Oceanography and Ocean Data Assimilation
Geospatial Data Grids

Keith Haines

**Environmental Systems Science Centre (ESSC),
Reading University**

Reading e-Science Centre

www.resc.reading.ac.uk

New technology for data distribution and integration (GRID)

- What is a "Grid"?
- Services on the internet
  - Web and Grid Services
  - Machine-Machine Communication Standards!
- Examples of some large/small Grid Projects
  - ESA, Particles Physics, Campus Grids
- Environmental Web Services
  - Data Sharing: OpenDAP and LAS
  - GIS and the Open GIS Consortium (OGC)
  - Data viewing tools:- Google Maps/Earth; OpenLayers
- Conclusions

Grids: a foundation for e-Research

- Integrating technology, enabling a whole-system approach to complex problems

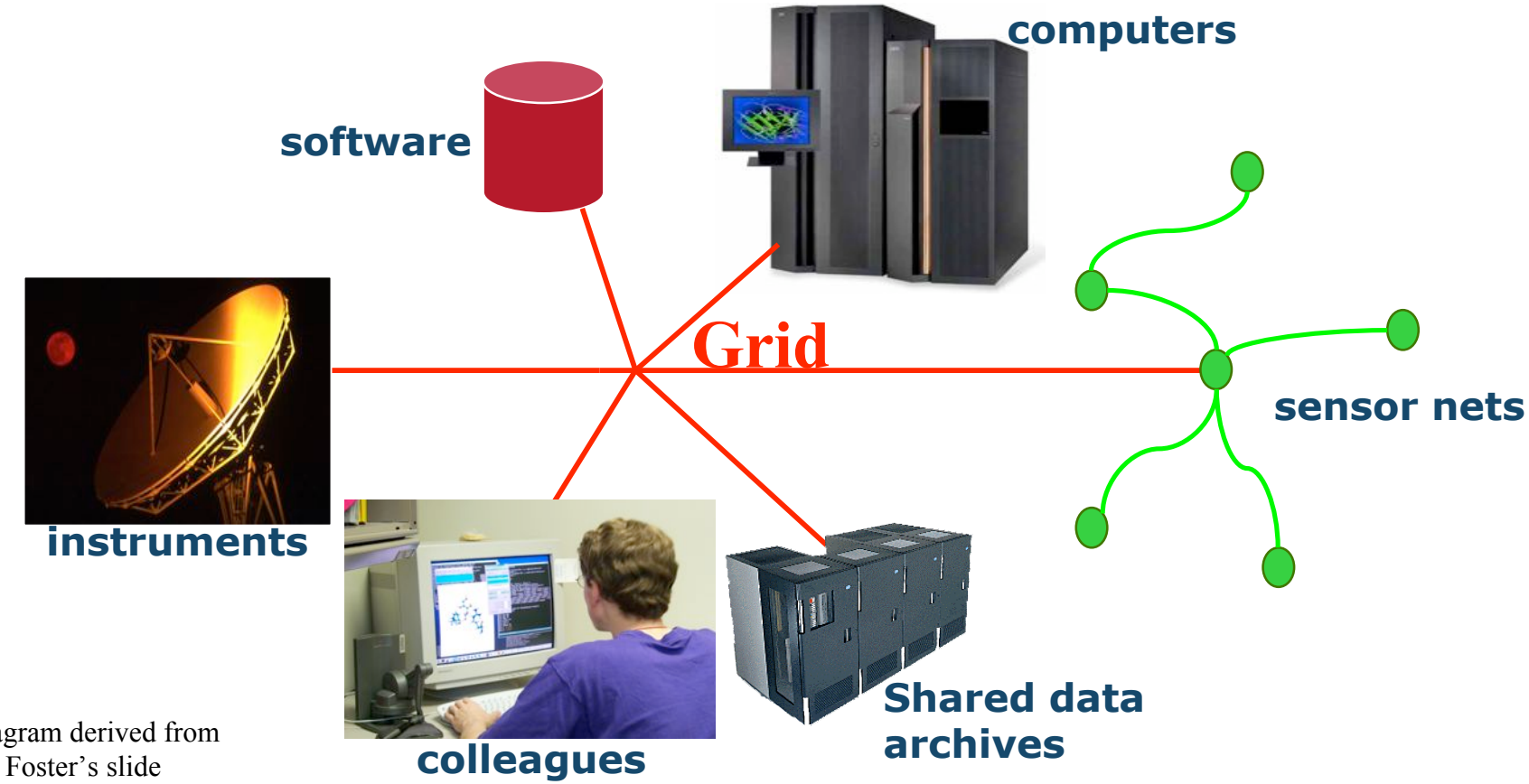


Diagram derived from Ian Foster's slide

Operational example on a Grid

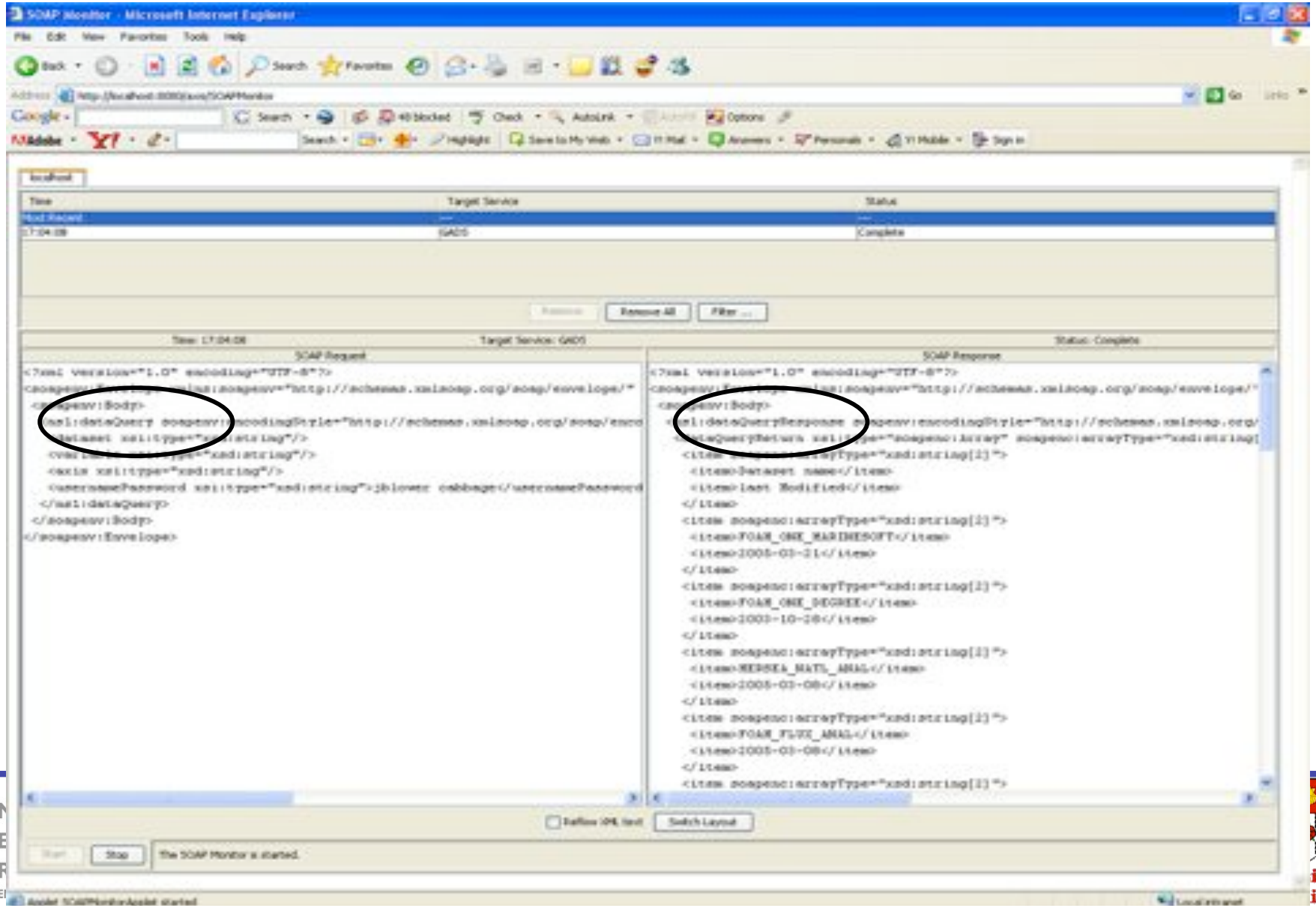
- Authorised Scientist schedules operations on all Resources with single sign on
- Instrument : Makes series of Measurements
- Instrument sends data to Permanent Data Store and to Computer Resource
- Computer resource makes diagnostics and compares with previous data in data store
- Particular result alerts scientist and requests Instrument to make more measurements before resuming normal schedule

- Many Grids only involve distributed computing resources and data stores
- Requires secure software for machine-machine communication across internet using Standard protocols

Web Services

- A Web Service can be thought of as a Subroutine that is available over the internet with an Input and an Output
- The I/O is coded in with a standard protocol called a SOAP message (Simple Object Access Protocol)
- A SOAP message is written in XML eXtensible Markup Language in such a way that it conforms the a Standard Format WSRF
- Web Services can be initiated by people providing Input on a web page, but they can also be initiated by other web services or applications running on the internet
- Can therefore string web services together, just like subroutines, in a Workflow. Various tools exist to build complex workflows; www.trianacode.org, taverna.sourceforge.net, BPEL..
- Services are normally managed locally by an Applications server such as Tomcat, JBoss, Jetty, IBM WebSphere, Oracle AppsSer.

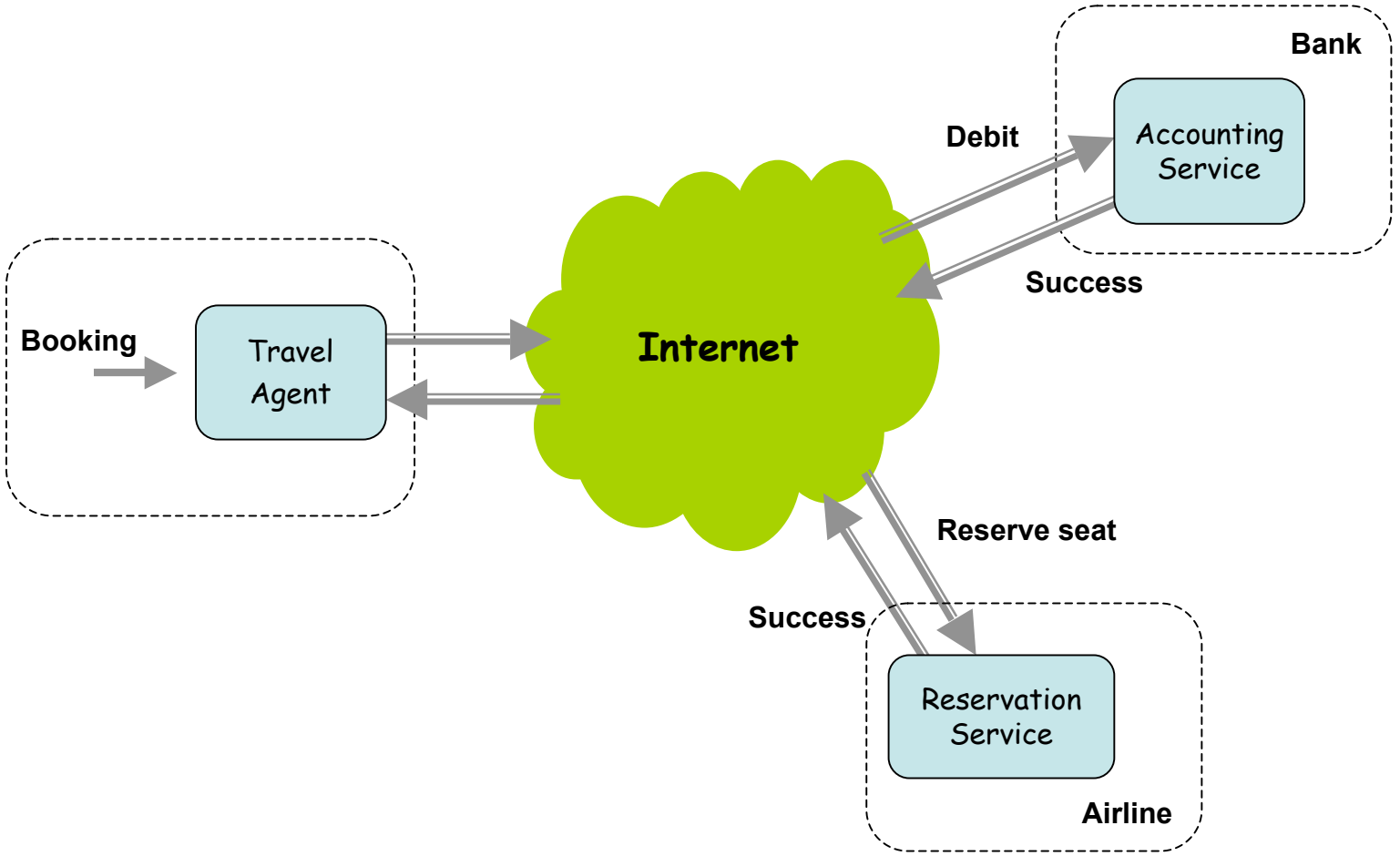
Example of dataQuery Request/Response Soap message



