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

Internet

Travel Agent
Booking

Reservation Service
Reserve seat
Success

Accounting Service
Debit
Success

Bank

Airline
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

http://gridportal.hep.ph.ic.ac.uk/rtm/
EU Grid project
UK National Grid Service

Welcome to the new look National Grid Service website!

As part of the on-going refocusing of the Grid Operations Support Centre, the GOSC website has changed both its content and imagery to reflect that of the NGS.

The UK's National Grid Service (NGS) provides a core e-infrastructure that underpins UK e-Research, providing standardised access to data management and compute resources and supporting collaborative computing across the UK.

The NGS also provides a national "gateway" to international collaborations through collaboration with related e-infrastructures internationally. The National Grid Service, funded by JISC and DfE, was created in October 2003. The service entered full production in September 2004 at which point the Grid Operations Support Centre was created with support from the EPSRC core e-Science programme. The GOSC was built upon earlier work of the DfI funded Grid Support Centre and the Grid Engineering Taskforce. Together the GOSC and
The many scales of grids

International instruments,

National datacentres, HPC, instruments

Institutes’ data;
Condor pools, clusters

Wider collaboration
greater resources

International grid (EGEE)

National grids (e.g. National Grid Service)

Regional grids (e.g. White Rose Grid) 3 UK Universities.

Campus grids

Desktop
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)
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
Measuring the Vegetation Change in Europe

The ENVIROS MERIS (Medium Resolution Imaging Spectrometer Instrument) Global Vegetation Index (GVI) was designed to assess and monitor the state and health of terrestrial vegetation using MERIS measurements acquired in space (at the so-called top of atmosphere). Using the blue, red, and near-infrared spectral bands of MERIS, as well as information on the angular geometry of illumination and observation, this algorithm was optimized to deliver the fraction of absorbed Photosynthetically Active Radiation (FAPAR) in terrestrial plant canopies. The FAPAR is used as an indicator of the state and productivity of vegetation and has been recognized as a fundamental surface parameter for environmental studies by international organizations. This quantity represents the fraction of the solar energy which is absorbed by vegetation and therefore plays the role of a battery during the plant photosynthetic process.

The grid on-Demand Services and Infrastructure generates maps of GVI for Europe on a monthly basis with a Time Composite Algorithm which selects, for each location, the value of GVI actually measured during the period that is the closest to the temporal average over the composing period for that location.
Welcome to the submission area for Category 1 data users on the Earth Observation GRID Processing on Demand environment.

ESA is offering all scientists 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 19 proposals will be selected for implementation in the frame of this Call.

- Invitation letter
- Call Main Text

Terms and Conditions for the Utilisation of Data under the ESA Category 1 scheme

Guidelines for the submission of proposals

Proposal Submission

More about G-POD

Contact the ESA ESPO Help and Online Desk for further information.
• 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.
• Over 2,500 simulations over a 45 year period showed a possible temperature increase of 2 - 11°C by 2050.

• 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
BOINC is free, open-source software for volunteer computing and desktop grid computing.

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

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

http://ndg.nerc.ac.uk/

ESA Summer School
Frascati August 2006
Climate Science Markup Language CSML

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.
- 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

www.nerc-essc.ac.uk/godiva

ESA Summer School
Frascati August 2006
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

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

(roughly in decreasing order of maturity)

Services can be composed to create a distributed geospatial application
The GODIVA2 Server

- 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

- 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
<?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

- 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
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