Science & Technology Facilities Council

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.

Science & Technology Facilities Council

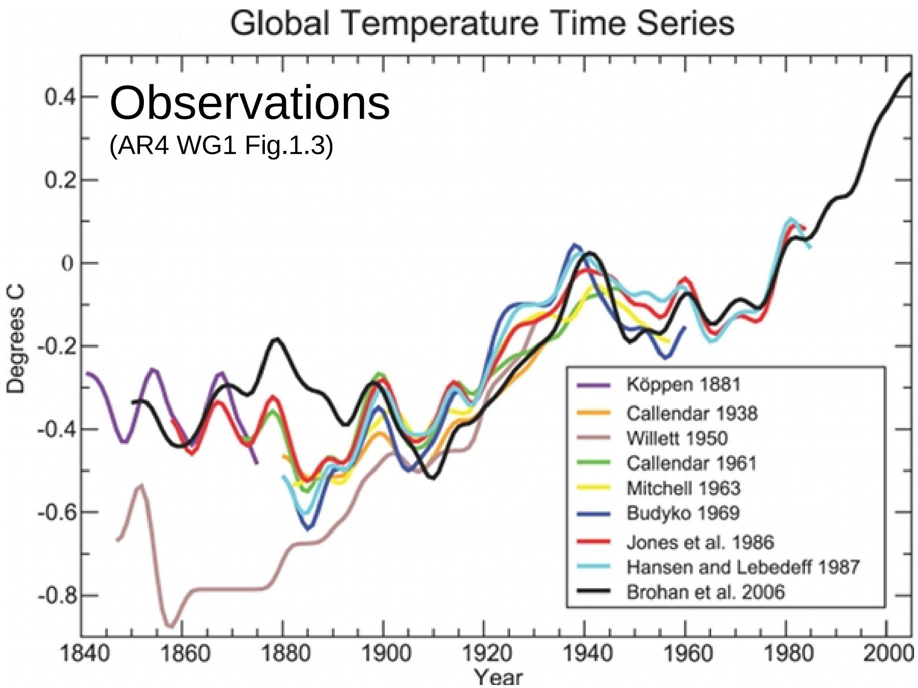
Outline

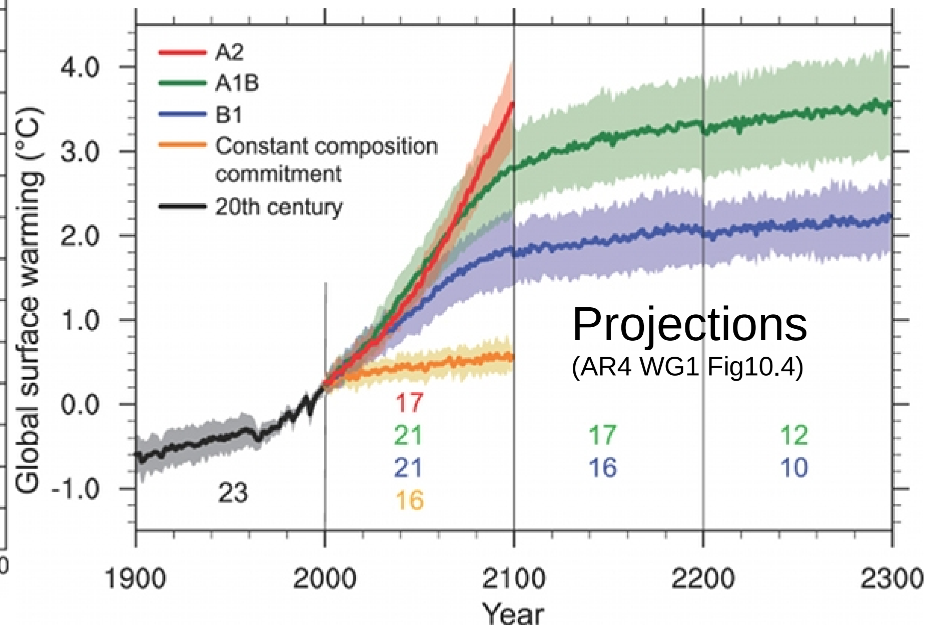
- The Climate Problem
  - Data Generating Infrastructure and the need for metadata
  - Evaluating Australia in CMIP3 Climate Models
  - Climate Model primer
  - The problem with understanding the differences between models and the simulations they produce.
- A Brief introduction to Metafor
- An introduction to the CMIP5 Information Ecosystem
  - Aims and objectives of CMIP5
  - Global problem: Global simulations simulated globally.
  - Global Deployment of information system
    - 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

Climate 101

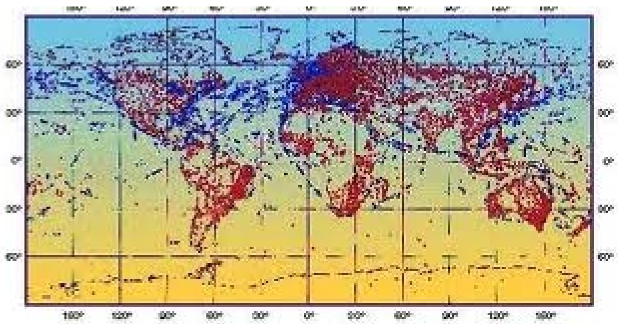


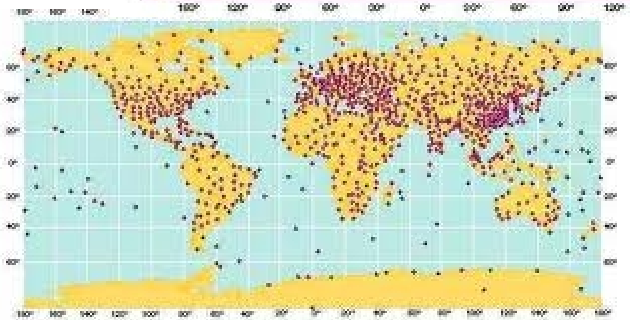


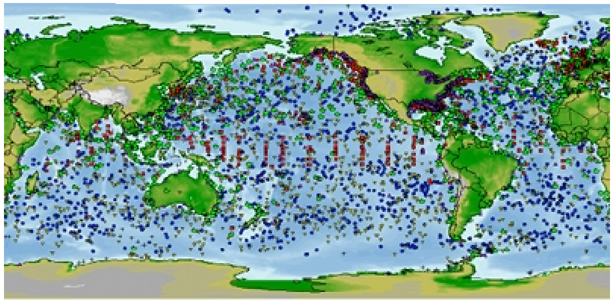
… all seems rather simple doesn't it?
Nice consumable curves ...
Enough for mitigation policy perhaps, but enough for adaptation policy?