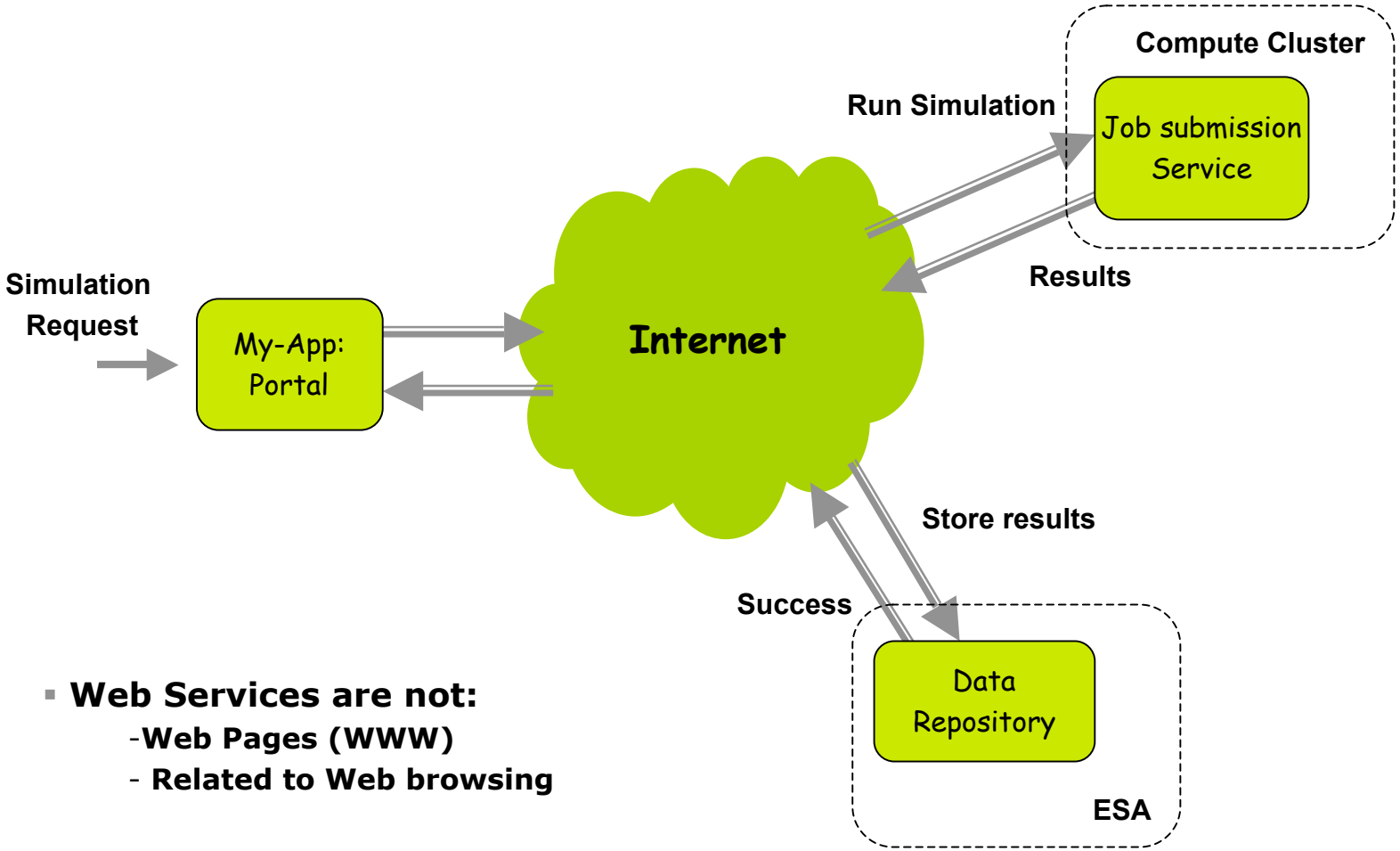
Web Services

- A Web Service can be thought of as a Subroutine that is available over the internet with an Input and an Output
- The I/O is coded in with a standard protocol called a SOAP message (Simple Object Access Protocol)
- A SOAP message is written in XML eXtensible Markup Language in such a way that it conforms the a Standard Format WSRF
- Web Services can be initiated by people providing Input on a web page, but they can also be initiated by other web services or applications running on the internet
- Can therefore string web services together, just like subroutines, in a Workflow. Various tools exist to build complex workflows; Triana, Taverna (Scientific), BPEL. (Business).
- Services are normally managed locally by an Applications server such as Tomcat, JBoss, Jetty, IBM WebSphere, Oracle AppsSer.

Web Services in e-Business



Web Services in e-Science



- **Web Services are not:**
  - **-Web Pages (WWW)**
  - **- Related to Web browsing**

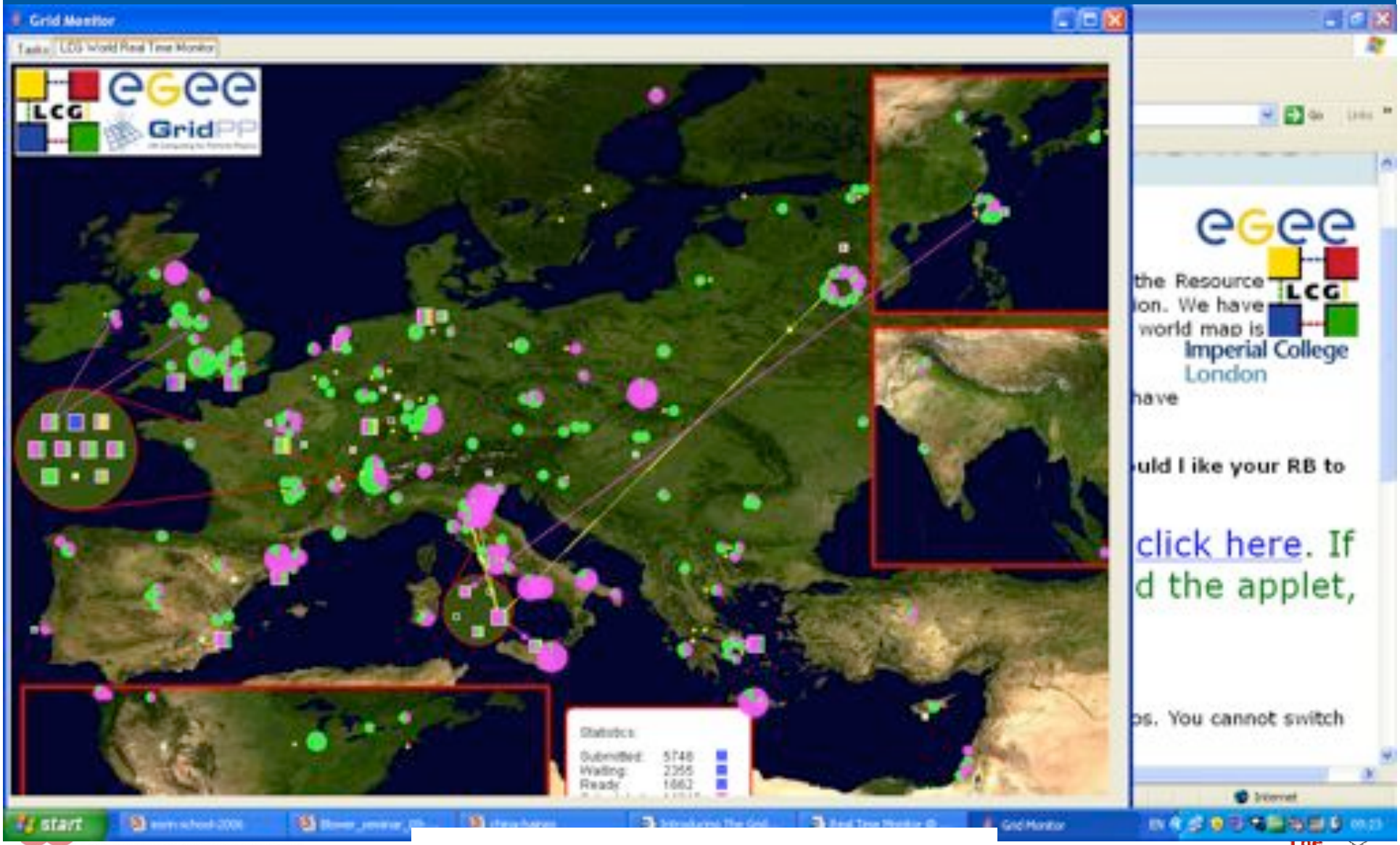
Web Services

- Web Services may
  - Provide metadata, data
  - Do some computations
- May require some Security on Web Services (who uses them, how much resource is requested/available). Standards WS-Security
- WS originally designed for Business use => limitations
- XML based Communication method not suitable for large I/O
  - Work around eg. by passing URL links to large I/O data volumes
- Web Services designed for instant responses. Not suitable for significant computing or “Batch mode” tasks
- Solution => Grid Services which are Web Services with “State”
  - Can therefore request information on how far they’ve got

Web and Grid Services

- Web and Grid Services are the basic building blocks of a “Grid” which is a set of distributed resources that can interoperate
- Key issue is Standards for machine-machine communication
  - Standards take a long time to agree upon (ISO)
  - Big computer companies have had tendency to try to set their own standards to lock users in
- Most “Grids” and Grid Service “middleware” packages are designed around access to computational resources, eg HPC, Compute clusters, Disk space, Databases
  - Big Science users are Particle Physics Community who plan to analyse LHC data on a global Grid from 2007 onwards
- Biggest problems in setting up a Computational Grid revolve around security => Digital certificates, **Globus toolkit,** Not so user friendly, g-Lite, other solutions eg. Inferno….

