Provenance, metadata, and e-infrastructure to support climate science

Bryan Lawrence
Rutherford Appleton Laboratory
reporting the efforts of dozens of other folks in major international projects
including, but not limited to
CMIP5 (Taylor, Stoufer)
Metafor (Guilyardi), IS-ENES (Joussaume)
Earth System Grid, Earth System Curator
(Balaji, DeLuca, Foster, Middleton, Williams)
(none of whom were consulted about the content of this talk)
&
The Global Organisation for Earth System Science Portals, and the new Earth System Grid Federation
Outline

• The Climate Problem
  – Data Generating Infrastructure and the need for metadata
  – Evaluating Australia in CMIP3 Climate Models
  – Climate Model primer
  – The problem with understanding the differences between models and the simulations they produce.

• A Brief introduction to Metafor
• An introduction to the CMIP5 Information Ecosystem
  – Aims and objectives of CMIP5
  – Global problem: Global simulations simulated globally.
  – Global Deployment of information systems.
    • The Earth System Grid Federation.
  – Quality Control and Assessment in CMIP5 and ESGF
  – Bringing the information flow together

• A tour through the CMIP5 information implementation.
  – Access control in an open world.
Outline

• The Climate Problem
  – Data Generating Infrastructure and the need for metadata
  – Evaluating Australia in CMIP3 Climate Models
  – Climate Model primer
  – The problem with understanding the differences between models and the simulations they produce.

• A Brief introduction to Metafor

• An introduction to the CMIP5 Information Ecosystem
  – Aims and objectives of CMIP5
  – Global problem: Global simulations simulated globally.
  – Global Deployment of information systems
    • The Earth System Grid Federation.
  – Quality Control and Assessment in CMIP5
  – Bringing the information flow together

• A tour through the CMIP5 information implementation
  – Access control in an open world.

Talk 3 of 3:
1: Information Network interoperability
2: Metadata Futures
   - including more details of greater role for RDF in the work discussed here

All to be on my blog
http://home.badc.rl.ac.uk/lawrence/talks
... all seems rather simple doesn't it?
Nice consumable curves ...
Enough for mitigation policy perhaps, but enough for adaptation policy?
In the beginning: observations

All linked up, with global data distribution.
World Meteorological Organisation have been doing e-infrastructure for years!

Images: from J. Lafeuille, 2006
NERC Observatories and Sensor Networks
NERC Mobile Research Sensors

Slide courtesy of Alan Gadian, NCAS
Old Weather/New Results

(Modern max & min over HMS Dorothea 1818
Brohan et al 2010)
Using data collected under the umbrella of the Atmospheric Circulation Reconstructions of the Earth (ACRE) initiative

Assimilating (only) surface observations of synoptic pressure, monthly sea surface temperature and sea ice distribution to produce

Data available from Jan 1871 to 2008 from NOAA ESRL … but:

1 GB/year/variable, 56 ensemble members (+mean and spread), 10 variables (there are more), 120 years = 70 TB …

http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html
• Neither of the last two examples would be possible without metadata
  - Ship logbooks with location, time, along with measurements
  - (Actually the measurements themselves were “metadata” for the ship logs.)
  - Station data with information about location and calibration

• But both demonstrate problems with lack of metadata too:
  - How were those ship measurements made, and with what accuracy?
  - Did that station move, and if so, did anyone write it down (movements often lead to discontinuities in data records)

• Research data systems generate a wealth of information, usually recorded for a specific task.
  - But that information, with sufficient information, can be repurposed, reinterpreted, and reused!

• But the sheer amount of data can overwhelm one's ability to reuse if one can't get at basic facts as to what was done, how, and why!
The scalability of real metadata

... but even this sort of metadata can be invisible (and hence, useless), if it's not machine readable.

Humans can't deal with thousands of such things (at least not without crowd sourcing, and that only works for “interesting” tasks).

Metadata needs to be machine readable.

WMO/TD 1250 (2006) (pdf)

(Research instruments often don't bother with this level of info, to the detriment of reuse)
Humans and Big Data

A person working full time for a year has about 1500 hours to do something. Moore's Law won't change that.

(In the UK 220 working days a year is about standard. Let's remove about 20 days for courses, staff meetings etc... so that leaves about 200 days or, for a working day of 7.5 hours, a working year of about 1500 hours.)

- What does a 50 TB dataset mean?
  - A single lat/lon map might be of order 50 Kb... so we have of the order of 10 billion maps. So, if we look at each map for 10s, one individual could quality control those maps in, say, two thousand years of work! Bring on crowd sourcing... (but not all problems are sexy)

  *We will never look at all our data.*

  *We need to do automatic quality control on ingestion.*

  *We have to provide tools so users can select what they want not download entire datasets*

  **Tools need metadata!**

- If it takes 2 minutes to find something, and have a quick look at it and, say, extract a parameter name, you can process 45,000 items a year, but no human could do that full time (repetitive boredom)! (Maybe 30K in two years?)

  *So, particularly with respect to observational data, we can't manually reprocess our files to create new information about the data we hold... we have to automate... automation needs compliant metadata...*

**Storage costs going down; metadata costs going up!**
Climate – Delving Deeper …

IPCC Fourth Assessment Report:

(Sorry: not much agreement in AR4 - No stipling)

Spatial and temporal subsetting … statistics over models …

Figure 10.12. Multi-model mean changes in (a) precipitation (mm day⁻¹), (b) soil moisture content (%), (c) runoff (mm day⁻¹) and (d) evaporation (mm day⁻¹). To indicate consistency in the sign of change, regions are stipled where at least 80% of models agree on the sign of the mean change. Changes are annual means for the SRES A1B scenario for the period 2080 to 2099 relative to 1980 to 1999. Soil moisture and runoff changes are shown at land points with valid data from at least 10 models. Details of the method and results for individual models can be found in the Supplementary Material for this chapter.
So why was Australia not stippled?