In the beginning: observations

WMO







Images: from J. Lafeuille, 2006

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

Science & Technology Facilities Council

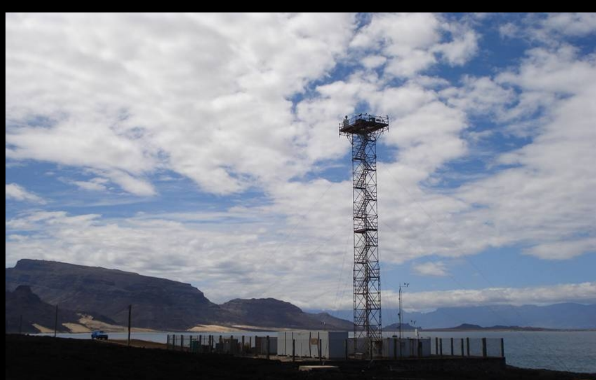
NERC Observatories and Sensor Networks











British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

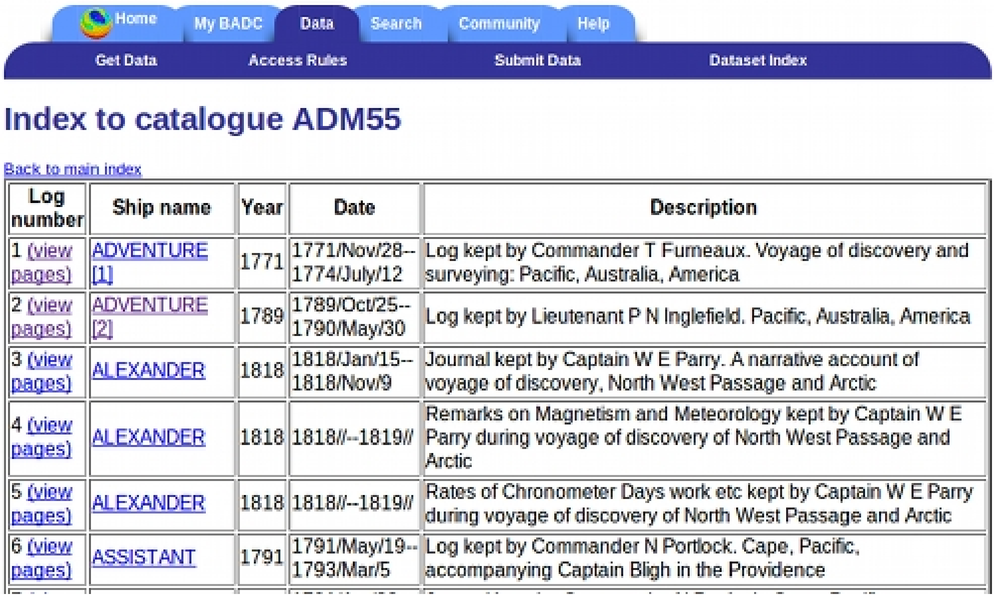
Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC Mobile Research Sensors

Slide courtesy of Alan Gadian, NCAS

Old Weather/New Results

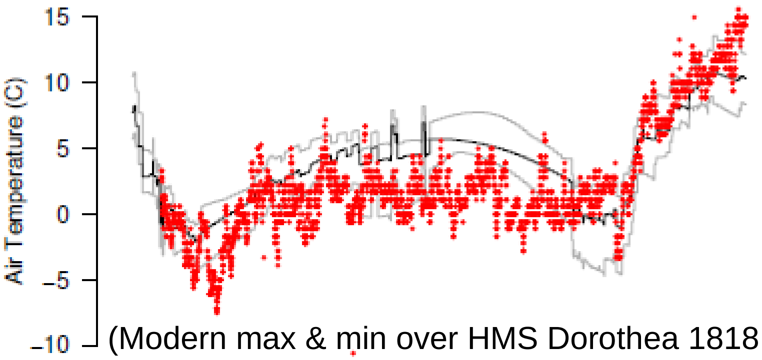


Index to catalogue ADM55

Back to main index

Log number	Ship name	Year	Date	Description
1 (view pages)	ADVENTURE [1]	1771	1771/Nov/28--1774/July/12	Log kept by Commander T Furneaux. Voyage of discovery and surveying: Pacific, Australia, America
2 (view pages)	ADVENTURE [2]	1789	1789/Oct/25--1790/May/30	Log kept by Lieutenant P N Inglefield. Pacific, Australia, America
3 (view pages)	ALEXANDER	1818	1818/Jan/15--1818/Nov/9	Journal kept by Captain W E Parry. A narrative account of voyage of discovery, North West Passage and Arctic
4 (view pages)	ALEXANDER	1818	1818//--1819//	Remarks on Magnetism and Meteorology kept by Captain W E Parry during voyage of discovery of North West Passage and Arctic
5 (view pages)	ALEXANDER	1818	1818//--1819//	Rates of Chronometer Days work etc kept by Captain W E Parry during voyage of discovery of North West Passage and Arctic
6 (view pages)	ASSISTANT	1791	1791/May/19--1793/Mar/5	Log kept by Commander N Portlock. Cape, Pacific, accompanying Captain Bligh in the Providence

Atmospheric Circulation
ACRE
Reconstructions over the Earth

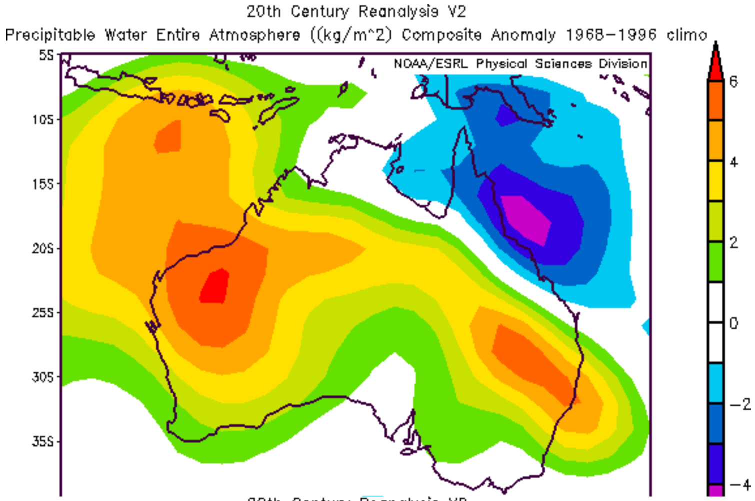


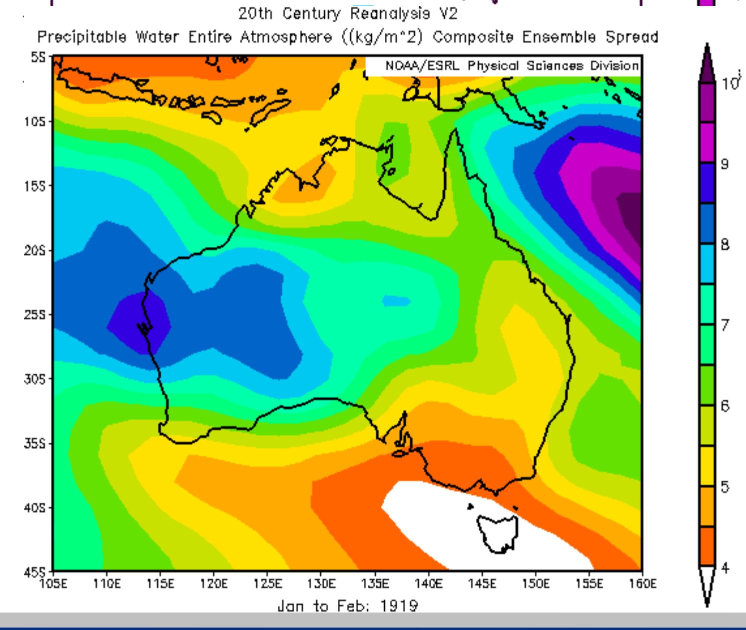
(Modern max & min over HMS Dorothea 1818
Brohan et al 2010)

20^{th} Century Reanalysis

- 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





… and so to metadata/provenance

- 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

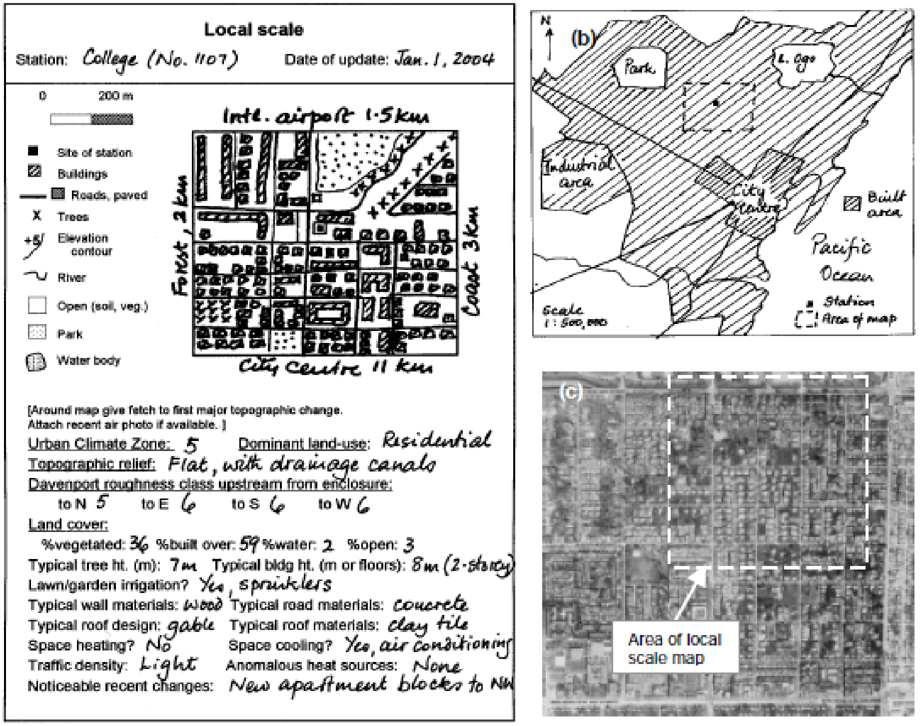


Figure 6 — Minimum information necessary to describe the local scale environment of an urban station, consisting of (a) template to document local setting, (b) sketch map to situate the station in the larger urban region, and (c) an aerial photograph.