Particle Physics Grid for LHC



RESEARCH COUNCIL
ENVIRONMENTAL SYSTEMS SCIENCE CENTRE

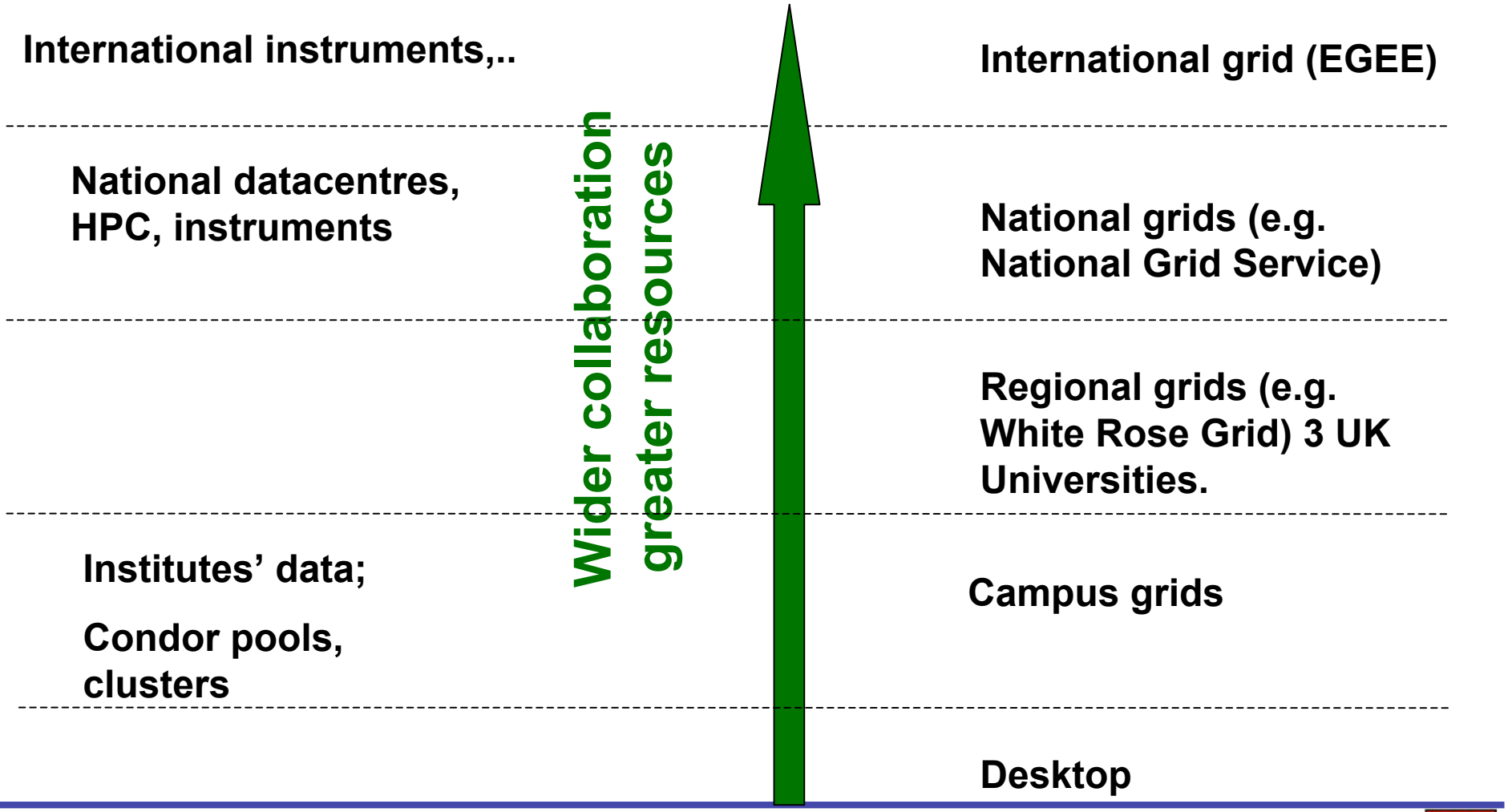
http://gridportal.hep.ph.ic.ac.uk/rtm/

The University of Reading

EU Grid project

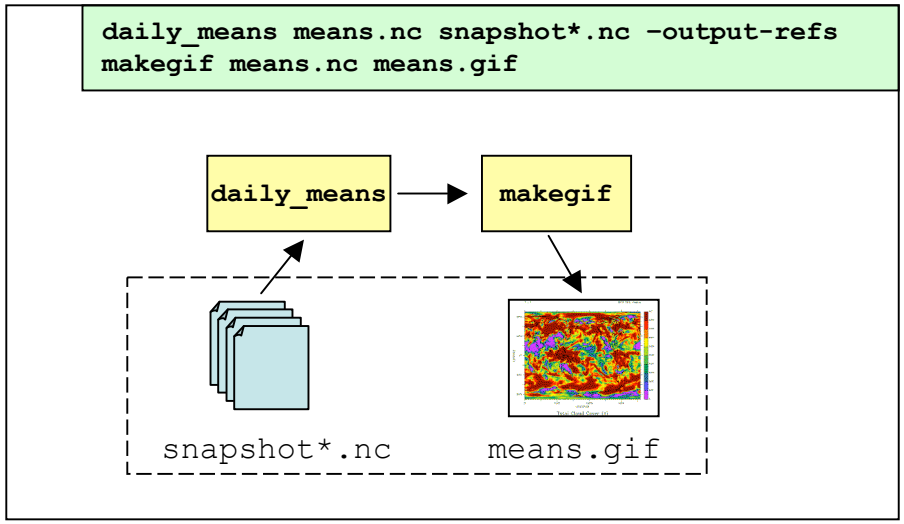
UK National Grid Service

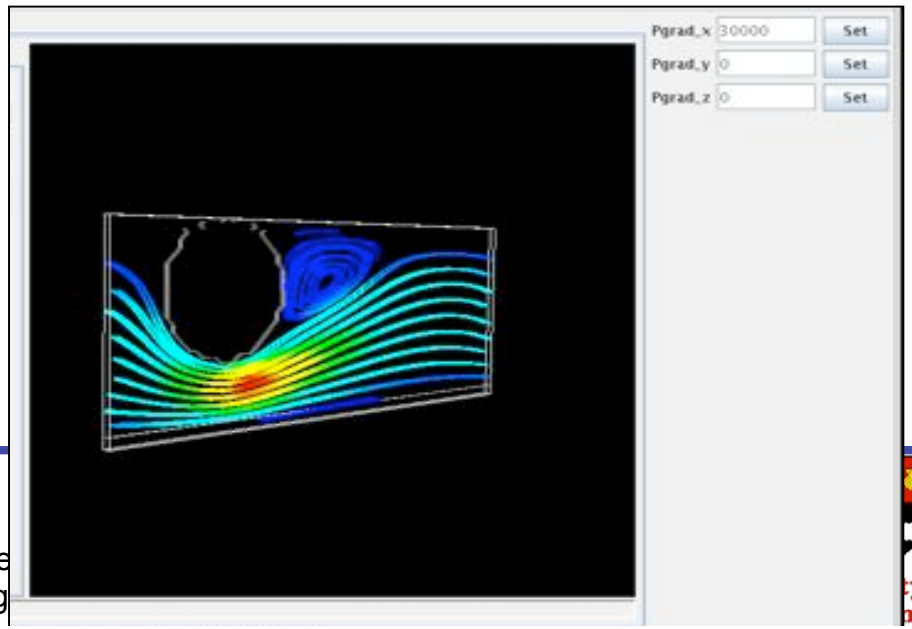
The many scales of grids



Styx Grid Services

- Easy-to-use, lightweight middleware for e-Science
- 5-minute installation
- Can expose existing executables as services
- Run them from the command line exactly as if they were local programs
- Create workflows with simple Unix-type shell scripts (above right) over distributed network using ssh security
- Been used for computational steering and collaborative visualization (below right)
- Blower et al 2006 www.resc.reading.ac.uk publics. http://jstyx.sf.net





Environmental Web and Grid Services?

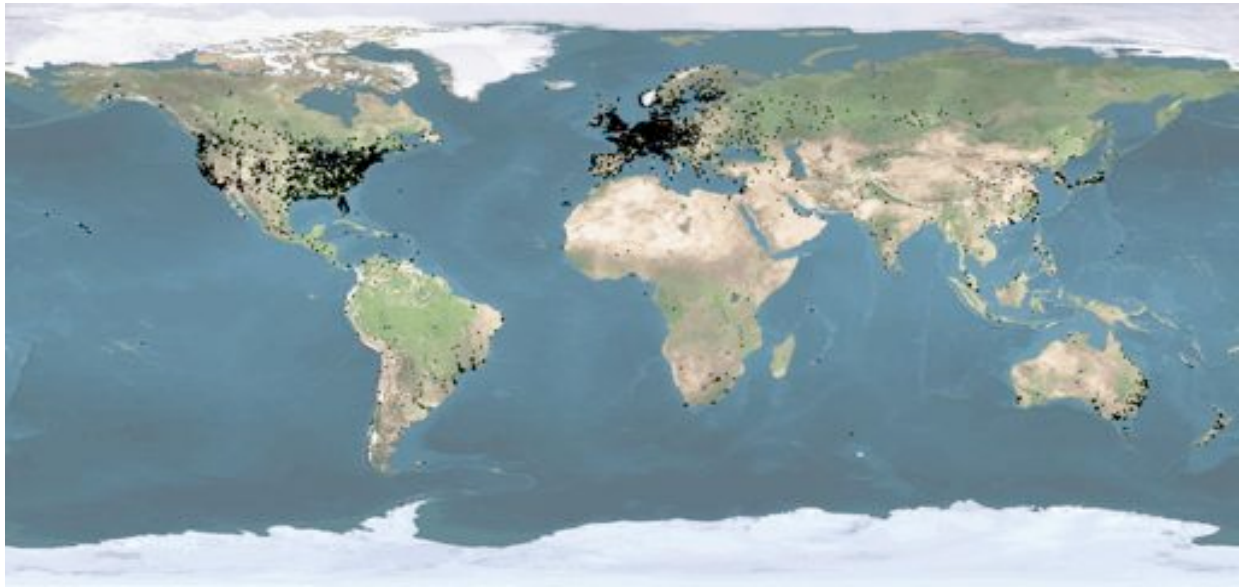
- Big problems in Environmental science revolve around Large and diverse data sets; In Situ, Satellite Observations, Model Output
- Need Standards to manage large geospatial data sets, in particular standards for METADATA; eg. Data Type, Provenance and Geospatial and Temporal structure
- ESA runs a "Grid on Demand"
  - Processing resources to provide client tailored ESA satellite data
  - Open call for Users
- UK has sponsored a number of "Environmental e-Science" projects
  - ClimatePrediction.net
  - NERC DataGrid
- Visualisation of Environmental data on the Internet
  - LAS/OPenDAP, GODIVA/GADS, OpenGIS, Google Earth, OpenLayers

Luigi.Fusco@esa.int

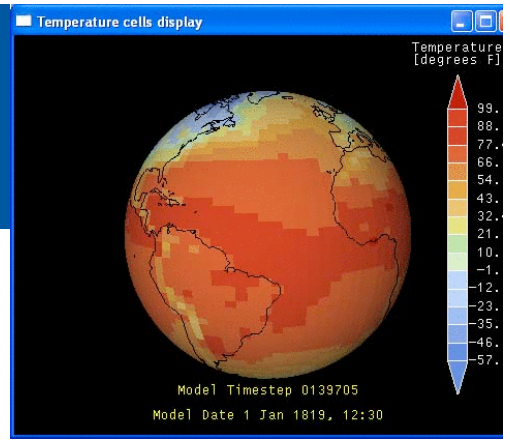

ESA EO Principal Investigator Portal - Microsoft Internet Explorer
esa
Earth Observation
Principal Investigator Portal
Contact us
Exploitation Results & News
EO Grid Processing on Demand
G-POD
Results
News
Search
Focus on PI
Round table
AO Submission
Cat-1 & Open AOs
Previous AOs
Update & Reporting
Services
About this site
ESA Data Policy
FAQ
Related Links
How to get ESA data
HOME
Welcome to the submission area for Category-1 data users on the Earth Observation GRID Processing on Demand environment
G-POD
ESA is offering all scientists with the possibility to perform bulk processing and/or validation of their own algorithms exploiting the large ESA earth-observation archive together with ESA available GRID computing and dynamic storage resources.
The deadline for proposal submission is 1st September 2006.
A maximum of 10 proposals will be selected for implementation in the frame of this Call.
Invitation letter
Call Main Text
Terms and Conditions for the Utilization of Data under the ESA Category-1 scheme
Guidelines for the submission of proposals
Proposal Submission
More about G-POD
Contact the ESA EOPI Help and Order Desk for further information.
start

climateprediction.net

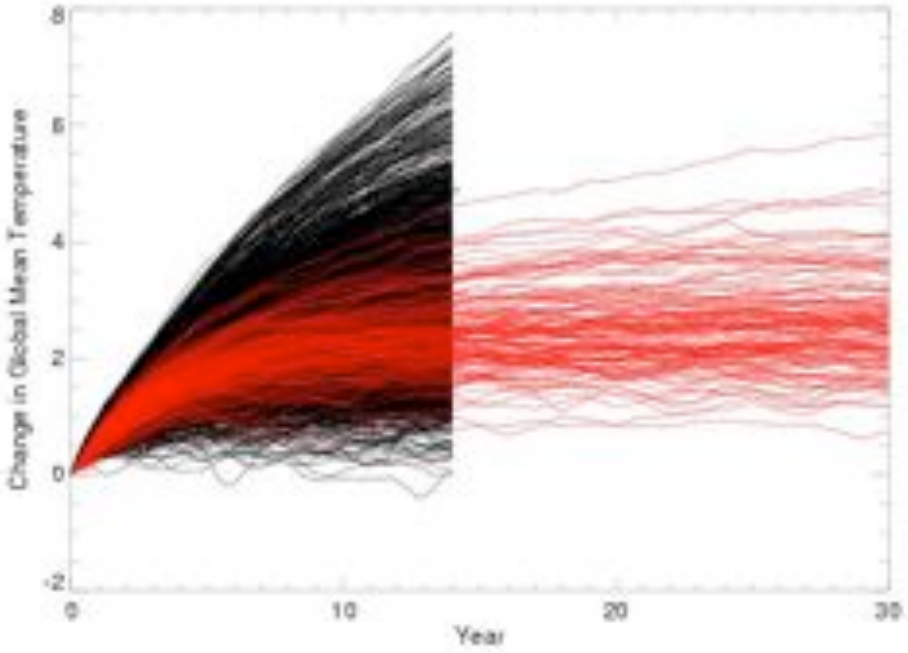
- **Distributed Global Collaborations**
- Hadley climate model cut down to run on single PC (cf. Seti@home)
- 105,000 people from 150 countries have donated 10,000 years of computing time to undertake climate change experiments.