*Interannual variability* means that when our projections need to start in the right state (and capture that variability correctly too).

*Model uncertainty* means that we may not believe our model(s) (any or all) have the relevant resolution and/or physics to capture important regional processes.

*Scenario uncertainty* means that we are not sure of the impact of different economic and emission futures.

(Australia is unlucky, some regions more predictable than others, global mean much more predictable than any region)

So what were the salient differences between the models? (Forget looking at the code, these models have millions of lines of code each!)
Rows: Models, and their output types. Columns: Experiments and Projections (Three layers of complexity: models, experiments, output ... each of which is itself complex)

**AR4: WG1 Table 10.4**

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Model, Country</th>
<th>Pre industri- control</th>
<th>Present day control</th>
<th>20th century Commit- ment</th>
<th>SRES A2</th>
<th>SRES A1B</th>
<th>SRES B1</th>
<th>1% to 2xCO2</th>
<th>1% to 4xCO2</th>
<th>Slab ocean control</th>
<th>2XCO2</th>
<th>AMIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BCC-CM1, China</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>BCCR-BCM2.0, Norway</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>CCSM3, USA *</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>CGCM3.1(T47), Canada</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>CGCM3.1(T63), Canada</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>CNRM-CM3, France</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>CSIRO-MK3.0, Australia</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>ECHAM5/MPI-OM, Germany</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>ECHO-G, Germany/Korea</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>FGOALS-g1.0, China</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>GFDL-CM2.0, USA</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>GFDL-CM2.1, USA</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>GISS-AOM, USA</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>GISS-EH, USA</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>GISS-ER, USA</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>INM-CM3.0, Russia</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>IPSL-CM4, France</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>MIROC3.2(hires), Japan</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>MIROC3.2(medres), Japan</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>MRI-CGCM2.3.2, Japan</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>PCM, USA</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>UKMO-HadCM3, UK</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>UKMO-HadGEM1, UK</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Digression: What is a model?

Primarily Mathematical (not statistical) representation of a complex system of climate processes.

Image: from J. Lafeuille, 2006
What is a Coupled Climate Model?

Answer: Lots of coupled partial differential equations solved via interactive numerical techniques. Grid resolution controls whether equations really represent processes or parameterised versions of them (which will have some statistical properties).
### State of the Art: Model Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>Institution</th>
<th>Atmosphere resolution</th>
<th>Ocean resolution</th>
<th>Length pctl</th>
<th>Length Ipcto2x</th>
<th>Length Ipcto4x</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSM3</td>
<td>NCAR (USA)</td>
<td>T85L26</td>
<td>1.125°×0.27°L40</td>
<td>230</td>
<td>150</td>
<td>n/a</td>
</tr>
<tr>
<td>CGCM3.1(T47)</td>
<td>CCCMA (Canada)</td>
<td>T47L31</td>
<td>1.85°×1.85°L29</td>
<td>500</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>CNRM-CM3</td>
<td>Meteo-France/CNRM (France)</td>
<td>T63L45</td>
<td>2°×0.5°L31</td>
<td>390</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>CSIRO-Mk3.0</td>
<td>CSIRO (Australia)</td>
<td>T63L18</td>
<td>1.875°×0.84°L31</td>
<td>380</td>
<td>10</td>
<td>n/a</td>
</tr>
<tr>
<td>ECHAM5/MPI-OM</td>
<td>MPI-M (Germany)</td>
<td>T63L31</td>
<td>1.5°×0.5°L40</td>
<td>332</td>
<td>100</td>
<td>81</td>
</tr>
<tr>
<td>FGOALS-g1.0</td>
<td>LASG/IAP (China)</td>
<td>T42L26</td>
<td>1°×1°L33</td>
<td>150</td>
<td>80</td>
<td>n/a</td>
</tr>
<tr>
<td>GFDL-CM2.0</td>
<td>GFDL (USA)</td>
<td>2.5°×2°L24</td>
<td>1°×0.33°L50</td>
<td>500</td>
<td>100</td>
<td>160</td>
</tr>
<tr>
<td>GFDL-CM2.1</td>
<td>GFDL (USA)</td>
<td>2.5°×2°L24</td>
<td>1°×0.33°L50</td>
<td>500</td>
<td>150</td>
<td>160</td>
</tr>
<tr>
<td>GISS-AOM</td>
<td>NASA/GISS (USA)</td>
<td>4°×3°L12</td>
<td>4°×3°L16</td>
<td>251</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>GISS-EH</td>
<td>NASA/GISS (USA)</td>
<td>5°×4°L20</td>
<td>2°×2°L16</td>
<td>500</td>
<td>80</td>
<td>150</td>
</tr>
<tr>
<td>GISS-ER</td>
<td>NASA/GISS (USA)</td>
<td>5°×4°L20</td>
<td>5°×4°L13</td>
<td>400</td>
<td>100</td>
<td>n/a</td>
</tr>
<tr>
<td>INM-CM3</td>
<td>INM (Russia)</td>
<td>5°×4°L21</td>
<td>2.5°×2°L33</td>
<td>330</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>IPSL-CM4</td>
<td>IPSL (France)</td>
<td>2.5°×3.75°L19</td>
<td>2°×0.5°L31</td>
<td>230</td>
<td>80</td>
<td>n/a</td>
</tr>
<tr>
<td>MIROC3.2(hires)</td>
<td>CCSR/NIES/FRCGC (Japan)</td>
<td>T106L56</td>
<td>0.28°×0.1875°L47</td>
<td>100</td>
<td>10</td>
<td>n/a</td>
</tr>
<tr>
<td>MIROC3.2(medres)</td>
<td>CCSR/NIES/FRCGC (Japan)</td>
<td>T42L20</td>
<td>1.4°×0.5°L43</td>
<td>500</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td>MRI-CGM2.3.2</td>
<td>MRI (Japan)</td>
<td>T42L30</td>
<td>2.5°×0.5°L23</td>
<td>350</td>
<td>150</td>
<td>n/a</td>
</tr>
<tr>
<td>PCM</td>
<td>NCAR (USA)</td>
<td>T42L18</td>
<td>0.66°×0.5°L32</td>
<td>350</td>
<td>96</td>
<td>n/a</td>
</tr>
<tr>
<td>UKMO-HadCM3</td>
<td>HadleyCentre (UK)</td>
<td>3.75°×2.5°L19</td>
<td>1.25°×1.25°L20</td>
<td>341</td>
<td>10</td>
<td>n/a</td>
</tr>
<tr>
<td>UKMO-HadGEM1</td>
<td>HadleyCentre (UK)</td>
<td>1.875°×1.25°L38</td>
<td>1°×0.33°L40</td>
<td>80</td>
<td>10</td>
<td>n/a</td>
</tr>
<tr>
<td>SINTEX T30</td>
<td>IPSL/INGV (France,Italy)</td>
<td>T30L19</td>
<td>2°×0.5°L31</td>
<td>200</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>SINTEX T106</td>
<td>INGV/IPS (Italy,France)</td>
<td>T106L19</td>
<td>2°×0.5°L31</td>
<td>100</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>SINTEX T106mod</td>
<td>IPSL/INGV (France,Italy)</td>
<td>T106L19</td>
<td>2°×0.5°L31</td>
<td>100</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>HadOA</td>
<td>CGAM/IPS (UK,France)</td>
<td>3.75°×2.5°L19</td>
<td>2°×0.5°L31</td>
<td>100</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

