MERIS algal_2 (2012-04-05 09:30)

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

Custom Zoom

120% Reset Fit



- 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)



FP7 OPEC development portal



AMT visualisation portal



HOME

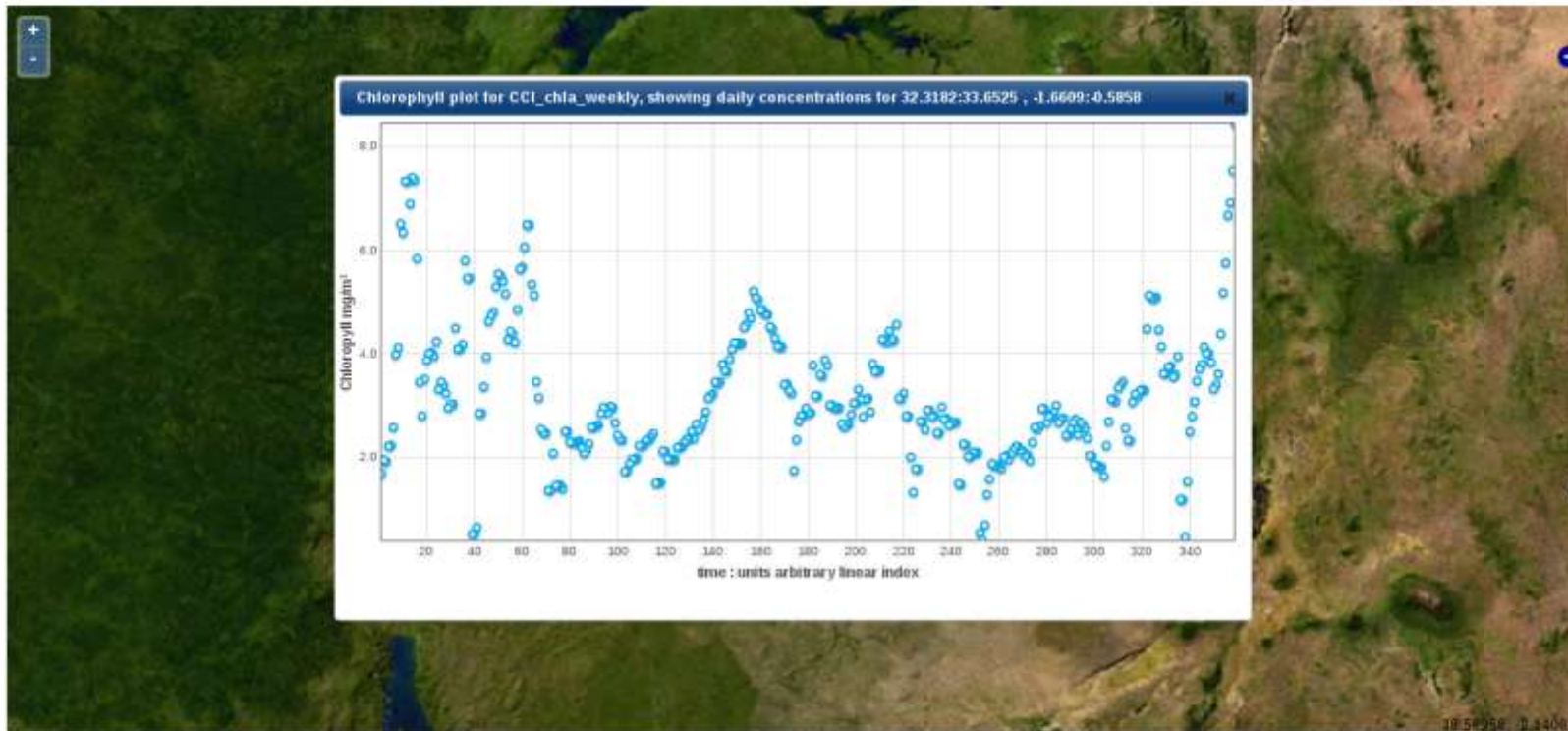
BETA AREA

WCS

WCPS

DATA

[open query form](#) ▼



38.50958454731:-2.1408041182631

Opacity Slider

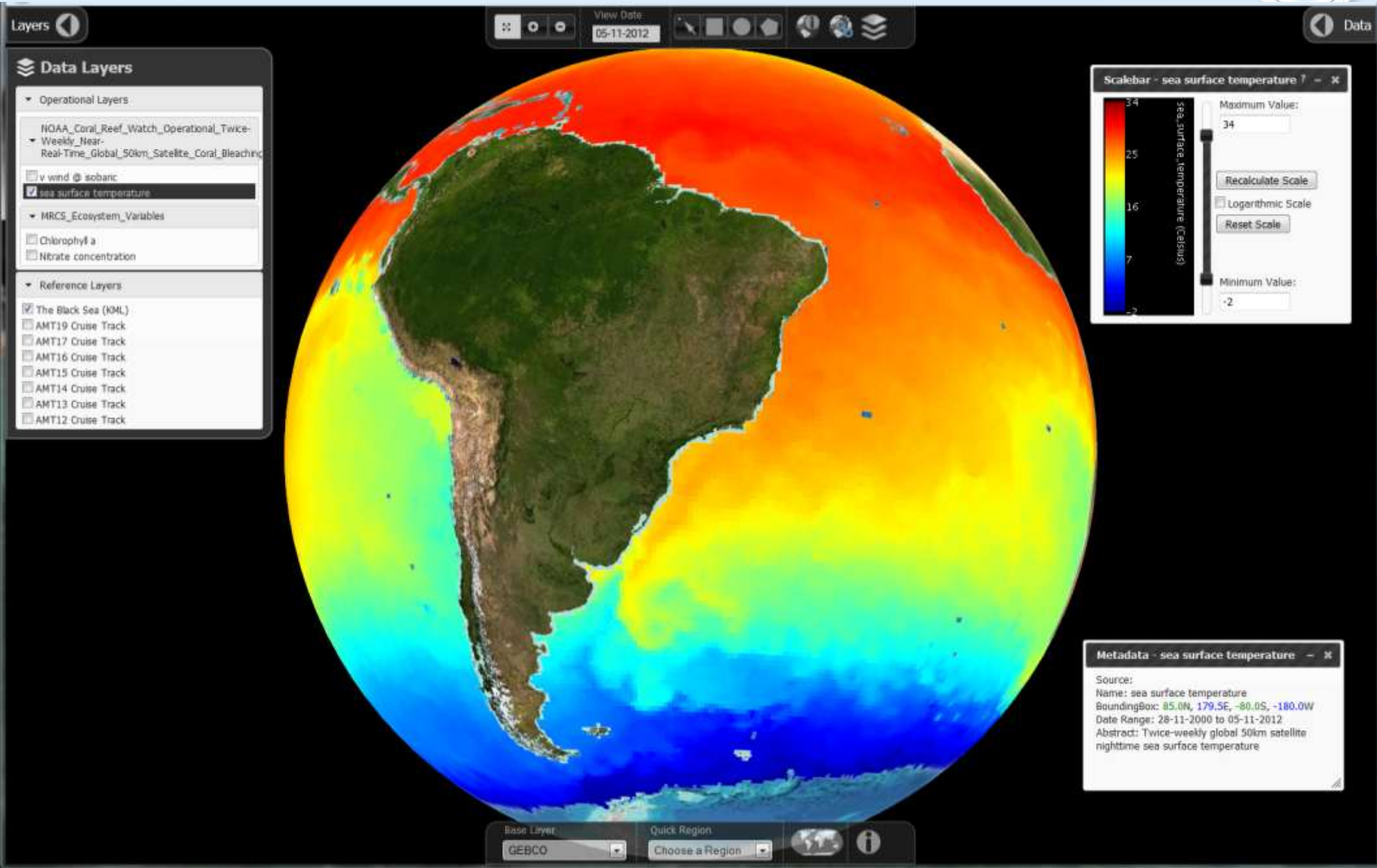
1.0

[Contact us](#) | [Site Map](#)

All materials copyright © Plymouth Marine Laboratory 2012.