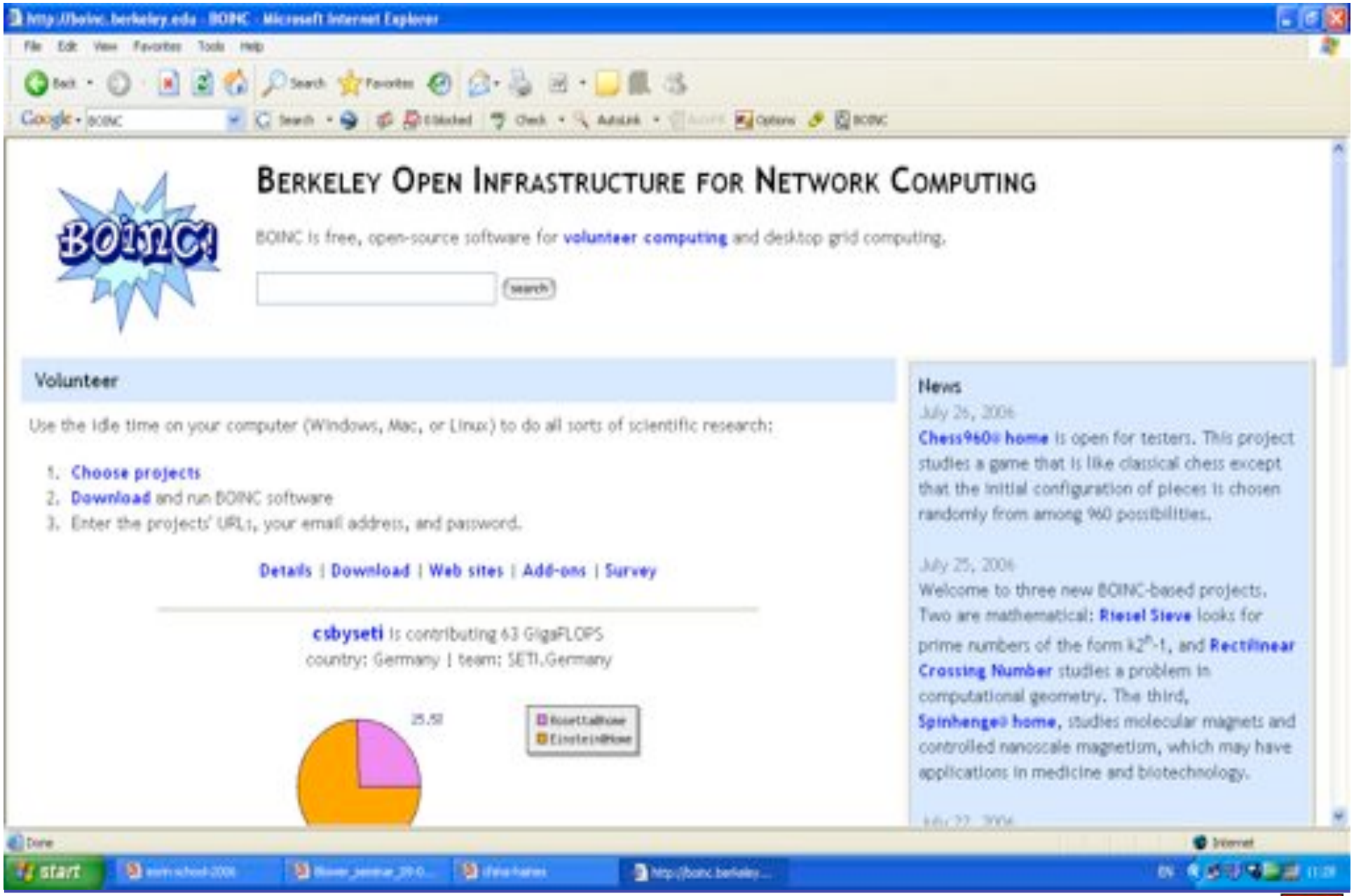
climateprediction.net



- Over 2,500 simulations over a 45 year period showed a possible temperature increase of 2 - 11°C by



- **Results from 2,579 15 year runs by climateprediction.net**
- **Results from 127 30 year runs of the Hadley model on the Met Office supercomputer**


BERKELEY OPEN INFRASTRUCTURE FOR NETWORK COMPUTING
BOINC is free, open-source software for volunteer computing and desktop grid computing.
search
Volunteer
Use the idle time on your computer (Windows, Mac, or Linux) to do all sorts of scientific research:
1. Choose projects
2. Download and run BOINC software
3. Enter the projects' URLs, your email address, and password.
Details | Download | Web sites | Add-ons | Survey
csbyseti is contributing 63 GigaFLOPS
country: Germany | team: SETI.Germany
25.58
News
July 26, 2006
Chess960® home is open for testers. This project studies a game that is like classical chess except that the initial configuration of pieces is chosen randomly from among 960 possibilities.
July 25, 2006
Welcome to three new BOINC-based projects. Two are mathematical: Riesel Sieve looks for prime numbers of the form k2^n-1, and Rectilinear Crossing Number studies a problem in computational geometry. The third, Spinhenge@ home, studies molecular magnets and controlled nanoscale magnetism, which may have applications in medicine and biotechnology.

NERC Data Grid (NDG)

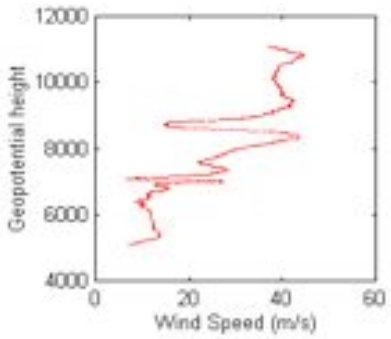
- NERC e-Science project led by British Atmospheric Data Centre (BADC)
- Developing software for data **discovery** and **delivery**
- Data will be distributed between NDG and other groups (NDG won't hold everything)
- Vast diversity of data types (all NERC data!)
- Rigidly standards-based (ISO)
  - Metadata is all-important: enables data discovery
  - Have created CSML (Climate Science Markup Language) – describes 7 feature types
- Producing whole array of OGC-compliant Web Services
  - Key task is to add proper security

Climate Science Markup Language CSML

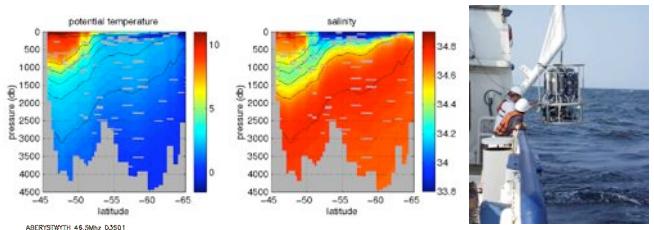
NATURAL ENVIRONMENT RESEARCH COUNCIL
ENVIRONMENTAL SYSTEMS SCIENCE CENTRE

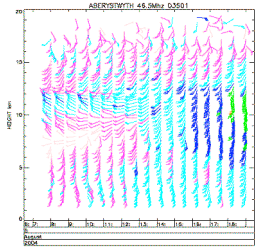
NDG

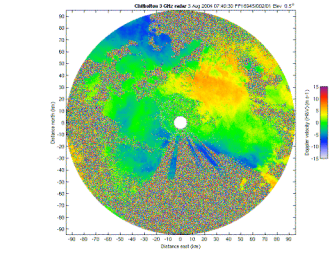
ProfileFeature



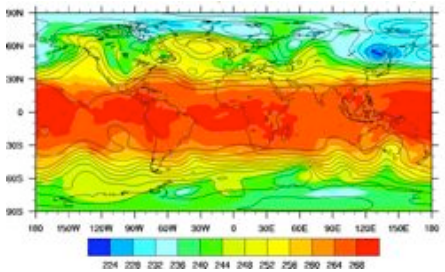
ProfileSeriesFeature







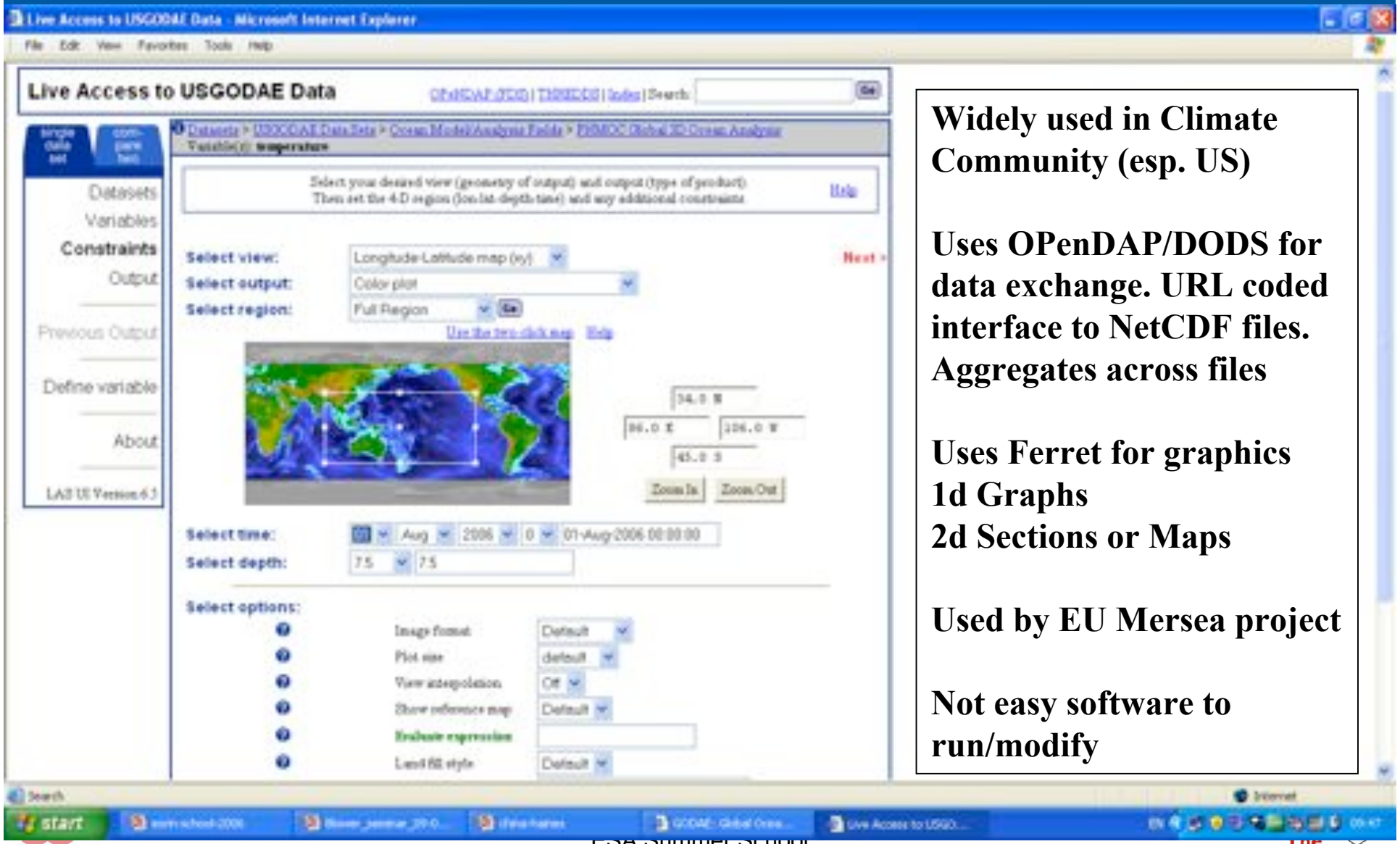
GridFeature



Visualising Environmental Data on the Internet

- Live Access Servers
- GODIVA Project
- Geographical Information Systems GIS
- OpenGIS Consortium (OGC) Standards
- Google Maps and Google Earth
- Open Source Counterparts

Live Access Server



Widely used in Climate Community (esp. US)

Uses OPenDAP/DODS for data exchange. URL coded interface to NetCDF files. Aggregates across files