The only flux corrected model is MRI-CGM2.3.2

1: Tabulate some interesting property (and author grafts hard to get the information)

State of the Art: Model Comparison

<table>
<thead>
<tr>
<th>Model label and climate sensitivity</th>
<th>Resolution</th>
<th>Institution a</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGCM3.1(T47) 3.6 K</td>
<td>96 × 48 L32 T47</td>
<td>Canadian Centre for Climate Modelling</td>
</tr>
<tr>
<td>CGCM3.1(T63) 3.4 K</td>
<td>128 × 64 L32 T63</td>
<td>Canadian Centre for Climate Modelling</td>
</tr>
<tr>
<td>CNRM-CM3 n/a</td>
<td>128 × 64 L45 T63</td>
<td>Centre National de Recherche Météorologie</td>
</tr>
<tr>
<td>ECHAM5/MIROC-OM 3.4 K</td>
<td>192 × 96 L31 T63</td>
<td>Max-Planck-Institut für Meteorologie</td>
</tr>
<tr>
<td>ECHO-G 3.2 K</td>
<td>96 × 48 L19 T30</td>
<td>Meteorological Institute of the University Research Institute, South Korea</td>
</tr>
<tr>
<td>GFDL-CM2.0 2.9 K</td>
<td>144 × 90 L24</td>
<td>Geophysical Fluid Dynamics Laboratory et al. (2006)</td>
</tr>
<tr>
<td>GFDL-CM2.1 3.4 K</td>
<td>144 × 90 L24</td>
<td>Geophysical Fluid Dynamics Laboratory et al. (2006)</td>
</tr>
<tr>
<td>GISS-AOM n/a</td>
<td>90 × 60 L12</td>
<td>Goddard Institute for Space Studies La</td>
</tr>
<tr>
<td>GISS-ER 2.7 K</td>
<td>72 × 45 L21</td>
<td>Institut Pierre-Simon Laplace, France</td>
</tr>
<tr>
<td>INM-CM3.0 2.1 K</td>
<td>72 × 45 L21</td>
<td>Institute of Numerical Mathematics, R tD</td>
</tr>
<tr>
<td>IPSL-CM4.0 4.4 K</td>
<td>96 × 72 L19</td>
<td>Institut Pierre-Simon Laplace, France</td>
</tr>
<tr>
<td>MIROC3.2(hires) 4.3 K</td>
<td>320 × 160 L56 T106</td>
<td>Center for Climate System Research, Jti</td>
</tr>
<tr>
<td>MIROC3.2(medres) 4.0 K</td>
<td>128 × 64 L20 T42</td>
<td>Meteorological Research Institute, Japan</td>
</tr>
<tr>
<td>MRI-CGCM3.2 3.2 K</td>
<td>128 × 64 L30 T42</td>
<td>National Center for Atmospheric Research</td>
</tr>
<tr>
<td>NAR-CO2CM3 2.7 K</td>
<td>256 × 128 L26 T85</td>
<td>National Center for Atmospheric Research</td>
</tr>
<tr>
<td>NAR-PCM 2.1 K</td>
<td>128 × 64 L26 T42</td>
<td>National Center for Atmospheric Research</td>
</tr>
</tbody>
</table>

**Model name**

- **Prognostic condensate and precipitation parameterization** (Zhang et al. 2003)
- **Precipitation occurs whenever the local relative humidity is supersaturated**
- **Statistical cloud scheme of Ricard and Royer** (1993)
- **Stratiform cloud condensate scheme from Rottstam** (2000)
- **Prognostic equations for the water phases, bulk cloud microphysics** (Lohmann and Roeckerer 1996)
- **Same as PCM**
- **Cloud microphysics from Rottstam (2000) and macrophysics from Tiedtke (1993)**
- **Subgrid-relative humidity-based scheme**
- **Prognostic stratiform cloud based on moisture convergence** (Del Genio et al. 1996)
- **Large-scale precipitation is calculated based on cloud water and ice contents (similar to Smith 1990)**
- **Mixed phase cloud scheme (Wilson and Ballard 1989)**
- **Stratiform cloud fraction is calculated as linear function of relative humidity**
- **Cloud cover and in-cloud water are deduced from the large-scale total water and moisture at saturation (Bony and Emmanuel 2001)**
- **Prognostic cloud water scheme based on Le Treut and Li (1991)**
- **Precipitation occurs whenever the local relative humidity is supersaturated**
- **Precipitation occurs whenever the local relative humidity is supersaturated**