WMO/TD 1250 (2006) (pdf)

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

… 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.

Humans and Big Data

A person working full time for a year has about 1500 hours to do something. Moore's Law wont 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)

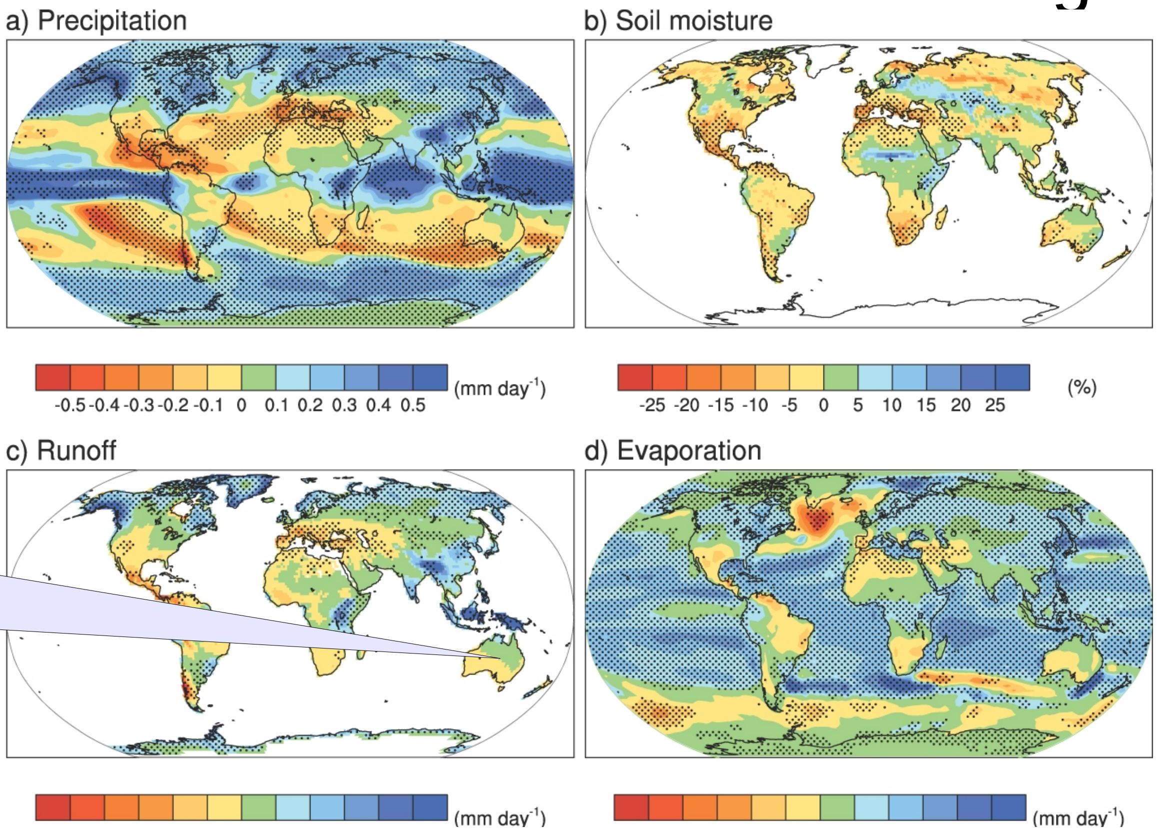


Figure 10.12. *Multi-model mean changes in (a) precipitation (mm day^{-1}), (b) soil moisture content (%), (c) runoff (mm day^{-1}) and (d) evaporation (mm day^{-1}). To indicate consistency in the sign of change, regions are stippled where at least 80% of models agree on the sign of the mean change. 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.*

Spatial and temporal subsetting … statistics over models ...

So why was Australia not stippled?

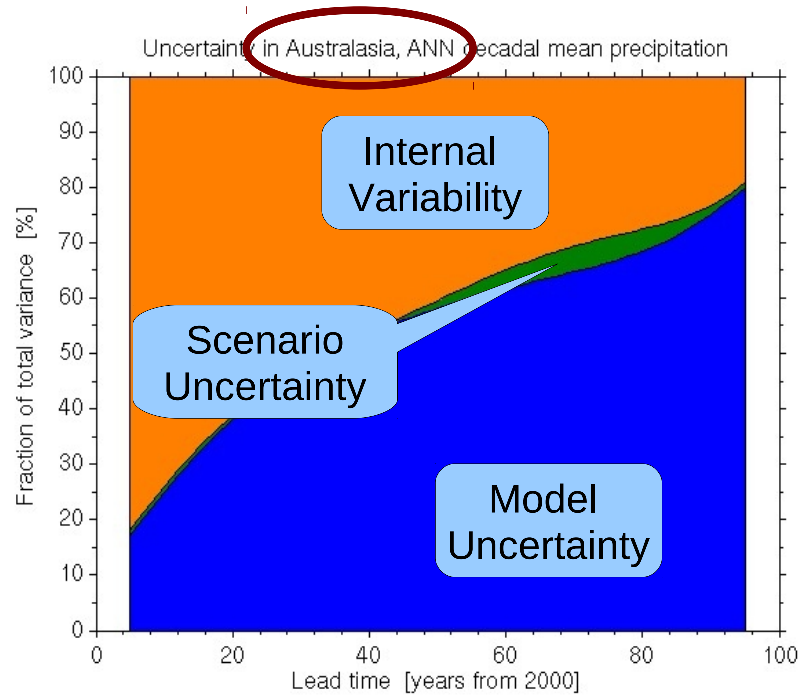
Interannual variability means that when our projections need to start in the right state (and capture that variability correctly too).

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.

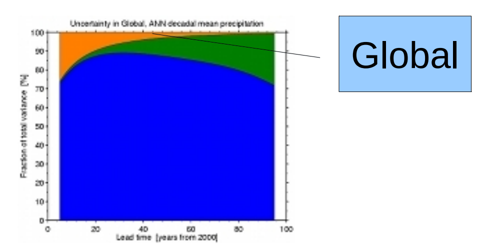
Scenario uncertainty means that we are not sure of the impact of different economic and emission futures.

(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!)*



Hawkins and Sutton, Climate Dynamics, 2010 (10.1007/s00382-010-0810-6)



...and deeper: CMIP3: What models did what?