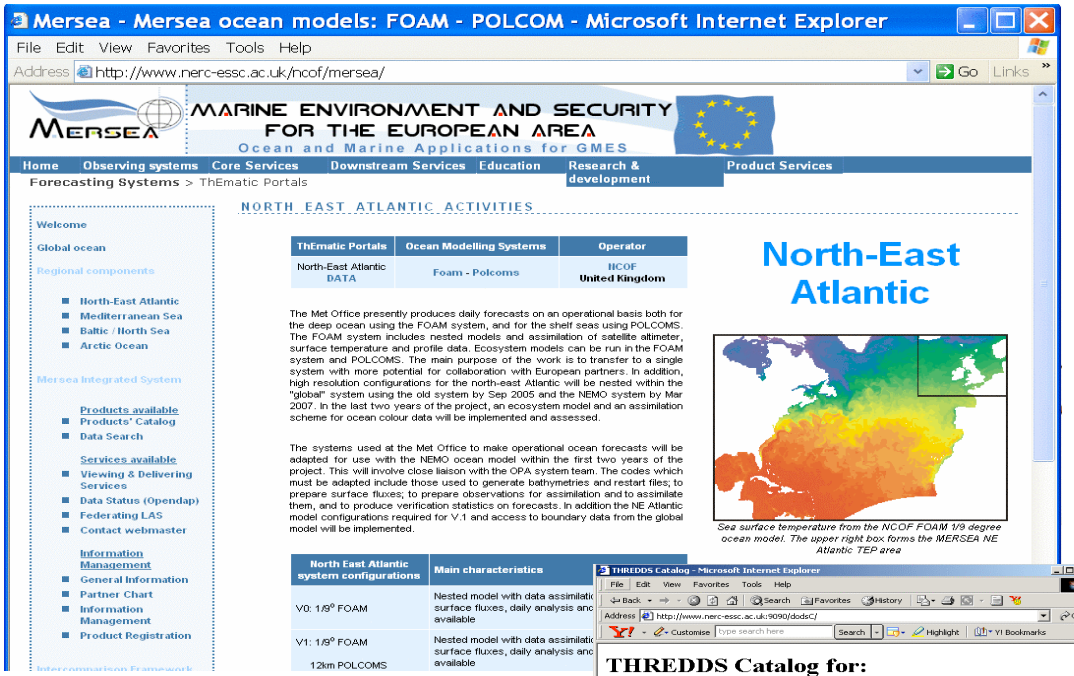
Uses Ferret for graphics
1d Graphs
2d Sections or Maps

Used by EU Mersea project

Not easy software to run/modify

Mersea: NE Atlantic ThEmatic Portal (TEP)

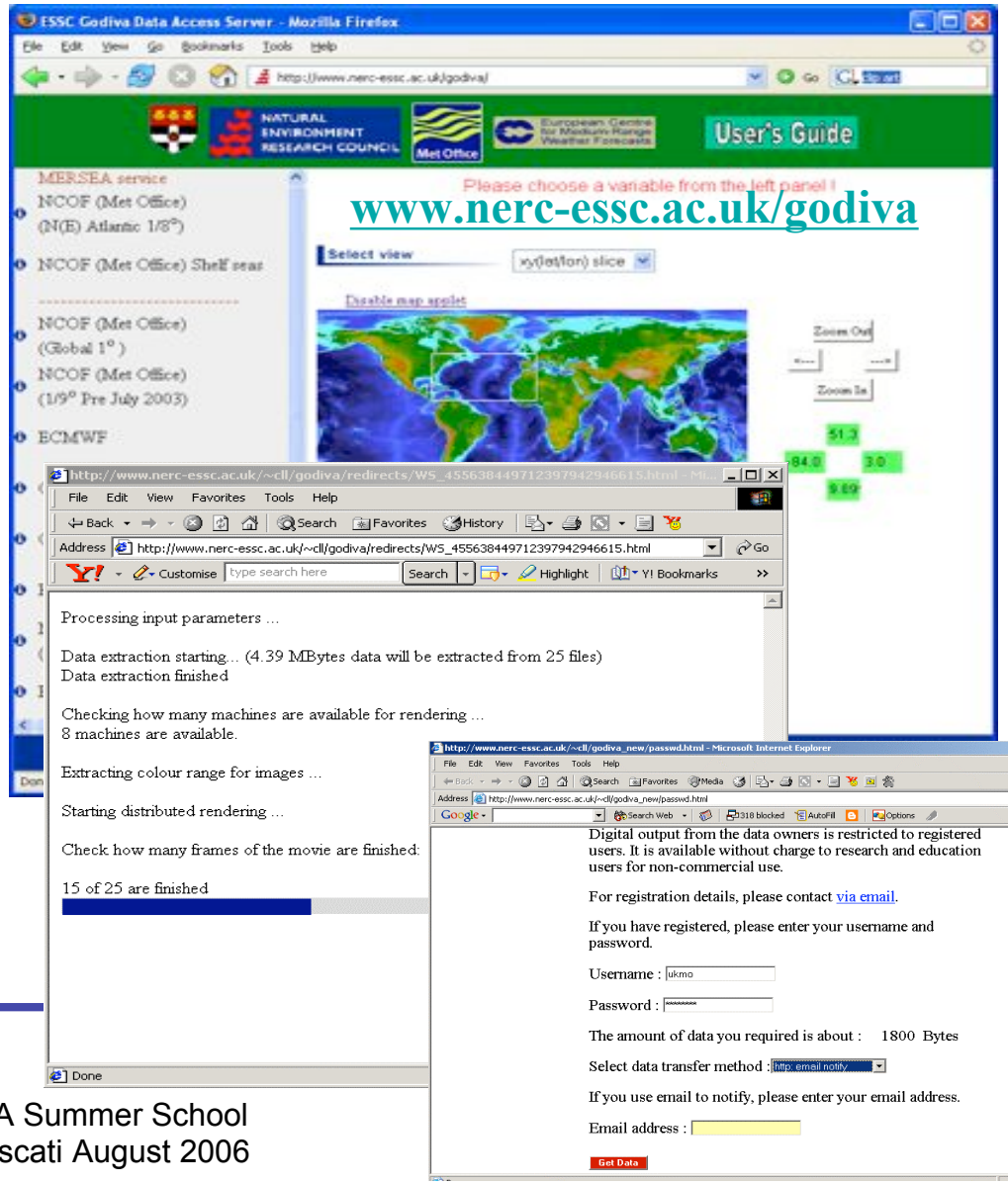
- In UK NCOF (National Centre for Ocean Forecasting) responsible for NE Atlantic Forecasting and Services within Mersea
- Range of models from 1/8° to 1 nm
- OPenDAP Server with 7 day forecasts
- Central Mersea Live Access Server
- Development and presentation of Application Services during EU ECOOP

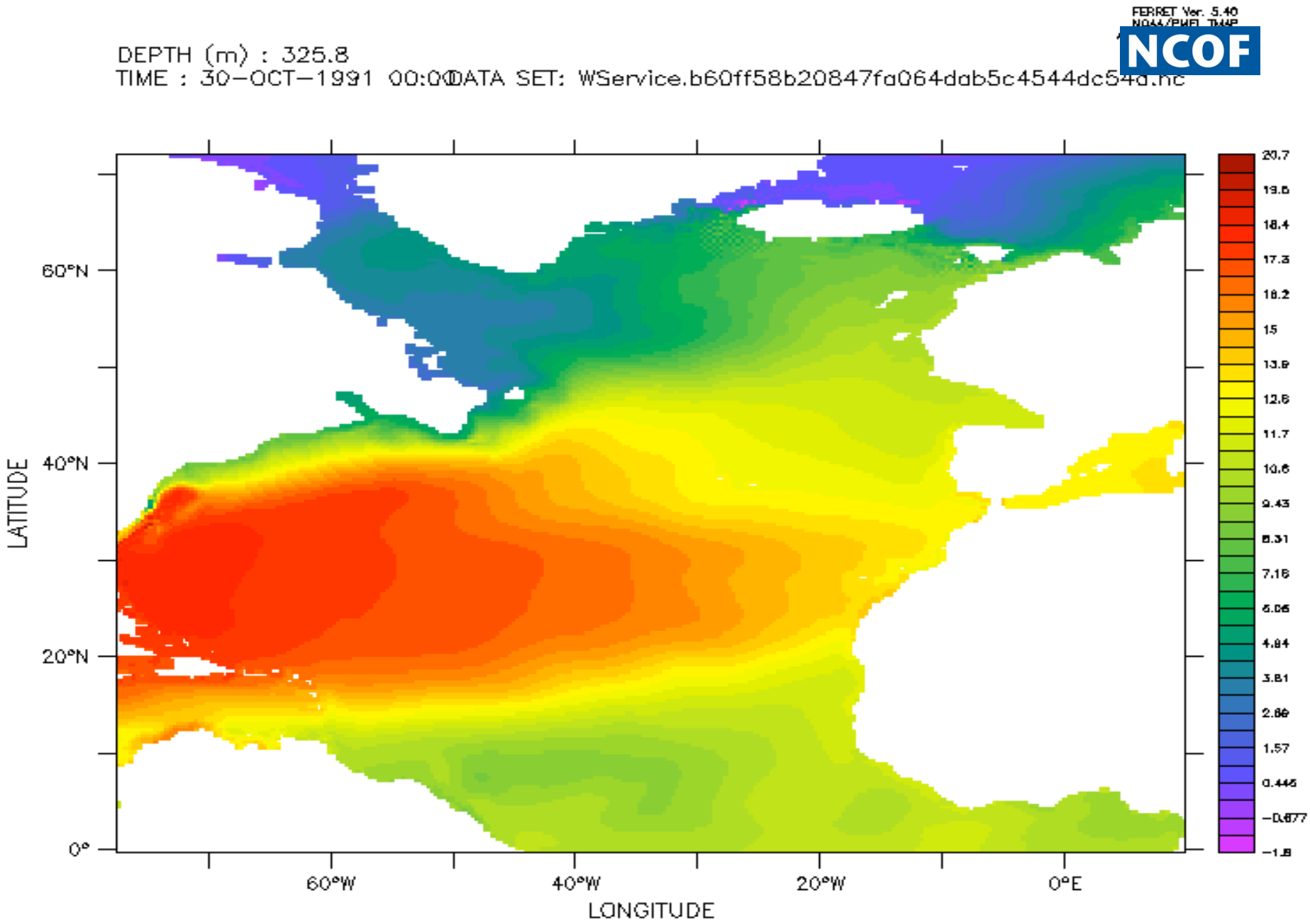


www.nerc-essc.ac.uk/ncof/mersea

GODIVA Project

- The GODIVA Web portal provides a graphical interface to data at Reading Univ.
- Web Service GADS (Grid Access to Data) instead of OPenDAP, Woolf et al (2003)
- Access to Operational Met Office ocean forecasts (last 5 years)
- Users can make simple visualisations
  - pictures and movies (uses Grid computing)
  - Secure data download
  - GADS WS can be accessed directly by third party software