- **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

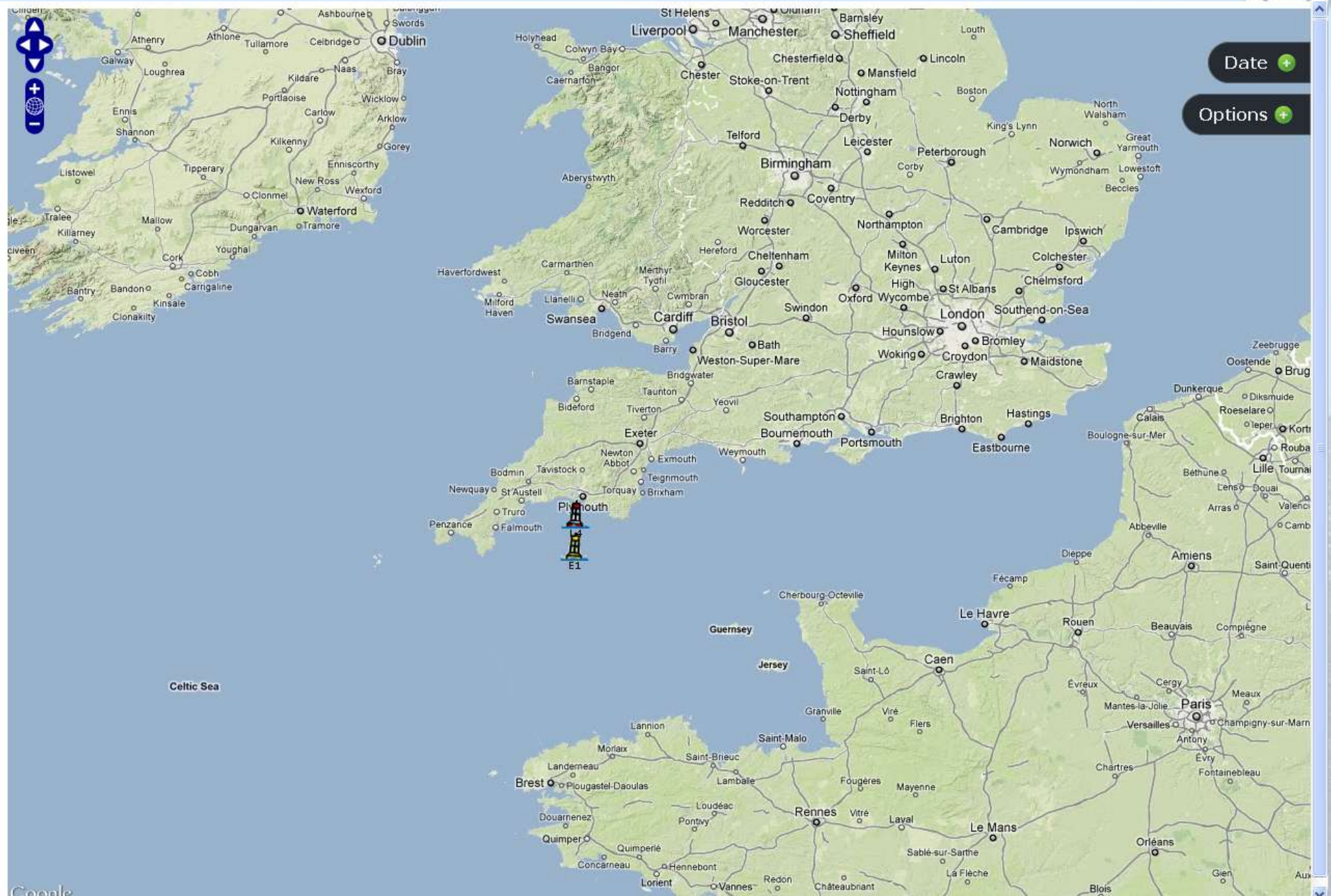