AR4: WG1 Table 10.4

- time-independent land surface
- >1 1 monthly-mean atmosphere
- daily-mean atmosphere
- 3-hourly atmosphere
- time-independent ocean
- >1 1 monthly-mean ocean
- >1 1 Extreme Indices
- Forcing
- ISCCP Simulator

Model ID	Model, Country	Pre industr. control	Present day control	20th century	Commit-ment	SRES A2	SRES A1B	SRES B1	1% to $2xCO_2$	1% to $4xCO_2$	Slab ocean control	$2XCO_2$	AMIP
1	BCC-CM1, China		2	4		2		2	1	1			4
2	BCCR-BCM2.0, Norway	1		1	1	1	1	1	1				
3	CCSM3, USA *	2	1	9	5	5	7	8	1	1	1	1	1
4	CGCM3.1(T47), Canada	1		5	5	5	5	4	1	1	1	1	
5	CGCM3.1(T63), Canada	1		1			1	1	1		1	1	
6	CNRM-CM3, France	1		1	1	1	1	1	1	1			1
7	CSIRO-MK3.0, Australia	2		3	1	1	1	1	1		1	1	
8	ECHAM5/MPI-OM, Germany	1		4	3	3	4	3	3	1	1	1	3
9	ECHO-G, Germany/Korea	1	1	5	4	3	3	3	1	1			
10	FGOALS-g1.0, China	3		3	3		3	3	3				3
11	GFDL-CM2.0, USA	1		3	1	1	1	1	1	1			
12	GFDL-CM2.1, USA	1		3	1	1	1	1	1	1			
13	GISS-AOM, USA	2		2			2	2					
14	GISS-EH, USA	1		5			4		1				
15	GISS-ER, USA	1		9	1	1	5	1	1	1	1	1	4
16	INM-CM3.0, Russia	1		1	1	1	1	1	1	1	1	1	1
17	IPSL-CM4, France	1	1	2	1	1	1	1	1	1			6
18	MIROC3.2(hires), Japan	1		1			1	1	1		1	1	1
19	MIROC3.2(medres), Japan	1		3	1	3	3	3	3	3	1	1	3
20	MRI-CGCM2.3.2, Japan	1	1	5	1	5	5	5	1	1	1	1	1
21	PCM, USA	1	1	4	3	4	4	4	5	1			1
22	UKMO-HadCM3, UK	2		2	1	1	1	1	1				
23	UKMO-HadGEM1, UK	1		1		1	1		1		1	1	1

Rows: Models, and their output types. Columns: Experiments and Projections

(Three layers of complexity: models, experiments, output … each of which is itself complex)

Digression: What is a model?

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

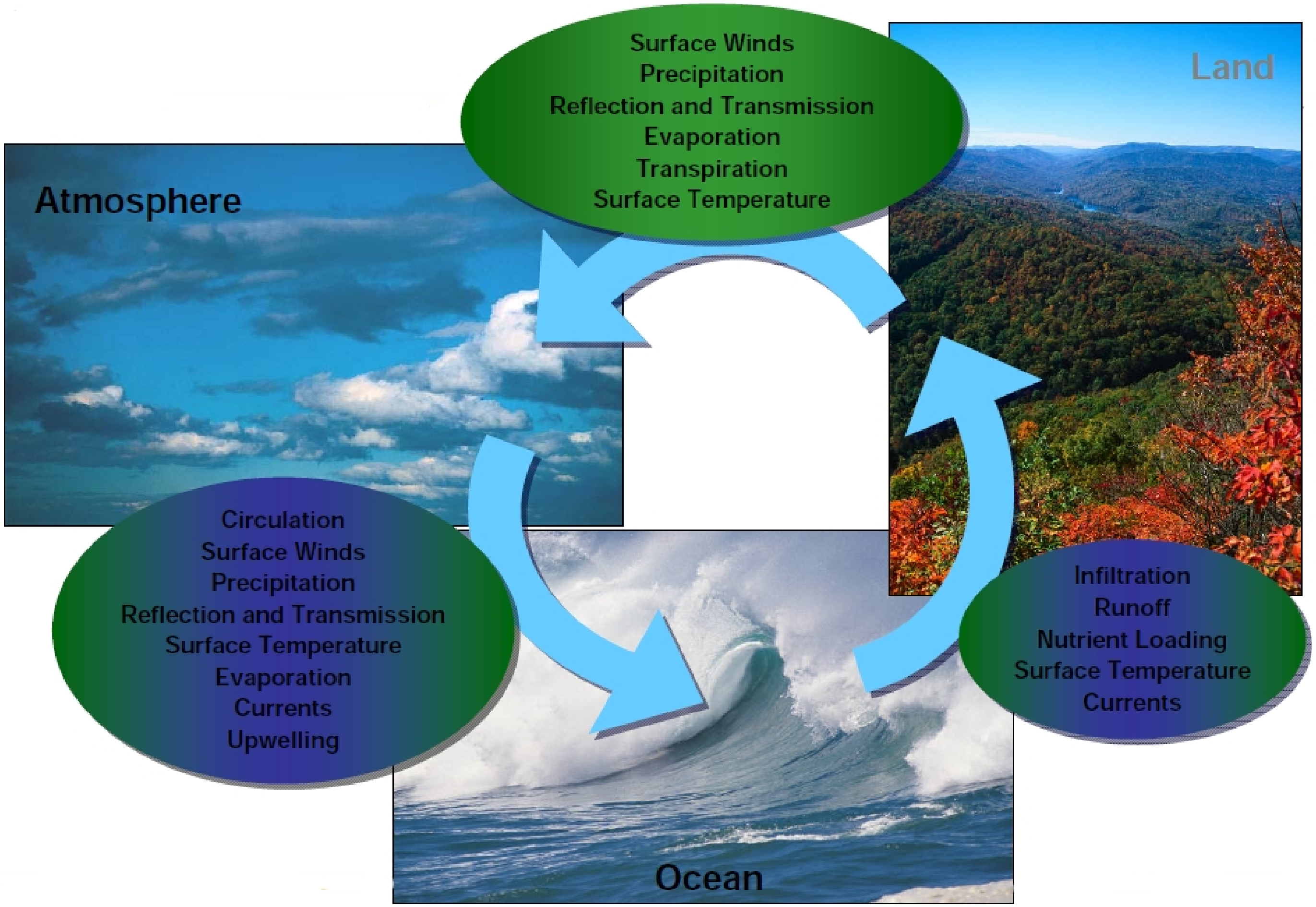
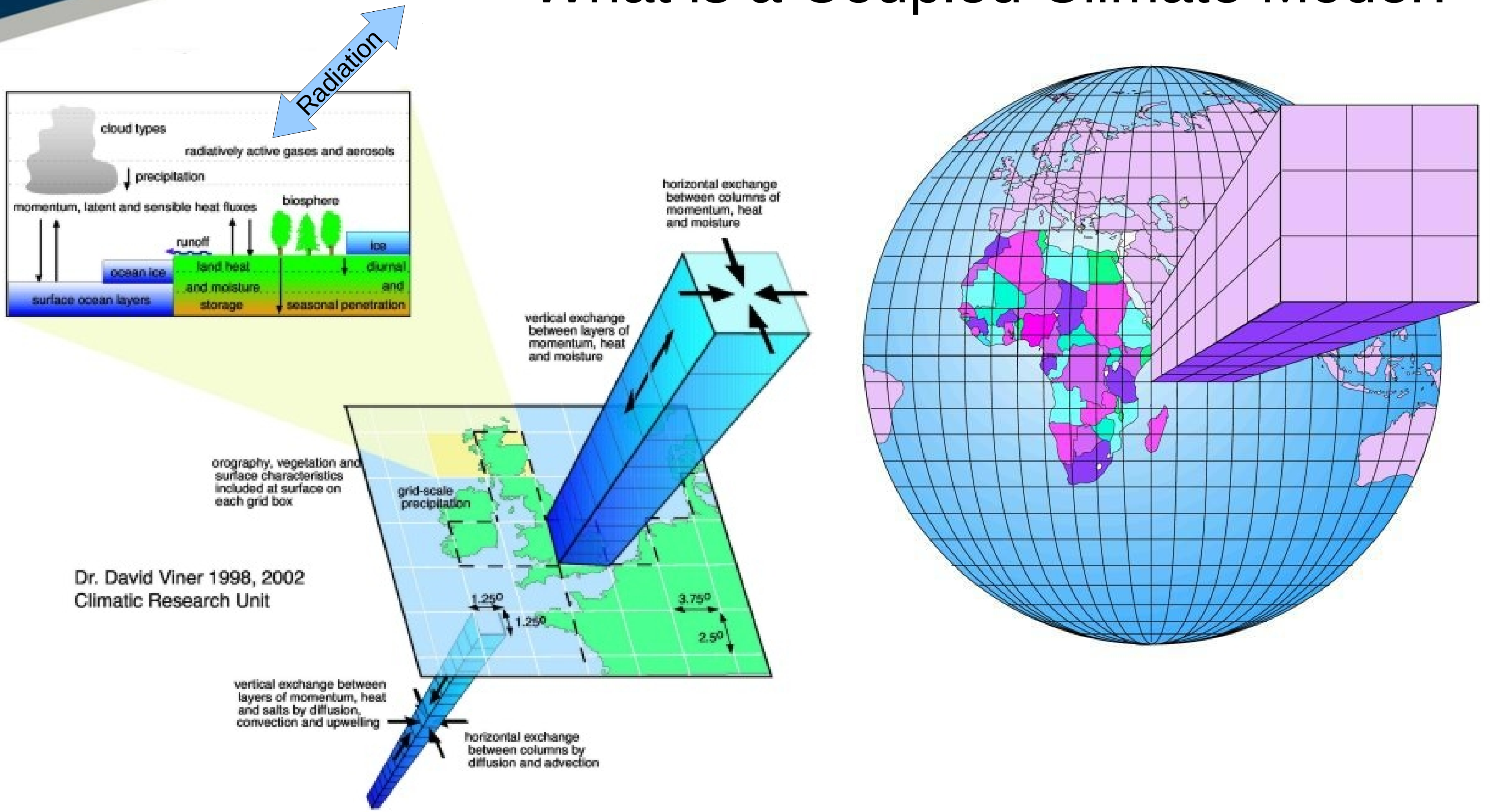