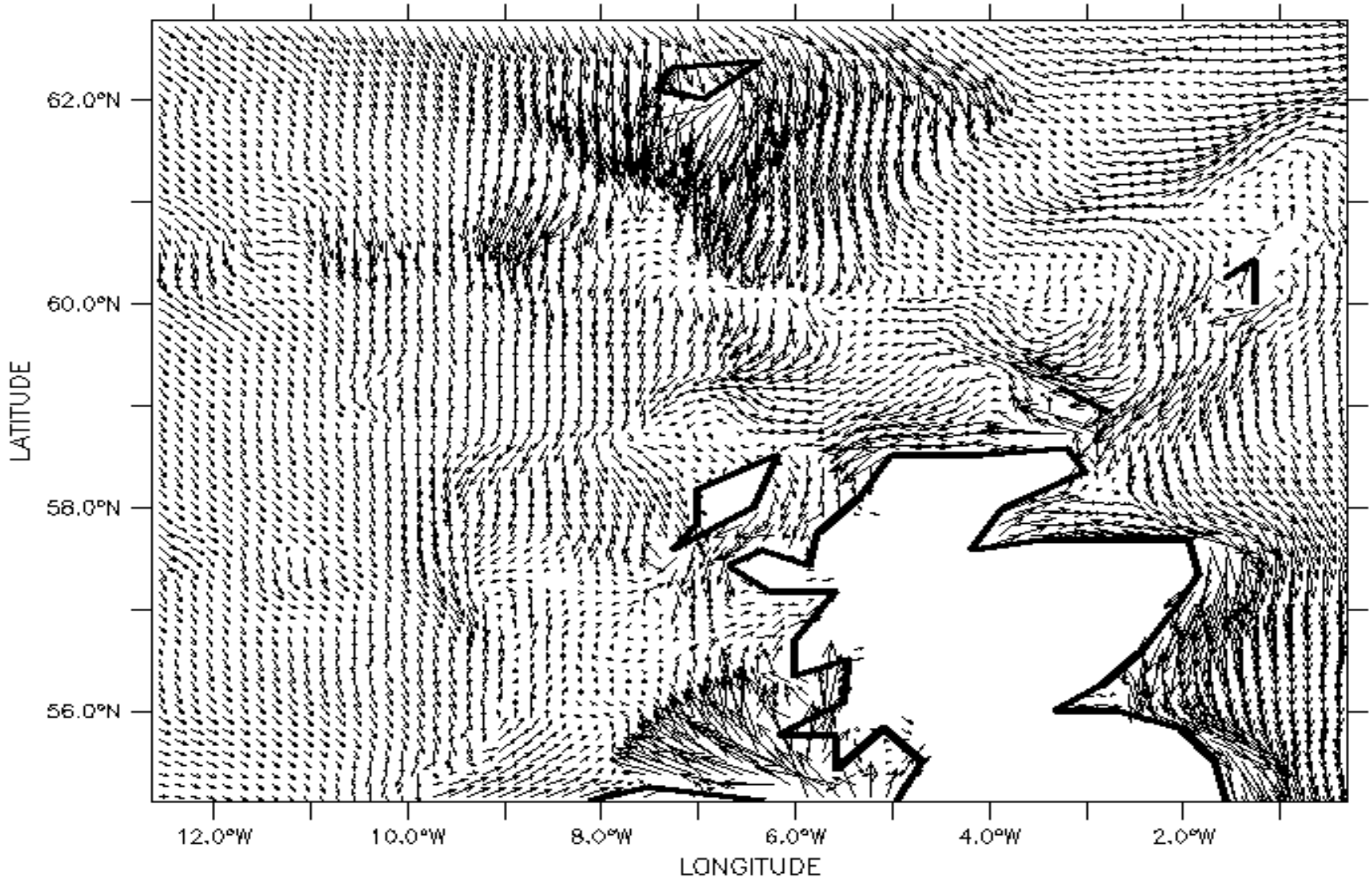



FERRET Ver. 5.40
NOAA/PMEL TMAP
NCOF
DEPTH (m) : 325.8
TIME : 30-OCT-1991 00:00
DATA SET: WService.b60ff58b20847fa064dab5c4544dc54d.nc
60°N
40°N
20°N
0°
LATITUDE
60°W
40°W
20°W
0°E
LONGITUDE
20.7
13.6
18.4
17.3
16.2
15
13.9
12.8
11.7
10.6
9.43
8.31
7.19
6.06
4.94
3.81
2.69
1.57
0.445
-0.677
-1.8
SEA_WATER_POTENTIAL_TEMPERATURE (C)

FERRET Ver. 5.40
NOAA/PMEL TMAP

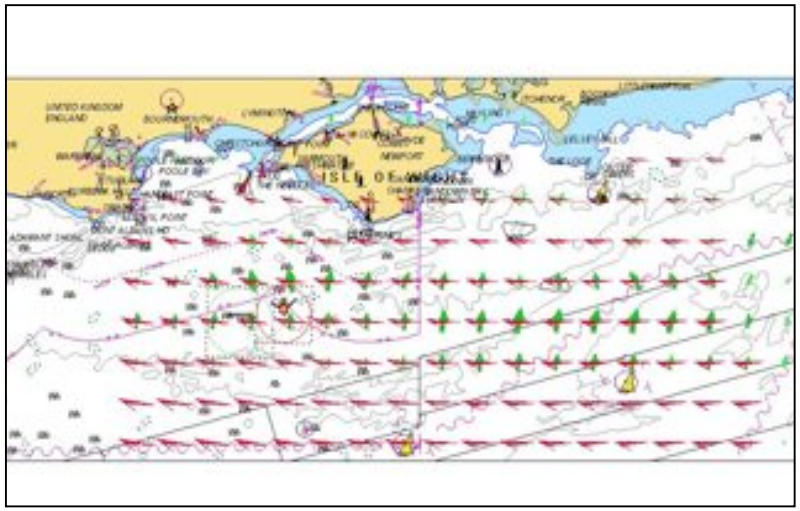
NCOF

TIME : 10-FEB-2006 01:00



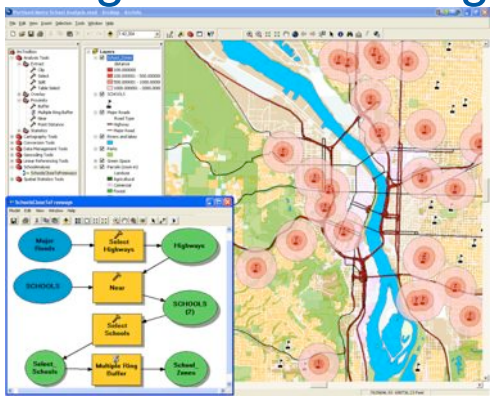
GADS application: Search and Rescue

- British Maritime Technology produce software (SARIS) to help the Coastguard with Search and Rescue
- Predicts drift patterns of people and objects that have fallen overboard
  - This significantly cuts the time to rescue
- Have worked with BMT to produce prototype that uses live Met Office data from GADS to improve its predictions
  - Uses forecasts of surface winds and surface currents
- Can also be applied to oil spills



Geographical Information Systems (GIS)

- Many companies produce GIS software for manipulating and visualizing geographical data
  - e.g. ArcInfo, Maptitude, many more
  - Big business!
- Very sophisticated and powerful
  - Spatial statistics, geoprocessing, mapping…
  - e.g. identify high-risk flood zones, assess effectiveness of ambulance centres etc…
- Historically very map-oriented (2-d or "2.5d")
  - Hence not so useful in ocean/atmosphere sciences (need 4-d)
- Vendors typically used proprietary formats and interfaces
  - Users "locked in" to a particular vendor, hard to share information
- The Open Geospatial Consortium (OGC) is addressing these issues



OGC Web Services

Web Service	Purpose
Web Map Server (WMS)	Serves map images (cf. Streetmap, Multimap)
Web Feature Server (WFS)	Serves geographical features (roads, rivers, hospital locations etc)
Web Coverage Server (WCS)	Serves multidimensional data (e.g. numerical model output)
Web Processing Server (WPS)	Processes data
Lots more in pipeline!	

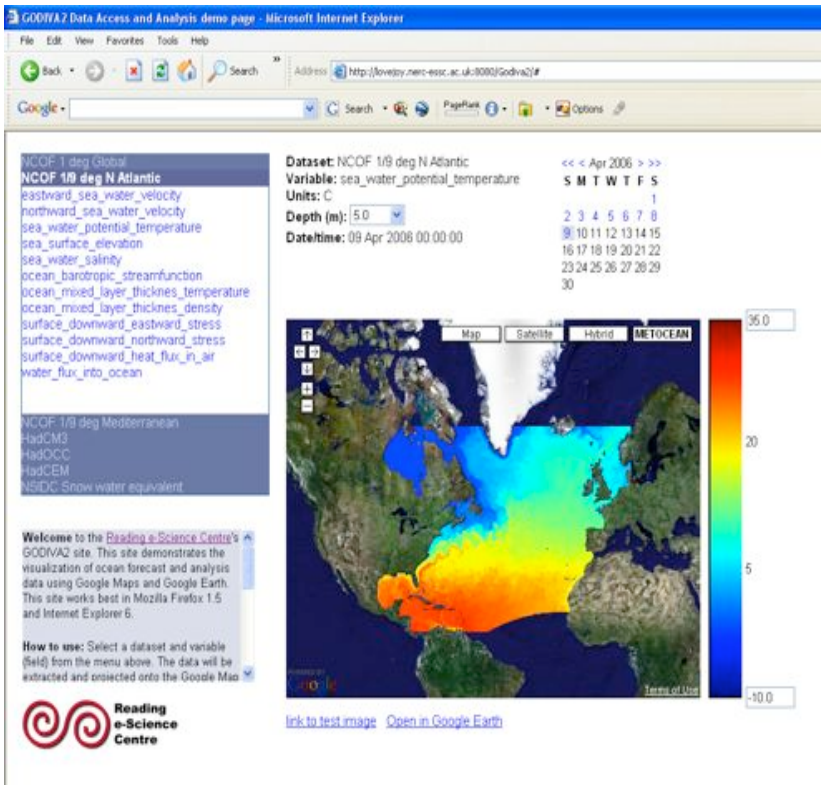
(roughly in decreasing order of maturity)

Services can be composed to create a distributed geospatial application

The GODIVA2 Server

Reading e-Science Centre

- Web Portal for Google Maps/Earth linked to GADS library
- Renders images from gridded NetCDF data
- Overlay data from different URLs in same image
- Highly interactive (pan & zoom)
- Only Mercator projections (Google maps)
- Loads javascript from Google site (API may change)
- Developing projects with
  - National Centre for Ocean Forecasting NCOF
  - EU Mersea (GMES Op. Oceanog.)
  - UK Hydrographic Office (Charts)
  - BMT Commercial Maritime Services



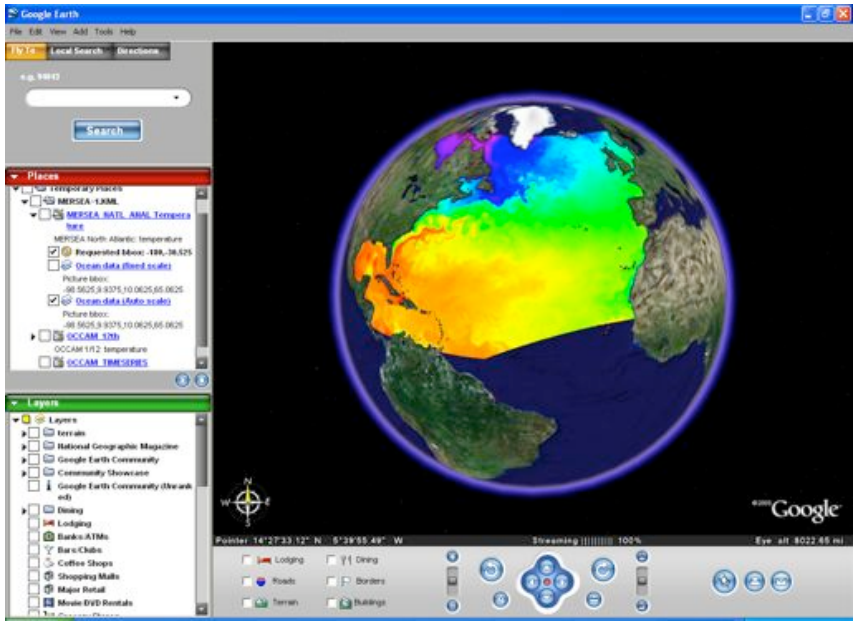
http://lovejoy.nerc-essc.ac.uk:8080/Godiva2

Google Earth

Reading e-Science Centre

- Can initiate from GODIVA2 portal
- Free closed-source executable for Windows, Mac, Linux
- Multiple GIS Overlays
- Reads images using KML metadata