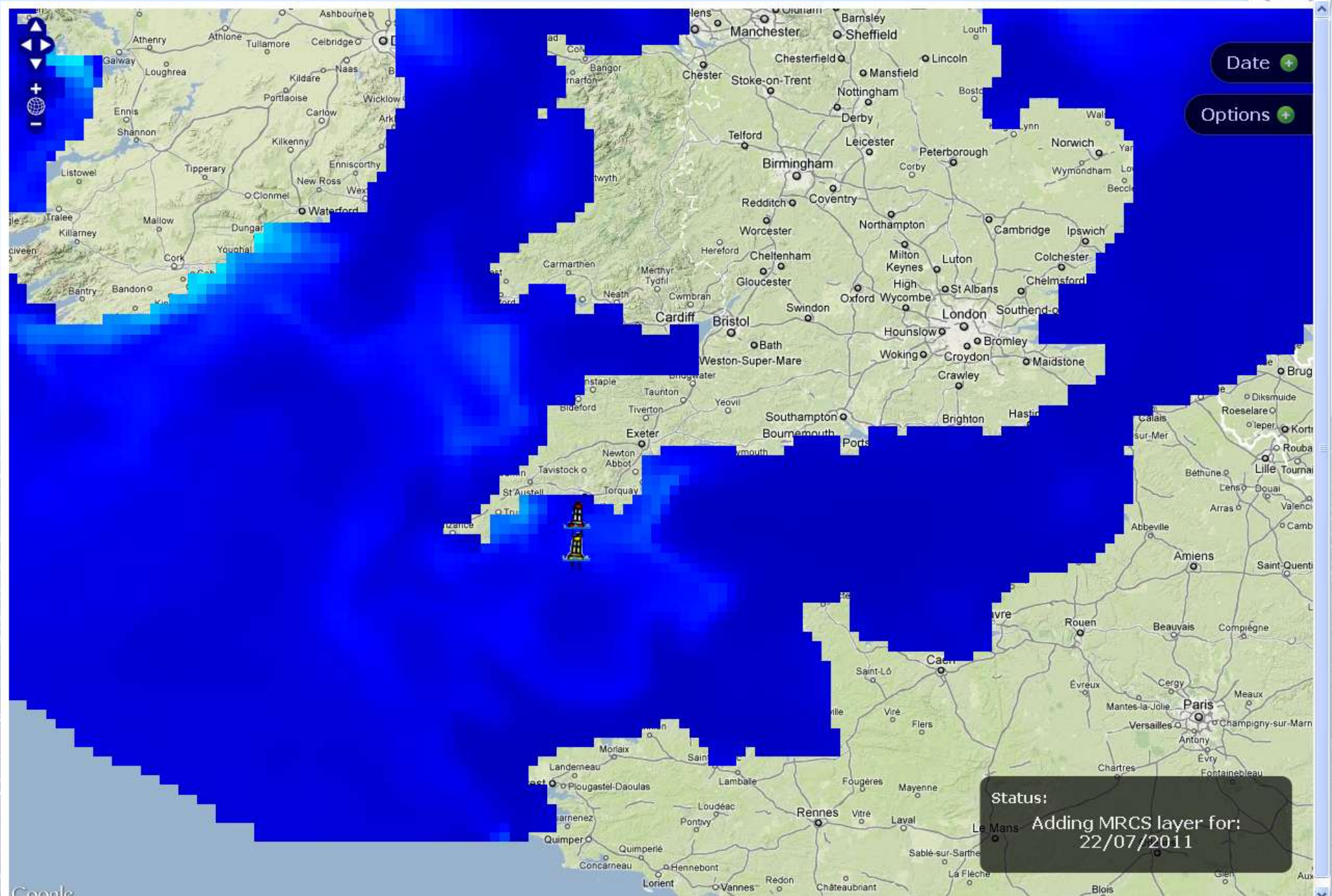


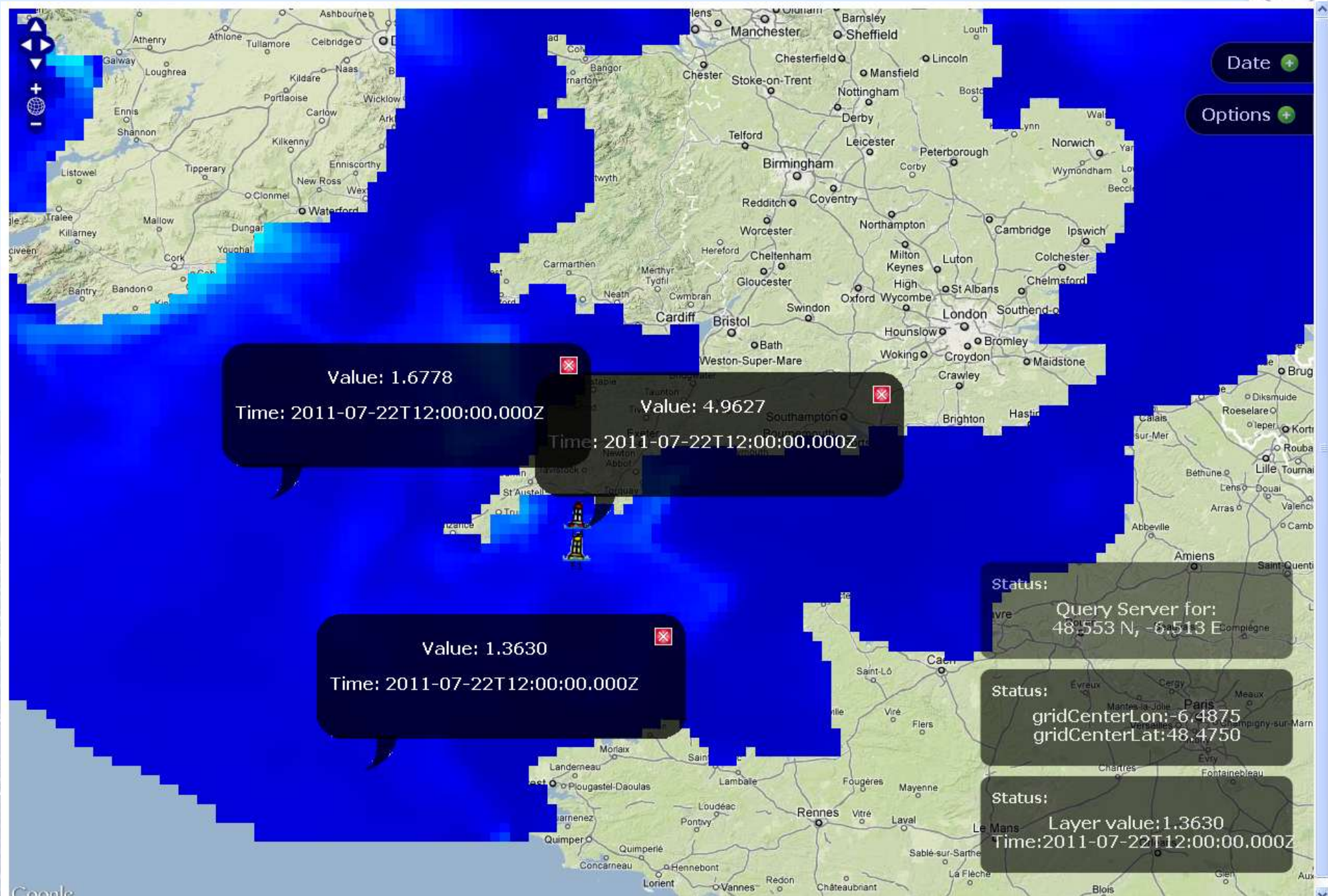
- GloboLakes will be an exciting challenge!
- Expect/hope to work well with Diversity II and GLaSS