**Kharin et al, Journal of Climate 2007 doi: 10.1175/JCLI4066.1**

**Dai, A., J. Climate 2006 doi: 10.1175/JCLI3884.1**

2: Provide some (slightly) organised citation material (and author and readers graft hard to get the information)
State of the art: Model Comparison

### Table 2: Main El Niño, mean state and seasonal cycle properties of the models (pre-industrial control)

<table>
<thead>
<tr>
<th>Model</th>
<th>Code</th>
<th>ElNiño amplitude</th>
<th>SST (°C) Niño3</th>
<th>zP (Pa) Niño4</th>
<th>SCRS (%)</th>
<th>ICS (%)</th>
<th>2xCO2 (%)</th>
<th>4xCO2 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs 1948-1975</td>
<td>A</td>
<td>0.71 ± 0.04</td>
<td>25.72 ± 0.49</td>
<td>-0.032 ± 0.003</td>
<td>10.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs 1976-2004</td>
<td>B</td>
<td>0.70 ± 0.04</td>
<td>26.03 ± 0.06</td>
<td>-0.026 ± 0.000</td>
<td>8.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCSM3</td>
<td>C</td>
<td>0.78 ± 0.04</td>
<td>25.29 ± 0.08</td>
<td>-0.038 ± 0.000</td>
<td>20</td>
<td>6.1</td>
<td>-13</td>
<td></td>
</tr>
<tr>
<td>CCSM3.1(T47)</td>
<td>D</td>
<td>0.42 ± 0.03</td>
<td>24.63 ± 0.15</td>
<td>-0.045 ± 0.002</td>
<td>44</td>
<td>11.6</td>
<td>+5</td>
<td>+2</td>
</tr>
<tr>
<td>CNRM-CM3</td>
<td>E</td>
<td>1.66 ± 0.21</td>
<td>23.43 ± 0.06</td>
<td>-0.026 ± 0.000</td>
<td>3</td>
<td>6.3</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>CSIRO-Mk3.0</td>
<td>F</td>
<td>0.90 ± 0.17</td>
<td>24.34 ± 0.23</td>
<td>-0.034 ± 0.000</td>
<td>20</td>
<td>7.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECHAM5/MPI-OM</td>
<td>G</td>
<td>1.16 ± 0.09</td>
<td>25.16 ± 0.06</td>
<td>-0.034 ± 0.001</td>
<td>13</td>
<td>7.3</td>
<td>+29</td>
<td>+31</td>
</tr>
<tr>
<td>FGOALS-g1.0</td>
<td>H</td>
<td>1.93 ± 0.34</td>
<td>26.37 ± 0.16</td>
<td>-0.028 ± 0.001</td>
<td>0</td>
<td>6.6</td>
<td>-27</td>
<td></td>
</tr>
<tr>
<td>GFDL-CM2.0</td>
<td>I</td>
<td>0.75 ± 0.19</td>
<td>24.74 ± 0.15</td>
<td>-0.043 ± 0.000</td>
<td>37</td>
<td>8.8</td>
<td>+20</td>
<td>+25</td>
</tr>
<tr>
<td>GFDL-CM2.1</td>
<td>J</td>
<td>1.32 ± 0.08</td>
<td>24.98 ± 0.14</td>
<td>-0.044 ± 0.000</td>
<td>12</td>
<td>12.8</td>
<td>+2</td>
<td>-18</td>
</tr>
<tr>
<td>GISS-AOM</td>
<td>K</td>
<td>0.17 ± 0.03</td>
<td>27.07 ± 0.01</td>
<td>-0.026 ± 0.000</td>
<td>45</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GISS-EH</td>
<td>L</td>
<td>0.86 ± 0.13</td>
<td>24.53 ± 0.13</td>
<td>-0.037 ± 0.001</td>
<td>24</td>
<td>0.8</td>
<td>-5</td>
<td>-21</td>
</tr>
<tr>
<td>INM-CM3</td>
<td>M</td>
<td>0.24 ± 0.01</td>
<td>28.16 ± 0.03</td>
<td>-0.026 ± 0.001</td>
<td>22</td>
<td>2.2</td>
<td>-21</td>
<td>+8</td>
</tr>
<tr>
<td>IPSL-CM4</td>
<td>N</td>
<td>0.92 ± 0.10</td>
<td>24.15 ± 0.09</td>
<td>-0.025 ± 0.001</td>
<td>23</td>
<td>6.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIROC3.2(hires)</td>
<td>O</td>
<td>1.00 ± 0.02</td>
<td>26.28 ± 0.08</td>
<td>-0.026 ± 0.000</td>
<td>13</td>
<td>5.9</td>
<td>-16</td>
<td></td>
</tr>
<tr>
<td>MIROC3.2(merdes)</td>
<td>P</td>
<td>0.35 ± 0.01</td>
<td>25.46 ± 0.14</td>
<td>-0.042 ± 0.002</td>
<td>86</td>
<td>15.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRI-CGM2.3.2</td>
<td>Q</td>
<td>0.44 ± 0.11</td>
<td>24.81 ± 0.03</td>
<td>-0.040 ± 0.000</td>
<td>60</td>
<td>10.7</td>
<td>+5</td>
<td>+2</td>
</tr>
<tr>
<td>PCM</td>
<td>R</td>
<td>0.70 ± 0.05</td>
<td>25.04 ± 0.04</td>
<td>-0.045 ± 0.000</td>
<td>35</td>
<td>16</td>
<td>+34</td>
<td>+77</td>
</tr>
<tr>
<td>UKMO-HadCM3</td>
<td>S</td>
<td>0.79 ± 0.07</td>
<td>25.58 ± 0.07</td>
<td>-0.045 ± 0.001</td>
<td>13</td>
<td>10.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKMO-HadGEM1</td>
<td>T</td>
<td>0.68 ± 0.17</td>
<td>23.69 ± 0.12</td>
<td>-0.064 ± 0.001</td>
<td>28</td>
<td>8.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTTEX30</td>
<td>U</td>
<td>0.61 ± 0.09</td>
<td>25.90 ± 0.08</td>
<td>-0.041 ± 0.001</td>
<td>13</td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTTEX106</td>
<td>V</td>
<td>0.74 ± 0.07</td>
<td>26.27 ± 0.16</td>
<td>-0.025 ± 0.002</td>
<td>5</td>
<td>7.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTTEX106mod</td>
<td>W</td>
<td>0.67 ± 0.06</td>
<td>26.84 ± 0.25</td>
<td>-0.041 ± 0.001</td>
<td>8</td>
<td>6.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadOPA</td>
<td>X</td>
<td>1.67 ± 0.14</td>
<td>27.46 ± 0.36</td>
<td>-0.035 ± 0.001</td>
<td>5</td>
<td>7.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SCRS is the seasonal cycle relative strength (in %). ICS is the summer interannual coupling strength (in 10^3 Pa°C). The El Niño amplitude change to doubling and quadrupling of CO2 (when compared to preindustrial) are shown in the last two columns. The El Niño amplitude is defined as the standard deviation of SST in the Niño3 region. Errors were estimated with a moving block bootstrap to account for serial correlation (windows: El Niño period of Fig. 1 for standard deviation and 10 month for means). The amplitude change values underlined