Examples



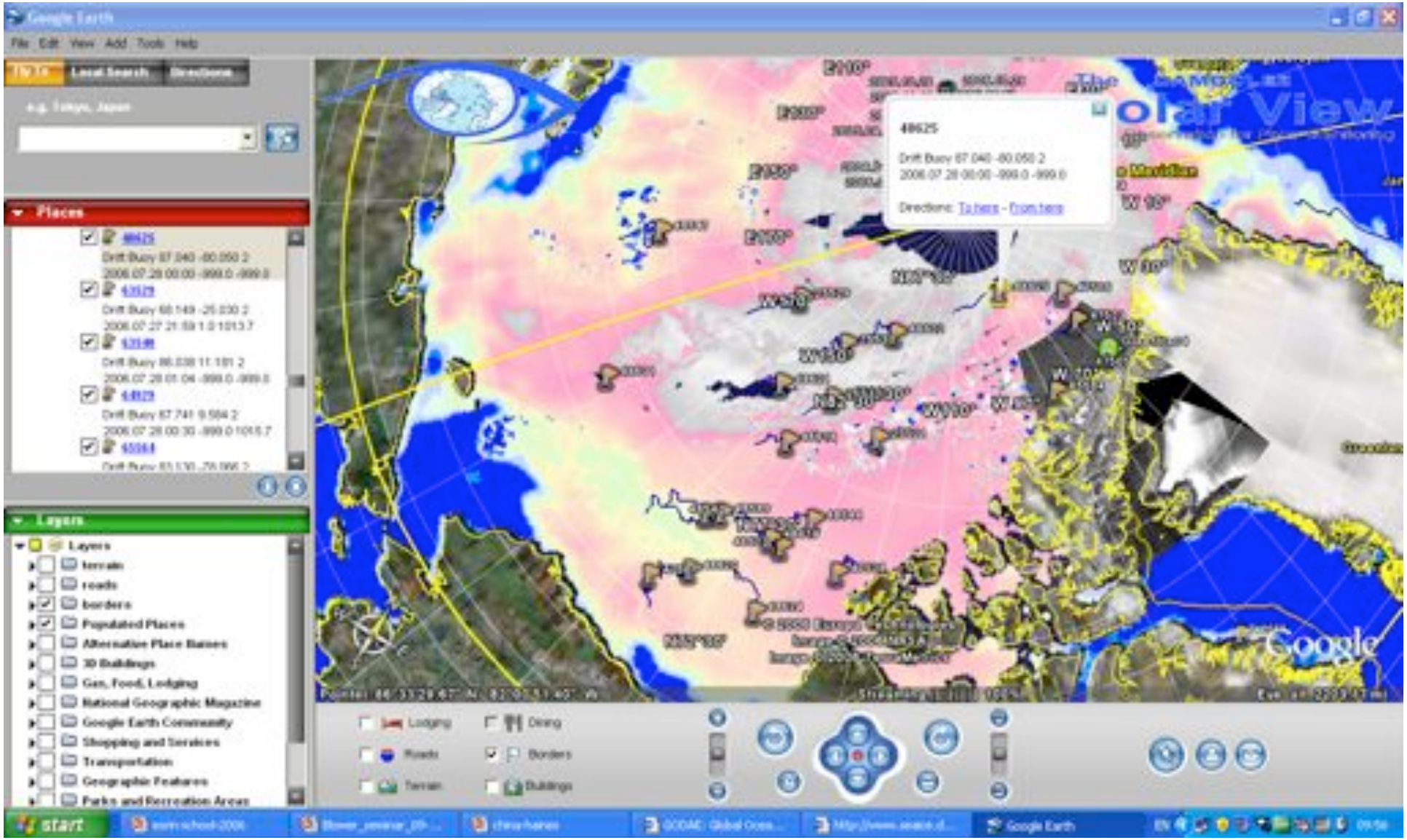
- Only Plate Carrée images
- Support for vector and raster data
- Support for vertical display increasing
- No support for time dimension
- Can add new datasets but currently can't change .exe

- But what about completely OpenSource licence software?