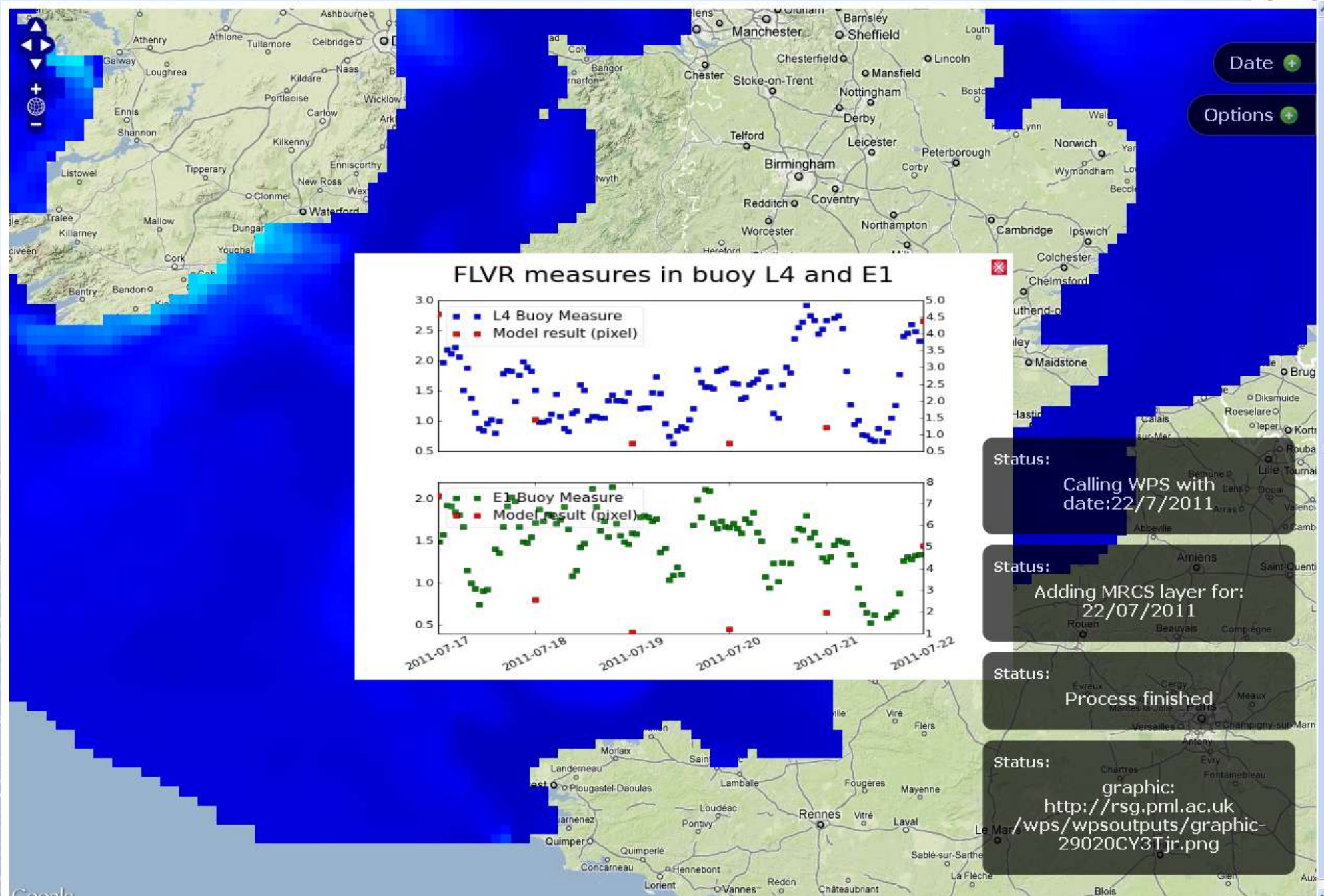
Advertisement

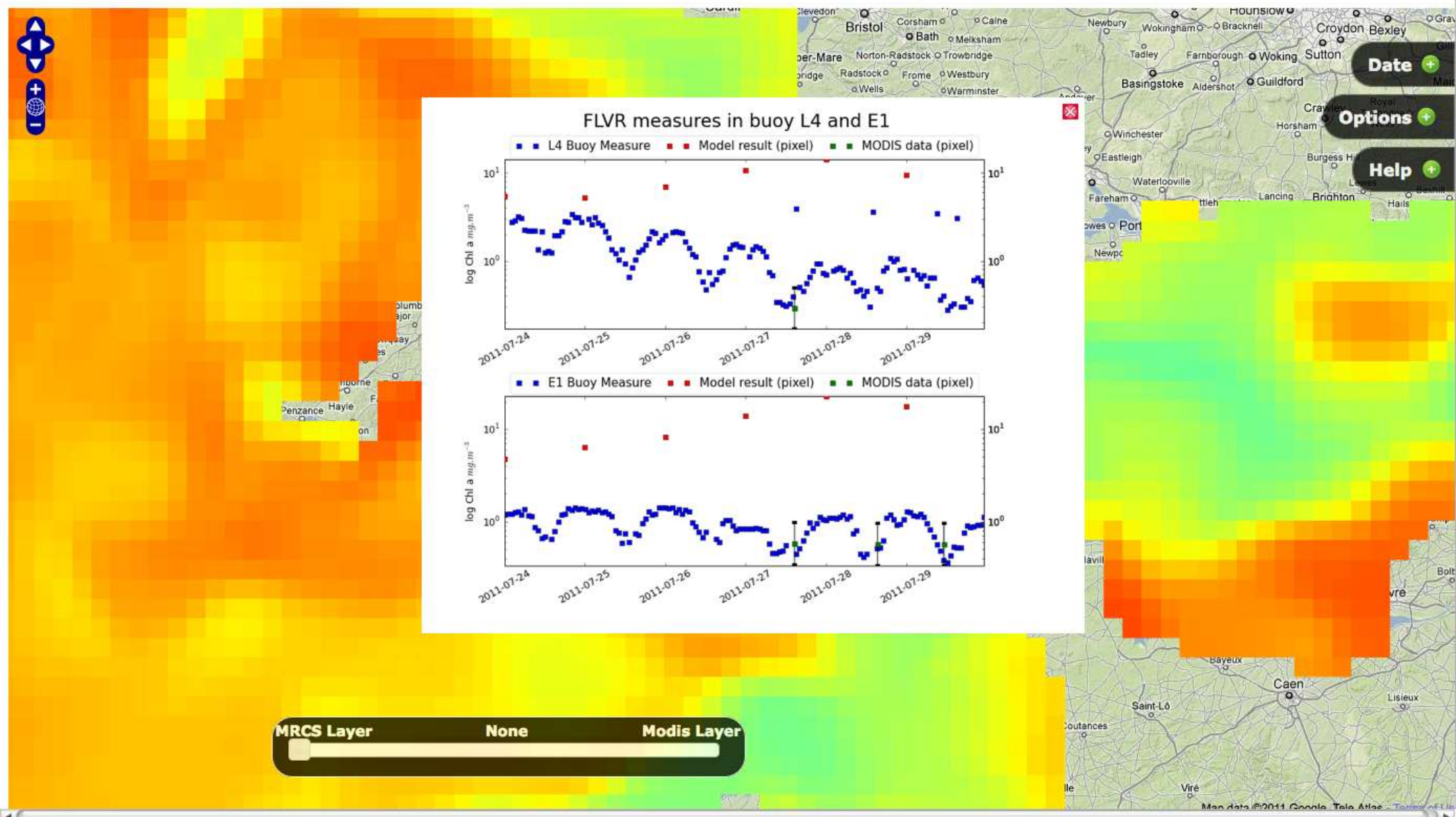
- The PML GloboLakes post-doc post is open
- Permanent / open-ended post ie not just fixed term
- If anyone knows of suitable candidates catch me whilst I'm here