3: Calculate and tabulate some interesting properties and bury in a table or figure

Why does this information detail matter?

... surely a technical paper can make lots of technical references, and those in the know, are, ... in the know?

By and large: the climate projections community is actually a group of communities: From next generation “experimenters”, to “big” GCM modellers, to regional modellers, impacts assessment modelling, to impacts and adaptation modelling. Information does not easily flow between communities!
We've already seen concepts:

• Experiments (e.g. specific scenario projections)
• Models, with specific (experiment dependent)
  – Resolution & Grids
  – combinations of sub-component models for specific processes.

and

• Simulations
  – Models run for specific experiments (and hence specific “boundary conditions”, e.g. CO2 projections)
  – with specific output variables, frequencies, and durations
  – run on specific platforms

Metafor:

• 2.2 M euro EC project to deliver
  – a “Common Information Model” to document these concepts, particularly in the context of supporting the next IPCC assessment report.
  – Associated infrastructure to collect and view such documentation, and
  – Build an accompanying governance structure ...
(Yes we know we shouldn't have this sort of detail in the UML, and it won't be ... shortly)
Scientific Properties:
Controlled Vocabularies developed with expert
community using mindmaps and some “rules” to aid
machine post processing ...

(everyone can use mindmaps: no \textit{a priori}
semantic technology knowledge required.)
A piece of the mindmap XML …
... vocabulary driven content in web based “human entry tool”

(A great advertisement for Python and Django)
CIM Tools

Consumption

(Spot the common factor: three groups, all python)
Provenance research and Metafor

There are a number of other major projects/paradigms addressing provenance in one way or another, including, but not limited to:

- The Open Provenance Model
- The Proof Markup Language,
- ISO19156 Observations and Measurements.

Metafor is a much more specialised activity than any of those, but the metafor concepts can be abstracted into their higher level concepts.

- In 2011, Metafor will be refactored to be O&M compliant, and we will develop an automated RDF serialisation (the current serialisation to RDF/OWL is not expected to remain stable).
- The OWL version of the Metafor CIM will subsequently be related to upper level provenance ontologies.
CMIP5: Fifth Coupled Model Intercomparison Project

- Global community activity under the auspices of the World Meteorological Organisation (WMO) via the World Climate Research Programme (WCRP)

- **Aim:**
  - to address outstanding scientific questions that arose as part of the AR4 process,
  - improve understanding of climate, and
  - to provide estimates of future climate change that will be useful to those considering its possible consequences.

- **Method:** standard set of model simulations in order to:
  - evaluate how realistic the models are in simulating the recent past,
  - provide projections of future climate change on two time scales, near term (out to about 2035) and long term (out to 2100 and beyond), and
  - understand some of the factors responsible for differences in model projections, including quantifying some key feedbacks such as those involving clouds and the carbon cycle.
Introduction to CMIP5: The Experiments

An important focus is model evaluation and understanding...

Example: CMIP5 long-term suite of experiments

Take home points here:

Many distinct experiments, with very different characteristics, which influence the configuration of the models, (what they can do, and how they should be interpreted).

(from Karl Taylor)
Introduction to CMIP5: The Experiments

Detection-Attribution (IDAG)

Integrated Assessment Consortium (IAM), connection to WG-III

Paleo (PMIP, IGBP-PAGES)

Take home point: - many different communities and projects

Cloud and moist processes (CFMIP-GCSS WGNE)

Carbon-climate feedbacks (C4MIP, IGBP-AIMES)

Chemistry, aerosols (SPARC, AC&C, CCMVal, aerocon)

(from Karl Taylor)
Simulations:
~90,000 years
~60 experiments
~20 modelling centres (from around the world) using
~30 major(*) model configurations
~2 million output “atomic” datasets
~10's of petabytes of output
~2 petabytes of CMIP5 requested output
~1 petabyte of CMIP5 “replicated” output
Which will be replicated at a number of sites (including ours), to start arriving in the next few months.