Example of a KML file

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://earth.google.com/kml/2.0">
<GroundOverlay>
  <description>Overlay shows Mount Etna erupting on
                                  July 13th, 2001.</description>
  <name>Large-scale overlay on terrain</name>
  <Icon>
    <href> http://bbs.keyhole.com/ubb/z0302a1700/etna.jpg </href>
  </Icon>
  <LatLonBox id="khLatLonBox751">
    <north>37.919</north>
    <south>37.465</south>
    <east>15.358</east>
    <west>14.601</west>
    <rotation>0</rotation>
  </LatLonBox>
</GroundOverlay>
</kml>
```

DAMOCLES Ice Buoys in Google E

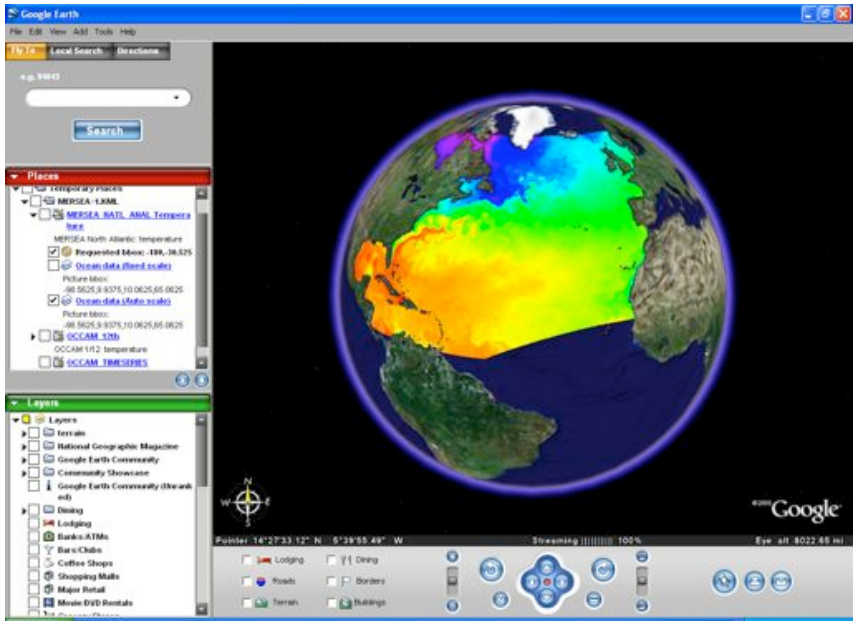


Google Earth

Reading e-Science Centre

- Can initiate from GODIVA2 portal
- Free closed-source executable for Windows, Mac, Linux
- Multiple GIS Overlays
- Reads images using KML metadata

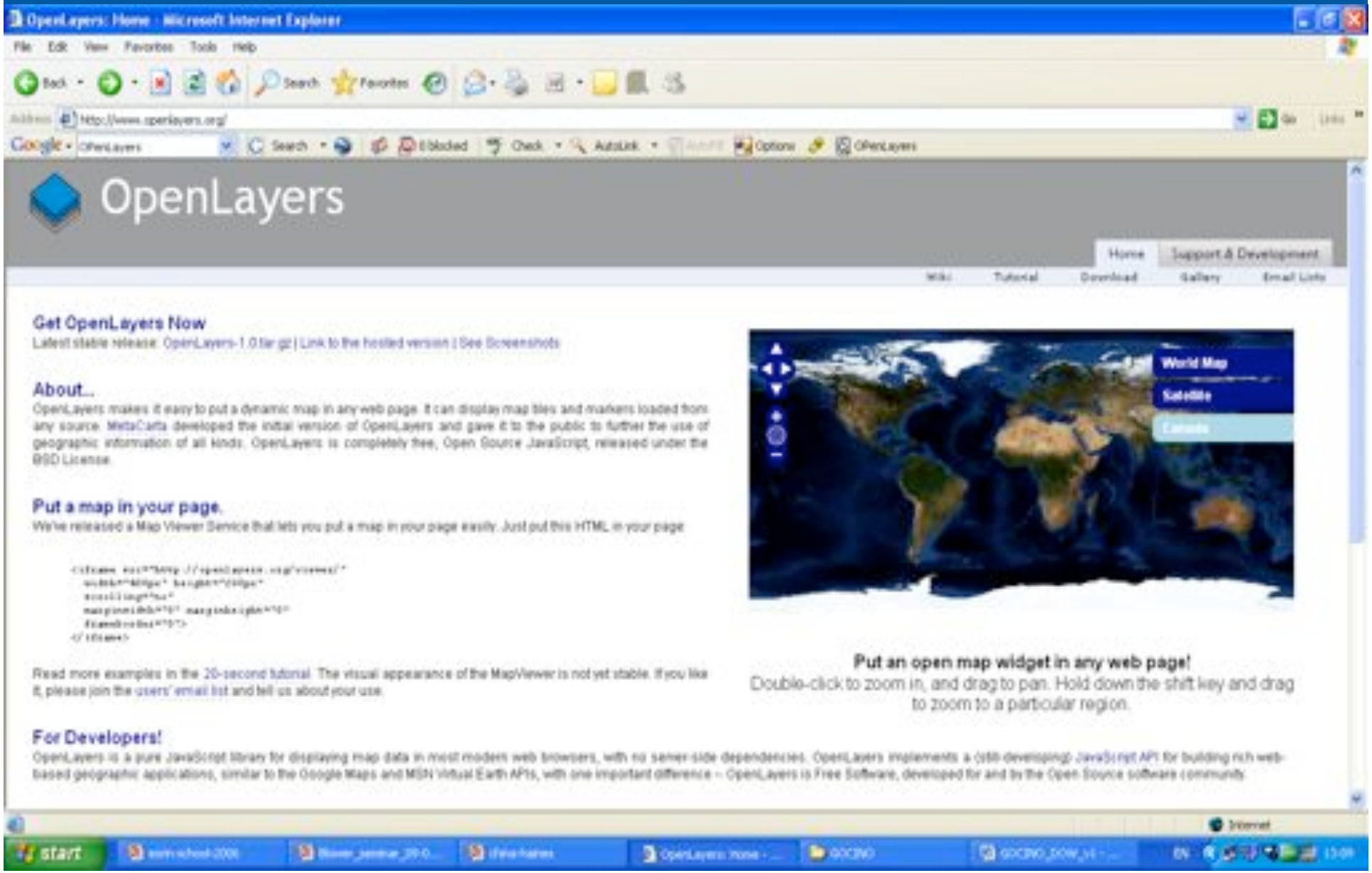
Examples

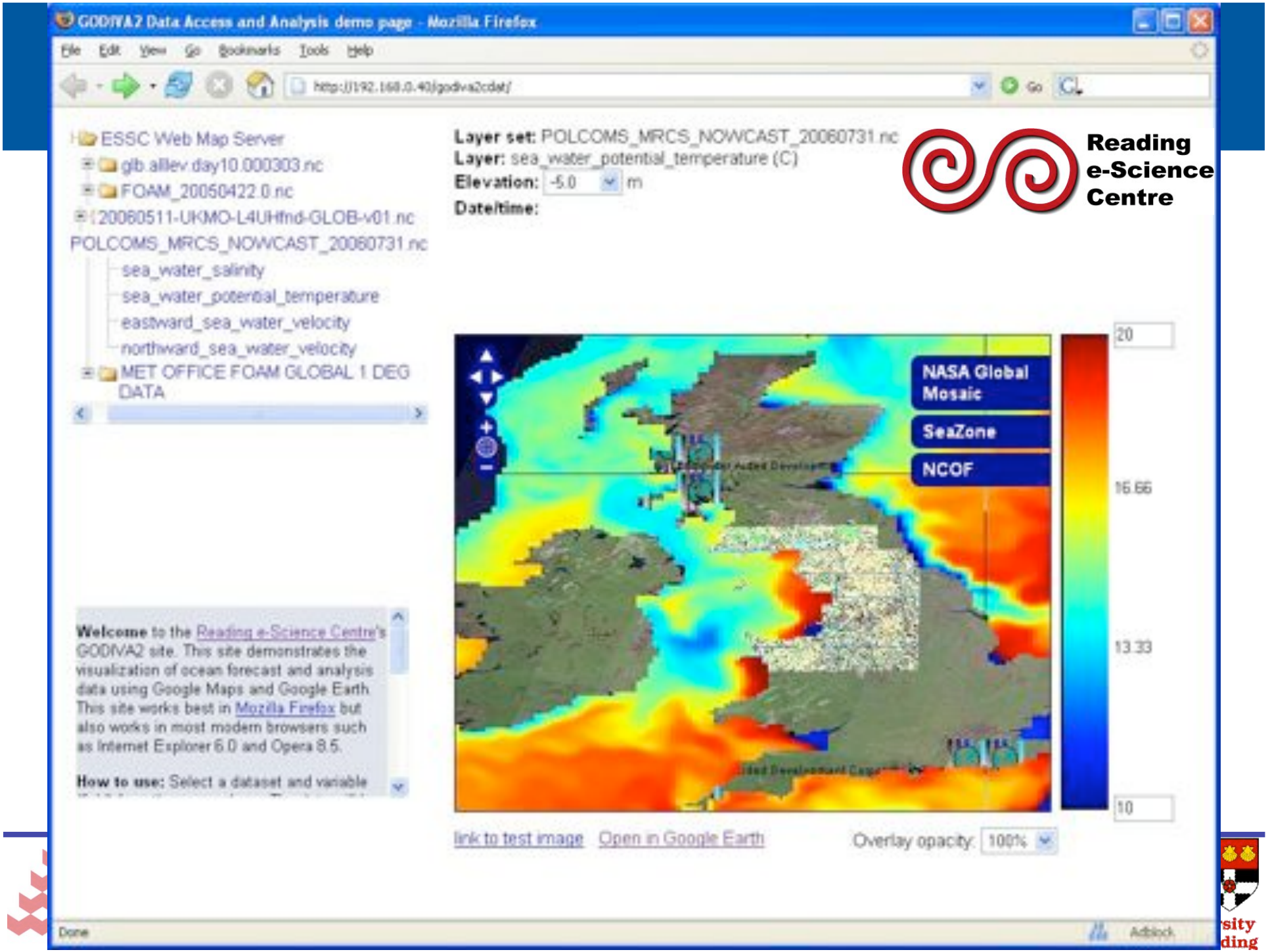


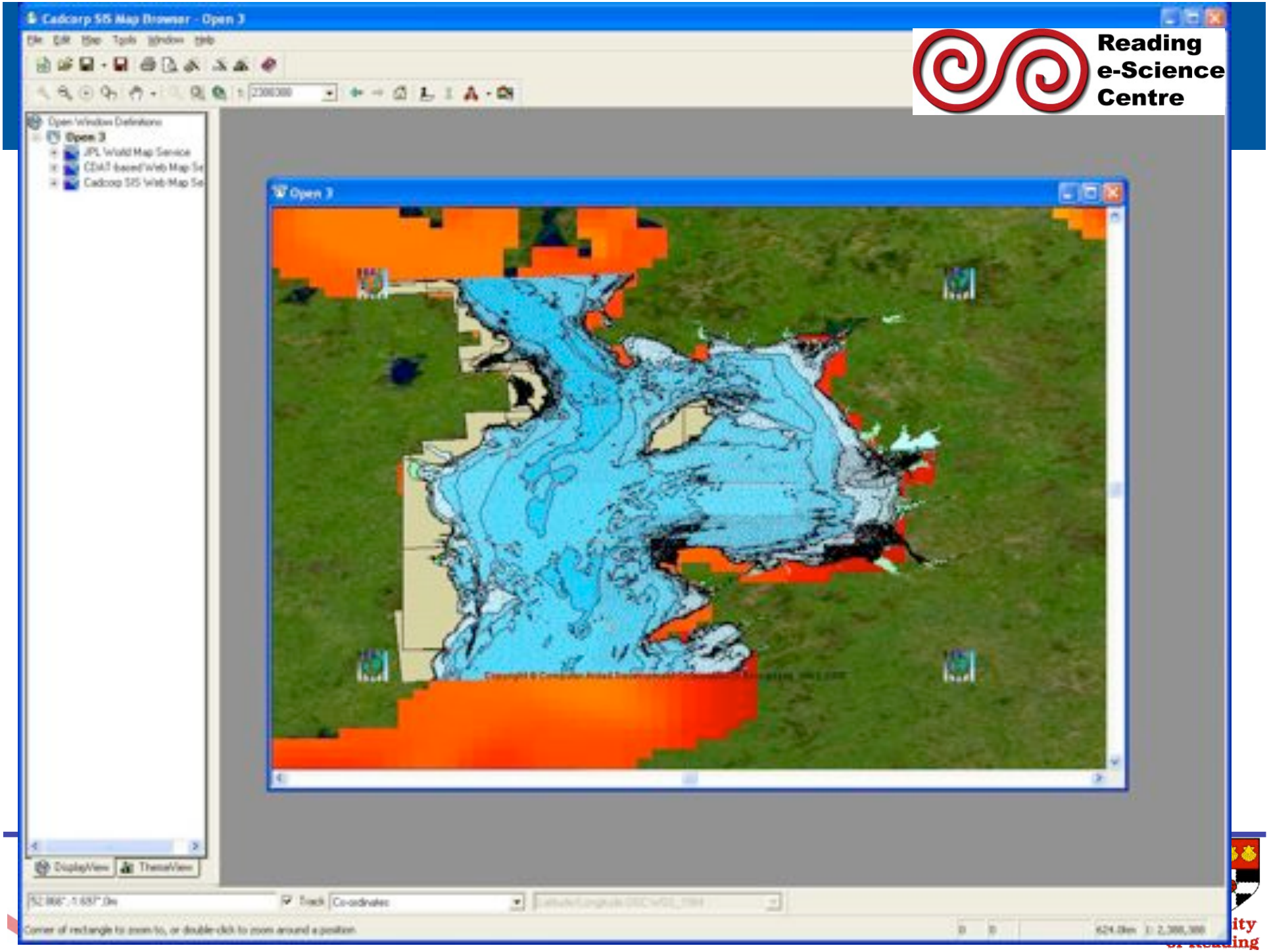
- Only Plate Carrée images
- Support for vector and raster data
- Support for vertical display increasing
- No support for time dimension
- Can add new datasets but currently can't change .exe

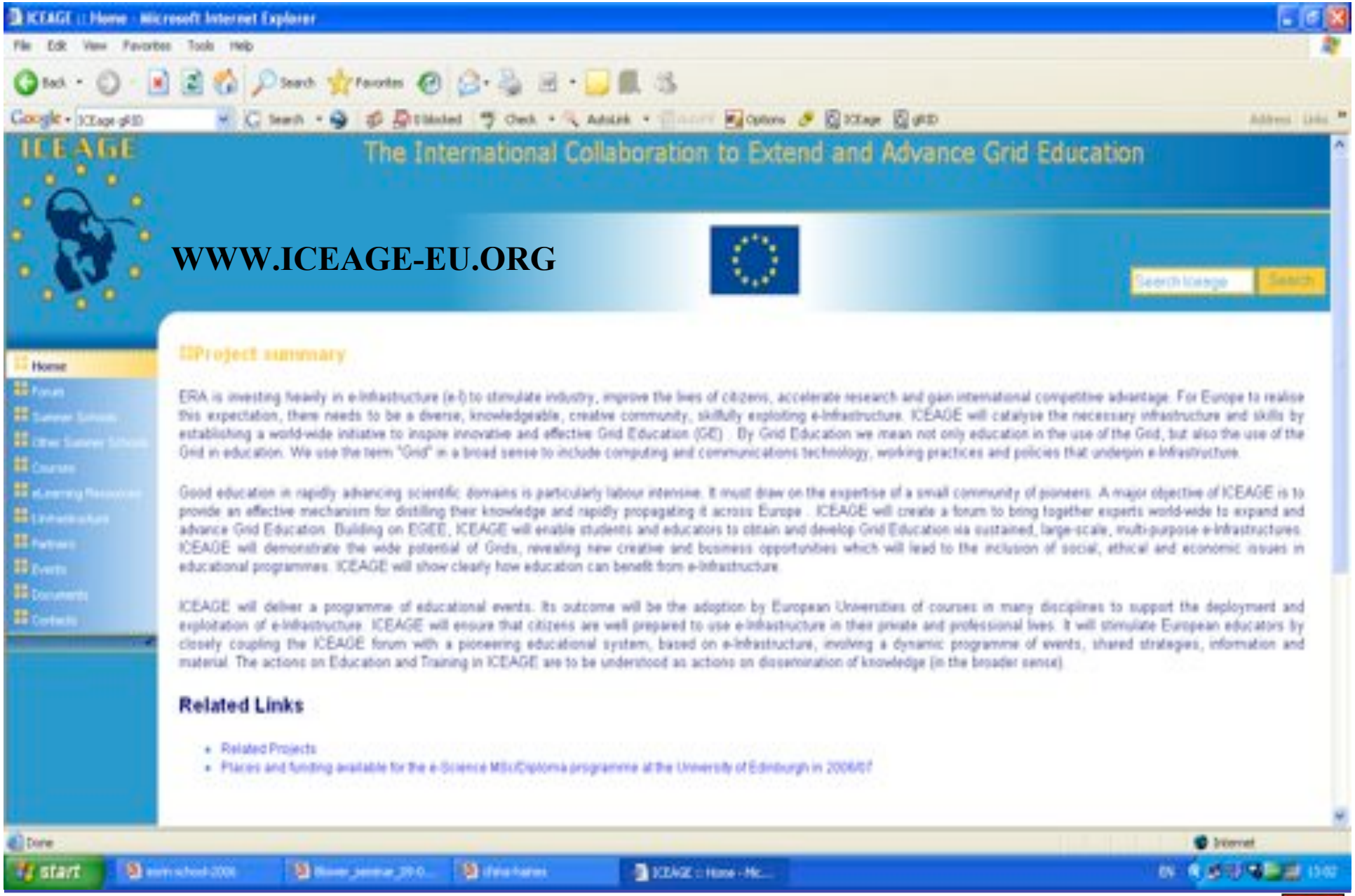
- But what about completely OpenSource licence software?

An OpenSource version of Google Maps




GODIVA2 Data Access and Analysis demo page - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
ESSC Web Map Server
glb allev day10.000303.nc
FOAM_20050422.0.nc
20060511-UKMO-L4UHfnd-GLOB-v01.nc
POLCOMS_MRCS_NOWCAST_20060731.nc
sea_water_salinity
sea_water_potential_temperature
eastward_sea_water_velocity
northward_sea_water_velocity
MET OFFICE FOAM GLOBAL 1 DEG DATA
Layer set: POLCOMS_MRCS_NOWCAST_20060731.nc
Layer: sea_water_potential_temperature (C)
Elevation: -5.0 m
Date/time:
Reading e-Science Centre
NASA Global Mosaic
SeaZone
NCOF
20
16.66
13.33
10
Welcome to the Reading e-Science Centre's GODIVA2 site. This site demonstrates the visualization of ocean forecast and analysis data using Google Maps and Google Earth This site works best in Mozilla Firefox but also works in most modern browsers such as Internet Explorer 6.0 and Opera 8.5.
How to use: Select a dataset and variable
link to test image
Open in Google Earth
Overlay opacity: 100%
Done
Adblock


Cadcorp SIS Map Browser - Open 3
Open 3
Reading
e-Science
Centre


ICEAGE :: Home - Microsoft Internet Explorer
ICEAGE
The International Collaboration to Extend and Advance Grid Education
WWW.ICEAGE-EU.ORG
Project summary
ERA is investing heavily in e-Infrastructure (e-I) to stimulate industry, improve the lives of citizens, accelerate research and gain international competitive advantage. For Europe to realise this expectation, there needs to be a diverse, knowledgeable, creative community, skilfully exploiting e-Infrastructure. ICEAGE will catalyse the necessary infrastructure and skills by establishing a world-wide initiative to inspire innovative and effective Grid Education (GE) . By Grid Education we mean not only education in the use of the Grid, but also the use of the Grid in education. We use the term "Grid" in a broad sense to include computing and communications technology, working practices and policies that underpin e-Infrastructure.
Good education in rapidly advancing scientific domains is particularly labour intensive. It must draw on the expertise of a small community of pioneers. A major objective of ICEAGE is to provide an effective mechanism for distilling their knowledge and rapidly propagating it across Europe . ICEAGE will create a forum to bring together experts world-wide to expand and advance Grid Education. Building on EGEE, ICEAGE will enable students and educators to obtain and develop Grid Education via sustained, large-scale, multi-purpose e-Infrastructures. ICEAGE will demonstrate the wide potential of Grids, revealing new creative and business opportunities which will lead to the inclusion of social, ethical and economic issues in educational programmes. ICEAGE will show clearly how education can benefit from e-Infrastructure.
ICEAGE will deliver a programme of educational events. Its outcome will be the adoption by European Universities of courses in many disciplines to support the deployment and exploitation of e-Infrastructure. ICEAGE will ensure that citizens are well prepared to use e-Infrastructure in their private and professional lives. It will stimulate European educators by closely coupling the ICEAGE forum with a pioneering educational system, based on e-Infrastructure, involving a dynamic programme of events, shared strategies, information and material. The actions on Education and Training in ICEAGE are to be understood as actions on dissemination of knowledge (in the broader sense).
Related Links
Related Projects
Places and funding available for the e-Science MSc/Diploma programme at the University of Edinburgh in 2006/07
Home
Forum
Courses
Partners
Events
Documents
Contacts
start

Conclusions: Getting Involved in Grid

- Stick to OGC standards if you are planning to handle Environmental data on the internet
- If you are keen software developer join OS groups like OpenLayers and help develop capability
- If you are interested in computational grid or software for setting up grids try
  - www.nesc.ac.uk or Open Middleware Infrastructure Institute (OMII) www.omii.ac.uk
- If would like to try using Grid computing set up by others try
  - www.eu-egee.org or seek access to the ESA Grid
- UK has an environmental e-science institute in Cambridge which runs training workshops specifically for environmental scientists www.niees.ac.uk; usually free to attend
- Or go to another European summer school! **www.iceage-eu.org**

End of third Lecture