Image: from J. Lafeuille, 2006

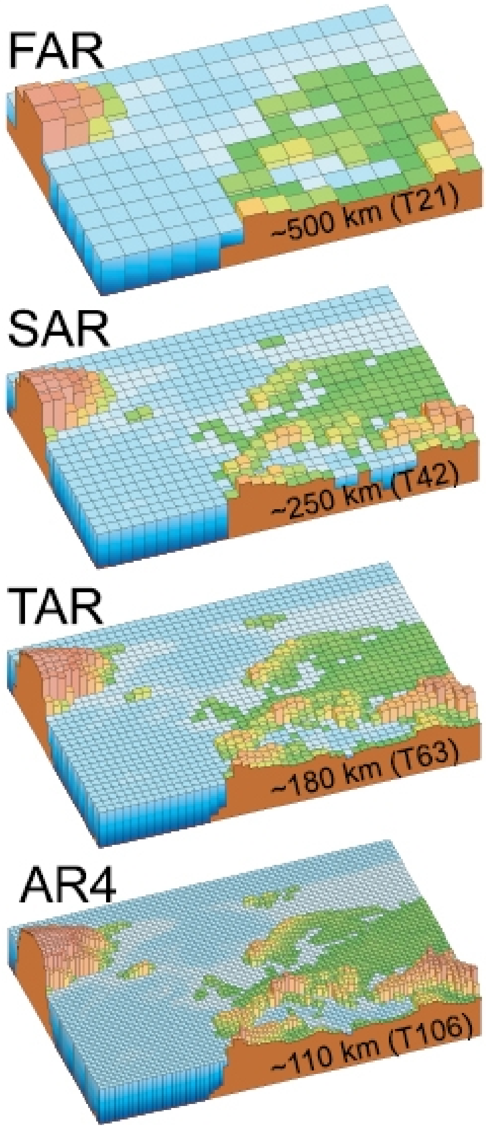
What is a Coupled Climate Model?

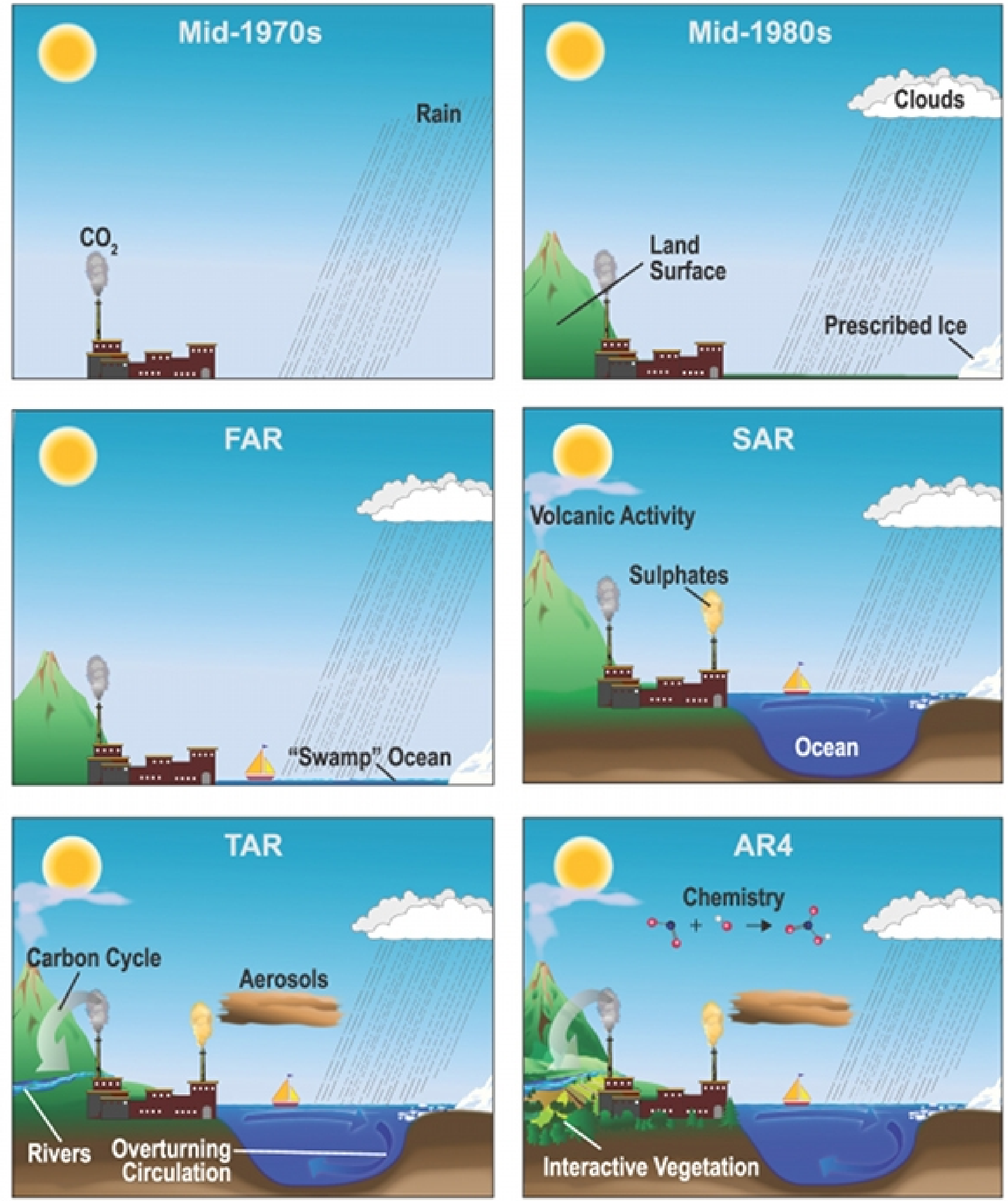


Answer: Lots of coupled partial differential equations solved via interative numerical techniques. Grid resolution controls whether equations really represent processes or parameterised verisions of them (which will have some statistical properties).