Of the replicants:
~ 220 TB decadal
~ 540 TB long term
~ 220 TB atmos-only
~80 TB of 3hourly data
~215 TB of ocean 3d monthly data!
~250 TB for the cloud feedbacks!
~10 TB of land-biochemistry (from the long term experiments alone).
# SI Prefixes

<table>
<thead>
<tr>
<th>SI prefix</th>
<th>Name</th>
<th>Power of 10 or 2</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>kilo</td>
<td>$10^3$</td>
<td>$2^{10}$</td>
</tr>
<tr>
<td>M</td>
<td>mega</td>
<td>$10^6$</td>
<td>$2^{20}$</td>
</tr>
<tr>
<td>G</td>
<td>giga</td>
<td>$10^9$</td>
<td>$2^{30}$</td>
</tr>
<tr>
<td>T</td>
<td>tera</td>
<td>$10^{12}$</td>
<td>$2^{40}$</td>
</tr>
<tr>
<td>P</td>
<td>peta</td>
<td>$10^{15}$</td>
<td>$2^{50}$</td>
</tr>
<tr>
<td>E</td>
<td>exa</td>
<td>$10^{18}$</td>
<td>$2^{60}$</td>
</tr>
<tr>
<td>Z</td>
<td>zetta</td>
<td>$10^{21}$</td>
<td>$2^{70}$</td>
</tr>
<tr>
<td>Y</td>
<td>yotta</td>
<td>$10^{24}$</td>
<td>$2^{80}$</td>
</tr>
</tbody>
</table>

*Stuart Feldman, Google*
Earth System Grid (ESG)

US Department of Energy funded project to support the delivery of CMIP5 data to the community.

Consists of

- distributed data node software (to publish data)
- Tools (Live Access Server, LAS, Bulk Data Mover, BDM, security systems etc)
- gateway software (to provide catalog and services)

Major “technical challenge”

Earth System Grid FEDERATION (ESGF)

Global initiative to deploy the ESG (and other) software to support:

- timely access to the data
- minimum international movement of the data
- long term access to significant versions of the CMIP5 data.

Major “social challenge” as well as “technical challenge”
Earth System Grid Data Nodes

Access Control

Thredds Data Server
GridFTP
Toolbox:
Live Access Server
(other) OpenDAP
servers

ESG Publisher inc CDAT
Bulk Data Mover (BDM)

Databases: Thredds Catalog, Postgres & Myproxy

Filesystem: preferably with a DRS layout

Deployed by “data providers” to “expose” their data via “Earth System Grid Gateways”
ESGF – Starts with Data Nodes

20 to 30, globally distributed, each with o(50-1000)TB
Originating Modelling Groups Deploy ESG Data Nodes (their own, or proxied, e.g. MOHC)

Metadata Harvest

ESGF Data Nodes with Gateways

Data nodes publish to Gateways

Gateways replicate metadata

(all data visible on all gateways)
ESGF: Nodes (Original and Replicates) & Gateways

Metadata Harvest

Data Replication

IPCC archive will persist all data with DOIs

“Core” requested data replicated to archives

Data replicated to Core Archives
CMIP5: Handling the metadata

Three streams of provenance metadata:
A) “archive” metadata
B) “browse” metadata
C) “character” metadata

A: **Archive** Metadata: three levels of information from the file system:
I. CF compliance in the NetCDF files
II. “Extra” CMIP5 required attributes including a unique identifier within each file.
III. Use of the Directory Reference Syntax (DRS) to help maintain version information.

Compliance enforced by ESG publisher.

B: **Browse** Metadata, added independently of the archive
• Exploiting Metafor controlled vocabularies via a customised “CMIP5 questionnaire”.

compliance enforced by CMIP5 quality control systems, leading to

C: **Character** Metadata
• Data assessment

Four concepts to follow up on:
1) A, B, C: metadata taxonomy
2) Metafor questionnaire
3) CMIP5 quality control
4) Combining the streams (the information pipeline to the Earth System Grid Gateways)
1) Metadata Taxonomy: Discovery, Documentation, Definition

In CMIP5 we haven't really addressed formal D (ISO19115 class) metadata yet.
2) Metafor Questionnaire
Metafor questionnaire: many parts ...
Metafor Quality Control

Specialises ISO19115 DQ package

CIM_Quality_Record

CIM_Issues

CIM_Reports

CIM_Measures

describes quality for

Remote Resource
(identified by URL and internal URI)

0..n

0..n

against

Log Files etc

Maps nicely onto the upcoming Open Annotation Model
(http://annotation.lanl.gov)

Extrinsic! Not intrinsic!
CMIP5 qctool (courtesy of Metafor)

This quality control tool allows one to associate two kinds of information with a remote resource:

- **issues**: that is things that are wrong with the remote resource, or which need looking into, and
- **reports**: comments on the consistency of the remote resource with various measures.

It is likely that in using it, you would start by creating one or more resource records, after which you can add issues and reports.

- Issues can include subissues, and can have zero or more partial resolutions and zero or one final resolution.
- Reports can include three types of results:
  - Conformance results (against the measure, with specification of how the conformance was evaluated),
  - Quantitative results, and
  - Uploaded or referenceable result plots, logfiles etc.

  All reports must include either a conformance or quantitative result, and may include zero to many uploaded results.

Future versions of this tool will have better help text here, along with diagrams showing the expected usage.
Quality Control Types

Producer Quality Control

• Modellers will be doing scientific and data completeness quality control before they even attempt to publish the data.

• ESGF will do a significant amount of automated quality control, coupled with scientific “spot checks”.

• The ESGF quality control will be according to a set of defined “qc levels”

Consumer Quality Control

• Consumers will be doing additional “spot checking” whether they know it or not. They will be able to raise “issues” against data.