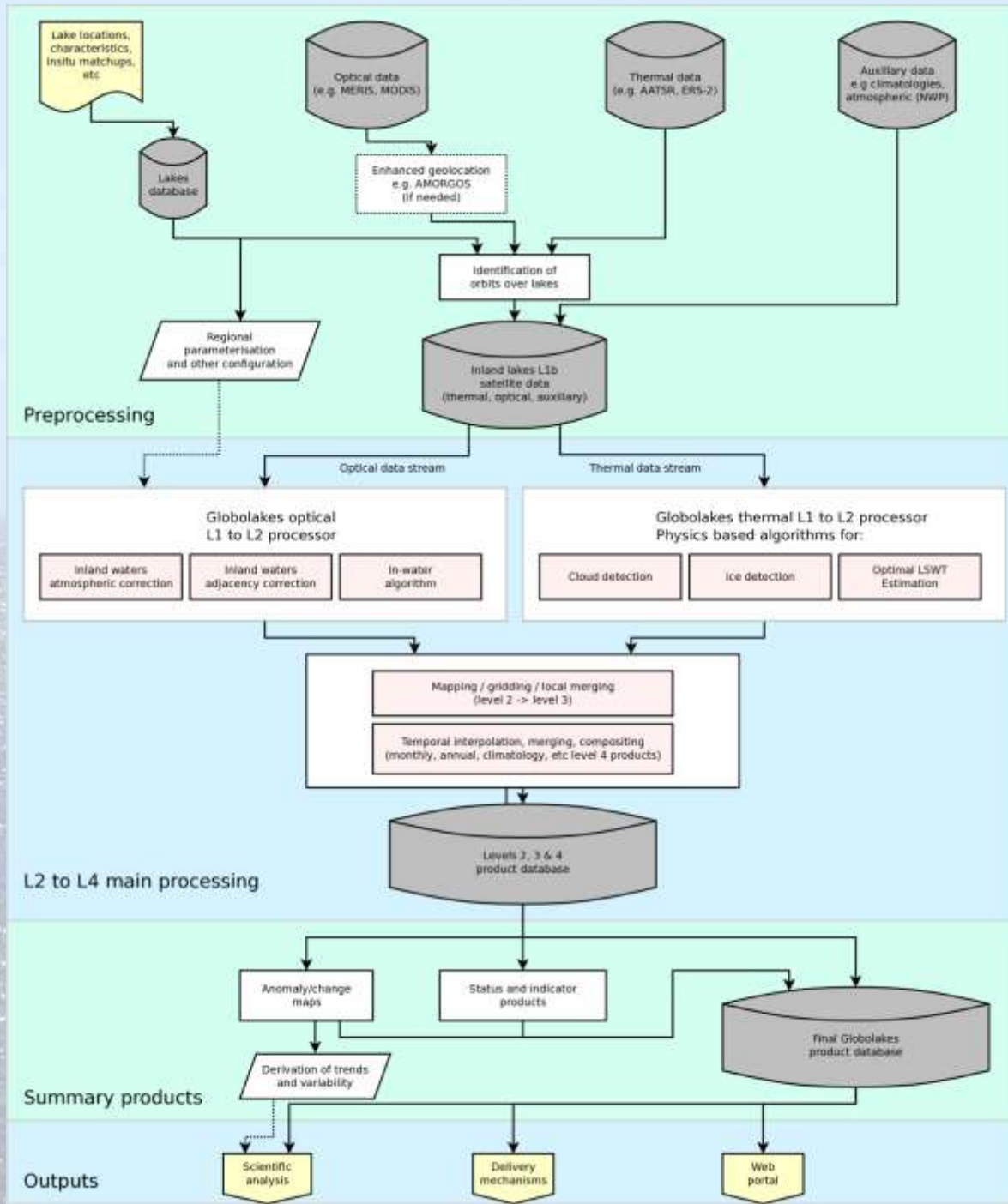












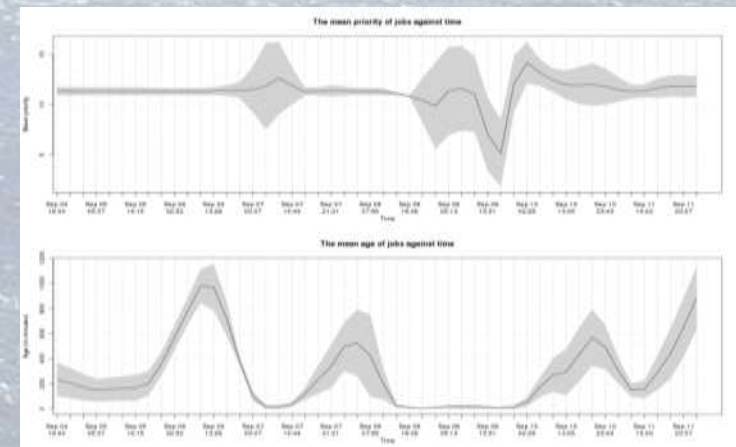
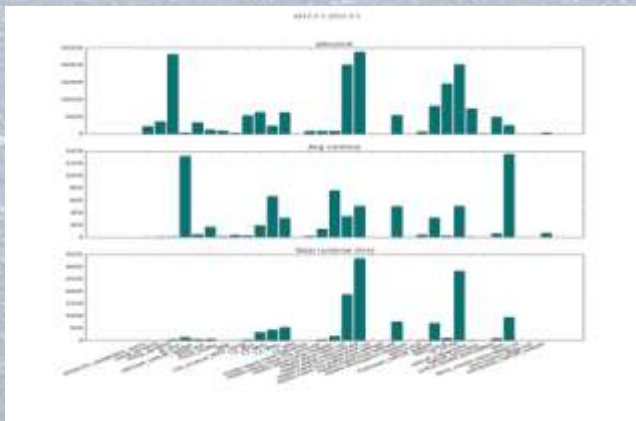
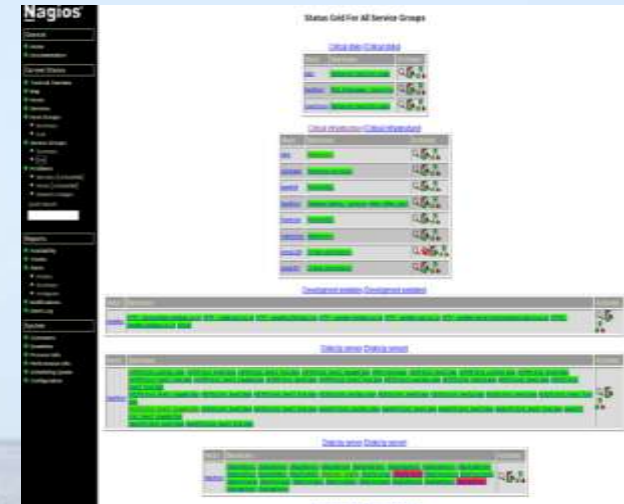
GloboLakes Processing Flow Chart



- **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