The World in Global Climate Models

FAR:1990
SAR:1995
TAR:2001
AR4:2007
AR5:2013





Science & Technology Facilities Council

State of the Art: Model Comparison

Table 1 The models used in the present study, including, configurations (near the equator) and number of years of simulations

Model	Institution	Atmosphere resolution	Ocean resolution	Length picntrl	Length 1pctto2x	Length 1pctto4x
CCSM3	NCAR (USA)	T85L26	1.125°×0.27°L40	230	150	n/a
CGCM3.1(T47)	CCCMA (Canada)	T47L31	1.85°×1.85°L29	500	150	150
CNRM-CM3	Meteo-France/CNRM (France)	T63L45	2°×0.5°L31	390	100	110
CSIRO-Mk3.0	CSIRO (Australia)	T63L18	1.875°×0.84°L31	380	10	n/a
ECHAM5/MPI-OM	MPI-M (Germany)	T63L31	1.5°×0.5°L40	332	100	81
FGOALS-g1.0	LASG/IAP (China)	T42L26	1°×1°L33	150	80	n/a
GFDL-CM2.0	GFDL (USA)	2.5°×2°L24	1°×0.33°L50	500	100	160
GFDL-CM2.1	GFDL (USA)	2.5°×2°L24	1°×0.33°L50	500	150	160
GISS-AOM	NASA/GISS (USA)	4°×3°L12	4°×3°L16	251	n/a	n/a
GISS-EH	NASA/GISS (USA)	5°×4°L20	2°×2°L16	500	80	150
GISS-ER	NASA/GISS (USA)	5°×4°L20	5°×4°L13	400	100	n/a
INM-CM3	INM (Russia)	5°×4°L21	2.5°×2°L33	330	n/a	n/a
IPSL-CM4	IPSL (France)	2.5°×3.75°L19	2°×0.5°L31	230	80	n/a
MIROC3.2(hires)	CCSR/NIES/FRCGC (Japan)	T106L56	0.28°×0.1875°L47	100	10	n/a
MIROC3.2(medres)	CCSR/NIES/FRCGC (Japan)	T42L20	1.4°×0.5°L43	500	100	150
MRI-CGM2.3.2	MRI (Japan)	T42L30	2.5°×0.5°L23	350	150	150
PCM	NCAR (USA)	T42L18	0.66°×0.5°L32	350	96	90
UKMO-HadCM3	HadleyCentre (UK)	3.75°×2.5°L19	1.25°×1.25°L20	341	10	n/a
UKMO-HadGEM1	HadleyCentre (UK)	1.875°×1.25°L38	1°×0.33°L40	80	10	n/a
SINTEX T30	IPSL/INGV (France,Italy)	T30L19	2°×0.5°L31	200	n/a	n/a
SINTEX T106	INGV/IPSL (Italy,France)	T106L19	2°×0.5°L31	100	n/a	n/a
SINTEX T106mod	IPSL/INGV (France,Italy)	T106L19	2°×0.5°L31	100	n/a	n/a
HadOPA	CGAM/IPSL (UK,France)	3.75°×2.5°L19	2°×0.5°L31	100	n/a	n/a

The only flux corrected model is MRI-CGM2.3.2

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

Guilyardi E. (2006): El Niño- mean state - seasonal cycle interactions in a multi-model ensemble. Clim. Dyn., 26:329-348, DOI: 10.1007/s00382-005-0084-6

Science & Technology Facilities Council

State of the Art: Model Comparison

TABLE 1. List of IPCC global coupled climate models analyzed in the present study and t
Model resolution is characterized by the size of a horizontal grid on which model output was
levels. Spectral models are also characterized by their spectral truncations. Equilibrium climat

Model label and climate sensitivity	Resolution	Institution a
CGCM3.1(T47) 3.6 K	96 × 48 L32 T47	Canadian Centre for Climate Modelling (http://www.cccma.ec.gc.ca/models/cgc
CGCM3.1(T63) 3.4 K	128 × 64 L32 T63	Canadian Centre for Climate Modelling (http://www.cccma.ec.gc.ca/models/cgc
CNRM-CM3 n/a	128 × 64 L45 T63	Centre National de Recherche Météoro manuscript submitted to *Climate Dyn*
ECHAM5/MPI-OM 3.4 K	192 × 96 L31 T63	Max-Planck-Institut für Meteorologie, C
ECHO-G 3.2 K	96 × 48 L19 T30	Meteorological Institute of the Universi Research Institute, South Korea (Mir
GFDL-CM2.0 2.9 K	144 × 90 L24	Geophysical Fluid Dynamics Laborator et al. 2006)
GFDL-CM2.1 3.4 K	144 × 90 L24	Geophysical Fluid Dynamics Laborator et al. 2006)
GISS-AOM n/a	90 × 60 L12	Goddard Institute for Space Studies Lal http://aom.giss.nasa.gov)
GISS-ER 2.7 K	72 × 46 L20	Goddard Institute for Space Studies Lal Russell et al. 2000)
INM-CM3.0 2.1 K	72 × 45 L21	Institute of Numerical Mathematics, Ru
IPSL-CM4.0 4.4 K	96 × 72 L19	Institut Pierre-Simon Laplace, France (http://dods.ipsl.jussieu.fr/omamce/IPS
MIROC3.2(hires) 4.3 K	320 × 160 L56 T106	Center for Climate System Research, Ja
MIROC3.2(medres) 4.0 K	128 × 64 L20 T42	Center for Climate System Research, Ja
MRI-CGCM2.3.2 3.2 K	128 × 64 L30 T42	Meteorological Research Institute, Japa
NCAR-CCSM3 2.7 K	256 × 128 L26 T85	National Center for Atmospheric Resea
NCAR-PCM 2.1 K	128 × 64 L26 T42	National Center for Atmospheric Resea et al. 2006)

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

TABLE 2. Description of model parameterizations for stratiform (i.e., large scale) and convective precipitation.

Model name	Stratiform precipitation	Convective precipitation
CCSM3, CCSM2	Prognostic condensate and precipitation parameterization (Zhang et al. 2003)	Simplified Arakawa and Schubert (1974) (cumulus ensemble) scheme developed by Zhang and McFarlane (1995)
CGCM3.1	Precipitation occurs whenever the local relative humidity is supersaturated	Zhang and McFarlane (1995) scheme
CNRM-CM3	Statistical cloud scheme of Ricard and Royer (1993)	Mass flux convection scheme with Kuo-type closure
CSIRO-Mk3.0	Stratiform cloud condensate scheme from Rotstayn (2000)	Bulk mass flux convection scheme with stability-dependent closure (Gregory and Rowntree 1990)
ECHAM5/MPI-OM	Prognostic equations for the water phases, bulk cloud microphysics (Lohmann and Roeckner 1996)	Bulk mass flux scheme (Tiedtke 1989) with modifications for deep convection according to Nordeng (1994)
FGOALS-g1.0	Same as PCM	Zhang and McFarlane (1995) scheme
GFDL-CM2.0, GFDL-CM2.1	Cloud microphysics from Rotstayn (2000) and macrophysics from Tiedtke (1993)	Relaxed Arakawa–Schubert scheme from Moorthi and Suarez (1992)
GISS-AOM	Subgrid-relative humidity-based scheme	Subgrid plume and buoyancy-based scheme (online at http://aom.giss.nasa.gov/DOC4X3/ATMOC4X3.TXT)
GISS-ER	Prognostic stratiform cloud based on moisture convergence (Del Genio et al. 1996)	Bulk mass flux scheme by Del Genio and Yao (1993)
HadCM3	Large-scale precipitation is calculated based on cloud water and ice contents (similar to Smith 1990)	Bulk mass flux scheme (Gregory and Rowntree 1990), with the improvement by Gregory et al. (1997)
HadGEM1	Mixed phase cloud scheme (Wilson and Ballard 1999)	Revised bulk mass flux scheme
INM-CM3.0	Stratiform cloud fraction is calculated as linear function of relative humidity	Lagged convective adjustment after Betts (1986), but with changed referenced profile for deep convection
IPSL-CM4	Cloud cover and in-cloud water are deduced from the large-scale total water and moisture at saturation (Bony and Emmanuel 2001)	Moist convection is treated using a modified version (Grandpeix et al. 2004) of the Emanuel (1991) scheme
MIROC3.2-medres MIROC3.2-hires	Prognostic cloud water scheme based on Le Treut and Li (1991)	Prognostic closure of Arakawa–Schubert based on Pan and Randall (1998) with relative humidity–based suppression (Emori et al. 2001)
MRI-CGCM2.3.2a	Precipitation occurs whenever the local relative humidity is supersaturated	Prognostic Arakawa–Schubert based on Pan and Randall (1998)
PCM	Precipitation occurs whenever the local relative humidity is supersaturated	Zhang and McFarlane (1995) scheme

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)