• They will also be able to define their own scientific measures, and enter information against specific models, and simulations. These data will be referencable and searchable
  – (avoiding the “buried in the table” problem demonstrated earlier)
CMIP5 Quality Control

<table>
<thead>
<tr>
<th>Label</th>
<th>Data</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>qc-1d</td>
<td>ESG publisher enforces some data checking</td>
<td>CF compliance</td>
</tr>
<tr>
<td>qc-1m</td>
<td>Questionnaire enforces some constraints and vocabularies, requires XSD validation.</td>
<td></td>
</tr>
<tr>
<td>qc-2m</td>
<td>Subjective examination by metafor team.</td>
<td></td>
</tr>
<tr>
<td>qc-2d</td>
<td>Automated examination with subjective spot checks: carried out at PCMDI, DKRZ and BADC.</td>
<td>Provisional DOI granted.</td>
</tr>
<tr>
<td>qc-3</td>
<td>Further subjective tests at DKRZ, author approval of all metadata and output. Final DOI granted.</td>
<td></td>
</tr>
</tbody>
</table>

Scientific Metrics
CMIP5 requires no scientific validation, but qc system will support data annotation against specific metrics of scientific interest.
CMIP5 Quality Control as a gatekeeper to global data flow and access:
- fail qc-1d: data not published
- fail qc-1m: no data access
CMIP5 Quality Control as a gatekeeper to global data flow and access:

- **fail qcl1-d:**
  - data not published
- **fail qcl1-m:**
  - no data access
- **pass qcl1-d**
  - Get data to a core data centre
- **Pass qcl2-d**
  - Start replication
- **Pass qcl2-m**
  - Provisional DOI
  - Start qcl3 process eventually gain permanent DOI

Here is a flowchart representation of the process:

1. **Modeling groups provide data & metadata input**
   - **Publish data to a gateway from the data's originating node (QCL1-D)**
   - **Is data in “requested” category?**
     - No action if **no**
     - **Is data at PCMDI, BADC, DKRZ?**
       - Yes: “Send” data to PCMDI
       - No: Does data pass QCL2-d test?
9. **Flag data as QCL1 compliant and give access to all users**
10. **Does data pass QCL2-M test?**
11. **Submit to DOI assignment process**
12. **Does data pass QCL2-d test?**
13. **Replicate data**
14. **Are data and metadata in place and self-consistent?**
15. **QCL1-M**
16. **Modellers complete METAFO questionnaires**
17. **Submit to DOI**

All creation tools provide atom feeds of their internal contents. XML documents are initially persisted locally and may be duplicated remotely (and persisted in different formats).

Generic editing with Geonetwork
<?xml version="1.0" encoding="UTF-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
  <id>http://ceda.ac.uk/feeds/cmip5/experiment/</id>
  <title>CMIP5 model experiment metadata</title>
  <subtitle>Metafor questionnaire - completed experiment documents</subtitle>
  <updated>2010-03-04T00:00:00Z</updated>
  <link href="http://ceda.ac.uk/feeds/cmip5/experiment/" rel="self"></link>
  <author><name>The metafor team</name></author>
  <generator version="r33" uri="http://code.google.com/p/django-atompub/">
django-atompub</generator>

  <entry>
    <id>urn:uuid:1fb380d2-2759-11df-924b-00163e9152a5</id>
    <title>5.5-1 esmFdbk1( 5.5-1 ESM feedback 1)</title>
    <updated>2010-03-04T00:00:00Z</updated>
    <published>2010-03-04T00:00:00Z</published>
    <summary>Impose conditions identical to 3.1::Control but radiation code sees CO2 concentration increase.</summary>
    <content src="/cmip5/experiment/1fb380d2-2759-11df-924b-00163e9152a5/1/" type="application/xml"></content>
  </entry>

  <entry>
    <id>urn:uuid:1fd2019c-2759-11df-924b-00163e9152a5</id>
    ...<br>
</feed>
Bringing it all together for CMIP5

The players:

1) NetCDF
   CF Conventions + CMIP5 extensions (orange)

2) Earth System Grid + Earth System Curator (green)

3) Metafor (yellow)

Conversion code (light blue/grey)

Glue: Atom
Earth System Grid Gateways

Earth system grid data nodes “publish” to a gateway (essentially the gateway harvests the information in their TDS catalog) and provide a search interface both to the harvested data, and to metadata harvested from the metafor questionnaire.
ESG Gateway hosted by the British Atmospheric Data Centre

Welcome to British Atmospheric Data Centre’s ESG gateway

This gateway is provided by the BADC on behalf of the European climate science community and the IPCC Data Distribution Centre.

BADC is part of the STFC Centre for Environmental Data Archival (CEDA) and exists to curate, and facilitate access to, data of importance to the environmental science community. BADC is primarily supported by the Natural Environment Research Council via the National Centre for Atmospheric Science (NCAS). The UK component of the IPCC Data Distribution Centre is supported by the UK Department of Energy and Climate Change.

CMIP5: The 5th Coupled Model Intercomparison Project

The main reason for this gateway is to provide access to the globally distributed data produced for CMIP5 along with the accompanying metadata.

The comprehensive scientific and technical metadata available for CMIP5 will far exceed the information available for previous CMIP efforts. The metadata descriptions are linked to CMIP5 model outputs, and vice versa. To view the metadata describing the simulations undertaken for CMIP5, select Simulations in the Search pull down menu on the left of the box above.
Earth System Grid

Advanced Search

Search Categories
- Project
- Experiment
  - Any Experiment
    - sresa1b
  - Realm
  - Variable
  - Any Variable
    - air temperature

Total Number of Results: 15
1-10 of 15 results | 11-15

1. cmip3_drs.output.BCCR BCM2.sresa1b.day.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

2. cmip3_drs.output.BCCR BCM2.sresa1b.mon.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

3. cmip3_drs.output.CCCMA CGCM3-1-T47.sresa1b.day.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

4. cmip3_drs.output.CCCMA CGCM3-1-T47.sresa1b.mon.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

5. cmip3_drs.output.GFDL CM2-1.sresa1b.3hr.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

6. cmip3_drs.output.GFDL CM2-1.sresa1b.day.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

7. cmip3_drs.output.GFDL CM2-1.sresa1b.mon.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

8. cmip3_drs.output.GFDL CM2-1.sresa1b.3hr.atmos
   Authorization: Guest Users
   Data Center: ESG-BADC

Files Download

Download all files for the selected datasets. Optionally use a wildcard expression to filter the filenames (example: use *.nc to select all files with extension nc).
### Download Data

#### Sub Select File Results

**File Name:**

```
*.*
```

Use `*` for a wildcard character. Regular Expressions will not work at this time.