Model	Code	ElNiño amplitude	SST (°C) Niño3	τ_x(Pa) Niño4	SCRS (%)	ICS	2×CO_2 (%)	4×CO_2 (%)
Observed		0.88 ± 0.02	25.87 ± 0.05	−0.029 ± 0.006	31	8.7		
Obs 1948–1975		0.71 ± 0.04	25.72 ± 0.49	−0.032 ± 0.003		10.4		
Obs 1976–2004		0.94 ± 0.36	26.03 ± 0.06	−0.026 ± 0.000		8.5		
CCSM3	A	0.78 ± 0.04	25.29 ± 0.08	−0.038 ± 0.000	20	6.1	−13	
CGCM3.1(T47)	B	0.42 ± 0.03	24.63 ± 0.15	−0.045 ± 0.002	41	11.6	+5	+2
CNRM-CM3	C	1.66 ± 0.21	23.43 ± 0.06	−0.026 ± 0.000	3	6.3	+1	$+\underline{7}$
CSIRO-Mk3.0	D	0.90 ± 0.17	24.34 ± 0.23	−0.034 ± 0.000	20	7.8		
ECHAM5/MPI-OM	E	1.16 ± 0.09	25.16 ± 0.06	−0.034 ± 0.001	13	7.3	$+\underline{29}$	$+\underline{31}$
FGOALS-g1.0	F	1.93 ± 0.34	26.57 ± 0.16	−0.028 ± 0.001	0	6.6	$-\underline{27}$	
GFDL-CM2.0	G	0.75 ± 0.19	24.74 ± 0.15	−0.043 ± 0.000	37	8.8	$+\underline{20}$	$+\underline{25}$
GFDL-CM2.1	H	1.32 ± 0.08	24.98 ± 0.14	−0.044 ± 0.000	12	12.8	+2	$-\underline{18}$
GISS-AOM	I	0.17 ± 0.03	27.07 ± 0.01	−0.036 ± 0.000	45	17		
GISS-EH	J	0.86 ± 0.13	24.53 ± 0.13	−0.037 ± 0.001	24	0.8	$-\underline{5}$	
GISS-ER	K	0.24 ± 0.01	28.16 ± 0.03	−0.026 ± 0.001	22	2.2	−21	+8
INM-CM3	L	0.92 ± 0.10	24.15 ± 0.09	−0.025 ± 0.001	23	6.2		
IPSL-CM4	M	1.00 ± 0.02	26.28 ± 0.08	−0.026 ± 0.000	13	5.9	$-\underline{16}$	
MIROC3.2(hires)	N	0.35 ± 0.01	25.46 ± 0.14	−0.042 ± 0.002	86	15.4		
MIROC3.2(medres)	O	0.44 ± 0.11	24.81 ± 0.03	−0.040 ± 0.000	60	10.7	+5	+2
MRI-CGM2.3.2	P	0.70 ± 0.05	25.04 ± 0.04	−0.045 ± 0.000	35	16	$+\underline{34}$	$+\underline{77}$
PCM	Q	0.89 ± 0.19	24.23 ± 0.11	−0.034 ± 0.001	11	6.1	$-\underline{8}$	$-\underline{13}$
UKMO-HadCM3	R	0.77 ± 0.09	25.58 ± 0.07	−0.045 ± 0.001	13	10.3		
UKMO-HadGEM1	S	0.68 ± 0.17	23.69 ± 0.12	−0.064 ± 0.001	28	8.9		
SINTEXT30	T	0.61 ± 0.09	25.90 ± 0.08	−0.041 ± 0.001	13	8.5		
SINTEXT106	U	0.74 ± 0.07	26.27 ± 0.16	−0.035 ± 0.002	5	7.0		
SINTEXT106mod	V	0.67 ± 0.06	26.84 ± 0.25	−0.041 ± 0.002	8	6.6		
HadOPA	W	1.67 ± 0.14	27.46 ± 0.36	−0.035 ± 0.001	5	7.5		

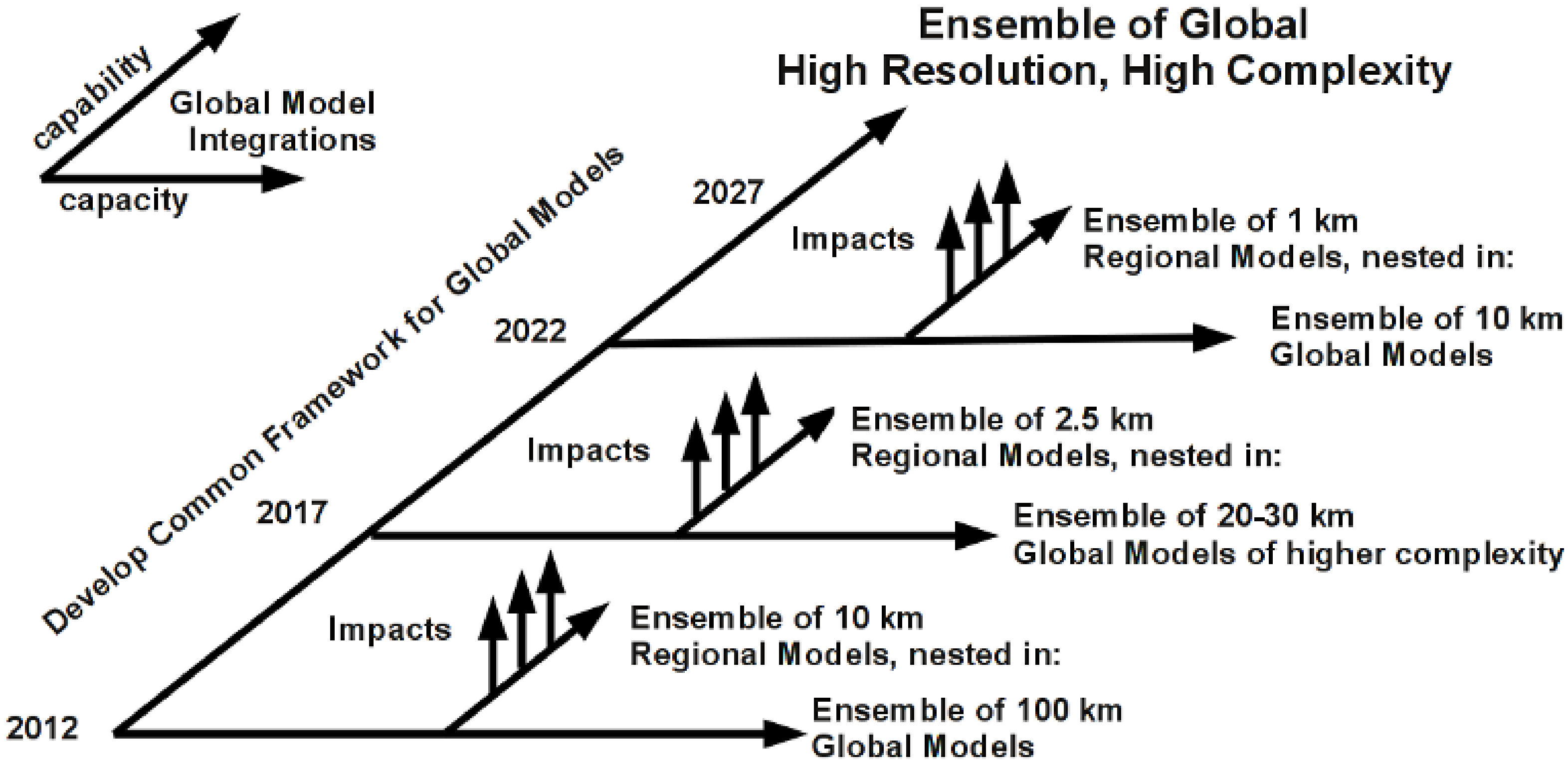
CSRS is the seasonal cycle relative strength (in %), ICS the summer interannual coupling strength (in 10^{-3} Pa/C). The El Niño amplitude change to doubling and quadrupling of CO_2 (when compared to picntrl) are shown in the last two columns. The El Niño amplitude is defined as the standard deviation of SST in the Nino3 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 months for means). The amplitude change values underlined

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

Guilyardi E. (2006): El Niño- mean state - seasonal cycle interactions in a multi-model ensemble. Clim. Dyn., 26:329-348, DOI: 10.1007/s00382-005-0084-6