**Variables:**

- Latent Heat Flux (hfls)
- Sensible Heat Flux (hfls)
- Specific Humidity (hus)
- Accumulated Total Precipitation (pr)
- Surface Pressure (ps)
- Mean Sea Level Pressure (psi)
- Surface Downward Longwave Radiation (tsds)
- Surface Upward Longwave Radiation (rues)
- Top-of-Atmosphere upward Longwave Radiation (nut)
- Surface Downward Shortwave Radiation (tsds)
- Surface Upward Shortwave Radiation (rues)
- Air Temperature (tas)
- Temperature 2m (tas)
- 2m max temperature (tasmax)
- 2m min temperature (tasmin)
- Zonal Wind Component (u)

#### File Download Selection

<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
<th>Format</th>
<th>Location</th>
<th>Direct Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2050-2065.nc</td>
<td>114.25 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2055-2065.nc</td>
<td>114.25 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2055-2065.nc</td>
<td>114.25 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2081-2090.nc</td>
<td>114.19 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2081-2090.nc</td>
<td>114.19 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2081-2090.nc</td>
<td>114.19 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2091-2100.nc</td>
<td>102.49 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
<tr>
<td>hfls_A2_BCM2_sresalb_r1_2091-2100.nc</td>
<td>102.49 MB</td>
<td>NetCDF</td>
<td>DISK</td>
<td>download</td>
</tr>
</tbody>
</table>

| hfls_A2_BCM2_sresalb_r1_2091-2100.nc          | 102.49 MB | NetCDF   | DISK     | download        |
Access Control and Delivery:
(1) Via Gateway

Wget scripts access secure data using myproxy and X509 certificates
Access Control and Delivery:
(2) Direct from TDS on Data Node

Catalog https://cmip-dn1.badc.rl.ac.uk/thredds/esgcat/2/cmip3_drs.output.UKMO.HADCM3.1pctto2x.mon.seaIce.v1.htm

Dataset:
cmip3_drs.output.UKMO.HADCM3.1pctto2x.mon.seaIce/sic_O1_HADCM3_1pctto2x_r1_1859_D1

- Data format: NetCDF
- Data size: 159.2 Mbytes
- Data type: GRID
- ID: cmip3_drs.output.UKMO.HADCM3.1pctto2x.mon.seaIce.v1.sic_O1_HADCM3_1pctto2x_r1_1859_Dec_to_1939_Dec.nc
- Restrict Access: esg-user

Access:
1. HTTPServer:
   /thredds/fileServer/cmip3_drs/output/UKMO/HADCM3/1pctto2x/mon/seaIce/sic/r1/v1/sic_O1_HADCM3_1pctto2x_r1_1859_Dec_to_1939_Dec.nc

Variables:
- Vocabulary [CF-1.0]:
  - sic = Sea Ice Concentration = sea_ice_area_fraction (%)
  - sit = Sea Ice Thickness = sea_ice_thickness (m)

Properties:
- file_id = "cmip3_drs.output.UKMO.HADCM3.1pctto2x.mon.seaIce.sic_O1_HADCM3_1pctto2x_r1_1859_Dec_to_1939_Dec.nc"
- file_version = "1"
- size = "159288492"
- mod_time = "2009-12-04 11:52:23"
- checksum = "e6a90b1eb5291c38c9c40d52f1828ce1f"
- checksum_type = "MD5"
Access Control and Delivery:
(2) Direct from TDS on Data Node
Access Control and Delivery:
(2) Direct from TDS on Data Node
ESG (and CEDA) have comprehensive access control middleware suitable for use in browsers and command line – federated globally!
(Back to the data): User Perspective

Data nodes will also deploy other tools: secure opendap coming soon (it's done, with modifications to the netcdf client libraries too) … it just needs to be configured to be visible.
Moving forward

• The Earth System Grid is a U.S. Project.
  - There will undoubtedly be successor projects
  - (Key role of ESG Curator and the NOAA Global Interoperability Project)
• The Earth System Grid Federation is a global activity,
  - led by the Global Organisation for Earth System Science Portals (GO-ESSP)
• In Europe, we are underpinning ESGF via two EC funded projects:
  - Metafor (which we have seen a lot of), and
  - IS-ENES (InfraStructure for a European Network for Earth Simulation)
    - (and much national work too of course)
• Metafor and IS-ENES are working on complementary information architectures
  - Metafor will finish in 2011, IS-ENS has some years to run.
  - (Metafor will leave an international governance system in place for the Common Information Model)
The Climate problem is one that integrates much of e-research, and in particular, the necessity for

- Major physical e-infrastructure (networks, supercomputers)
- Comprehensive information architectures covering the gamut of the information life cycle, including annotation (particularly of quality)
  
  ... and hard work populating these information objects, particularly with provenance detail.
- Sophisticated tooling to produce and consume the data and information objects
- State of the art access control techniques

Major distributed systems are social challenges as much as technical challenges.

The Fifth Coupled Model Intercomparison project (CMIP5) provides an exemplar of most of these things, supported as it is, by a major global federation of activities.