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

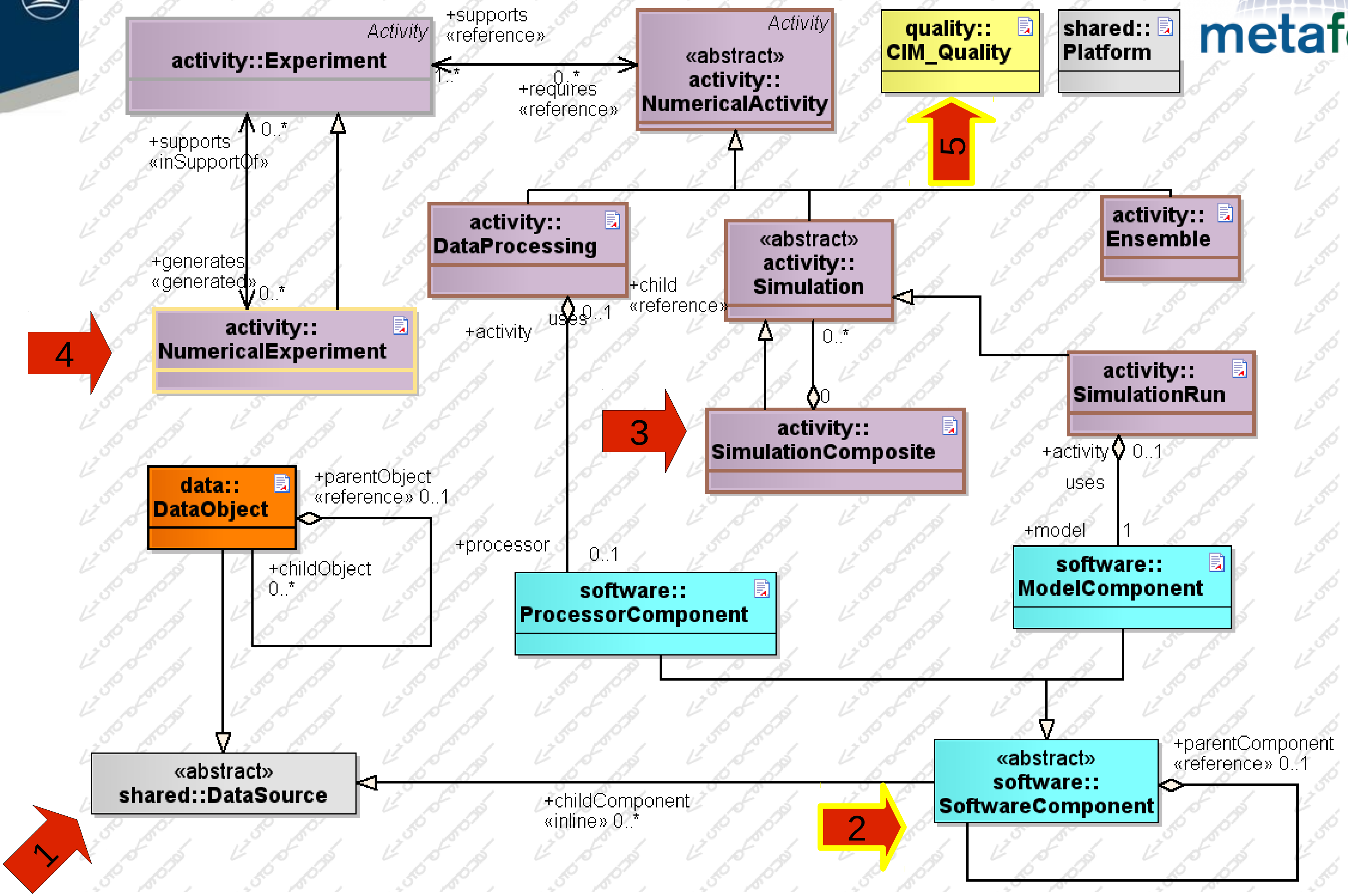
- Simulations
  - Models run for specific experiments (and hence specific "boundary conditions", e.g. CO_2 projections)
  - with specific output variables, frequencies, and durations
  - run on specific platforms

Metafor:

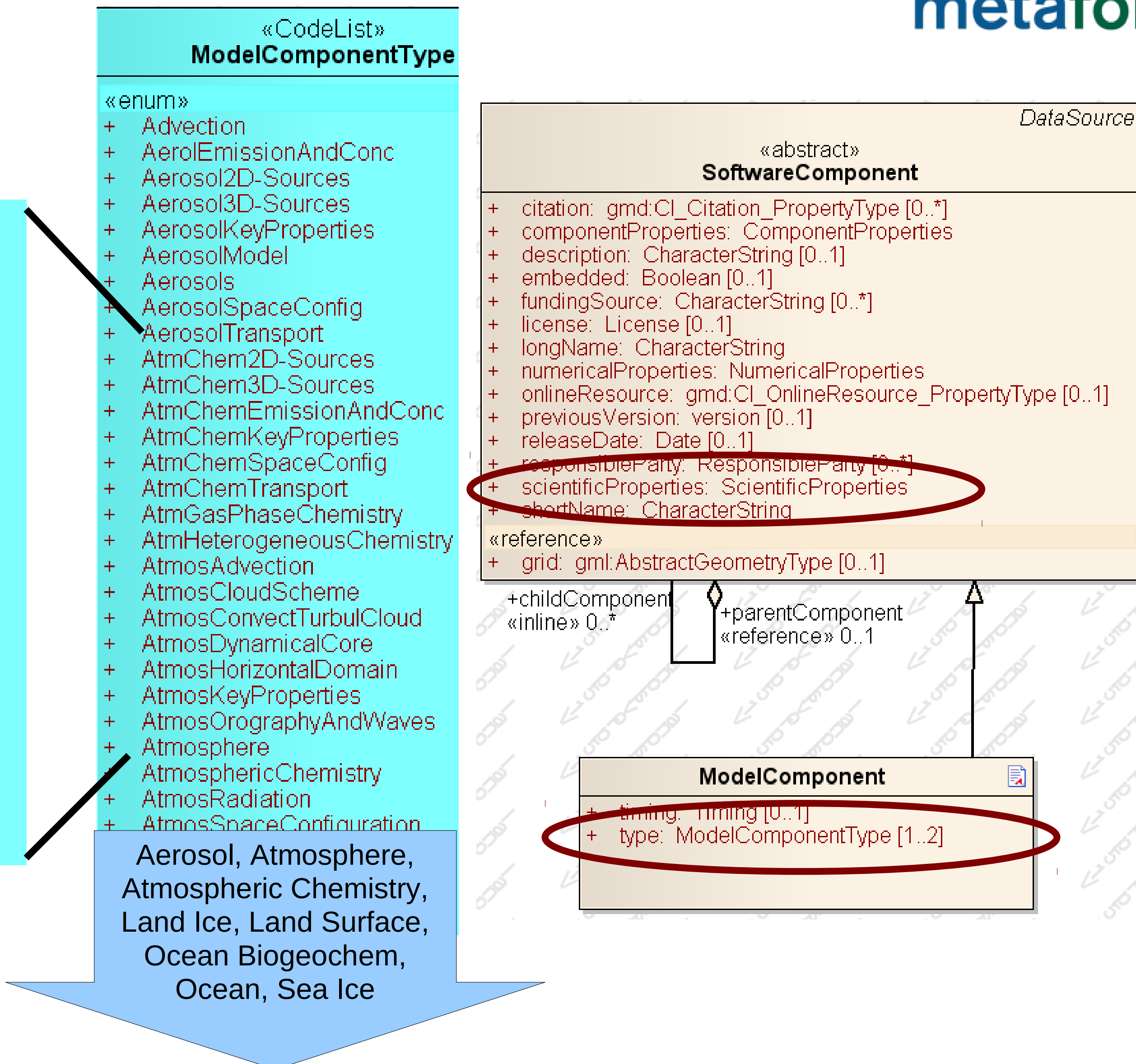
- 2.2 M euro EC project to deliver
  - a "Common Information Model" to document these concepts, particularly in the context of supporting the next IPCC assessment report.
  - Associated infrastructure to collect and view such documentation, and
  - Build an accompanying governance structure ...

Science & Technology

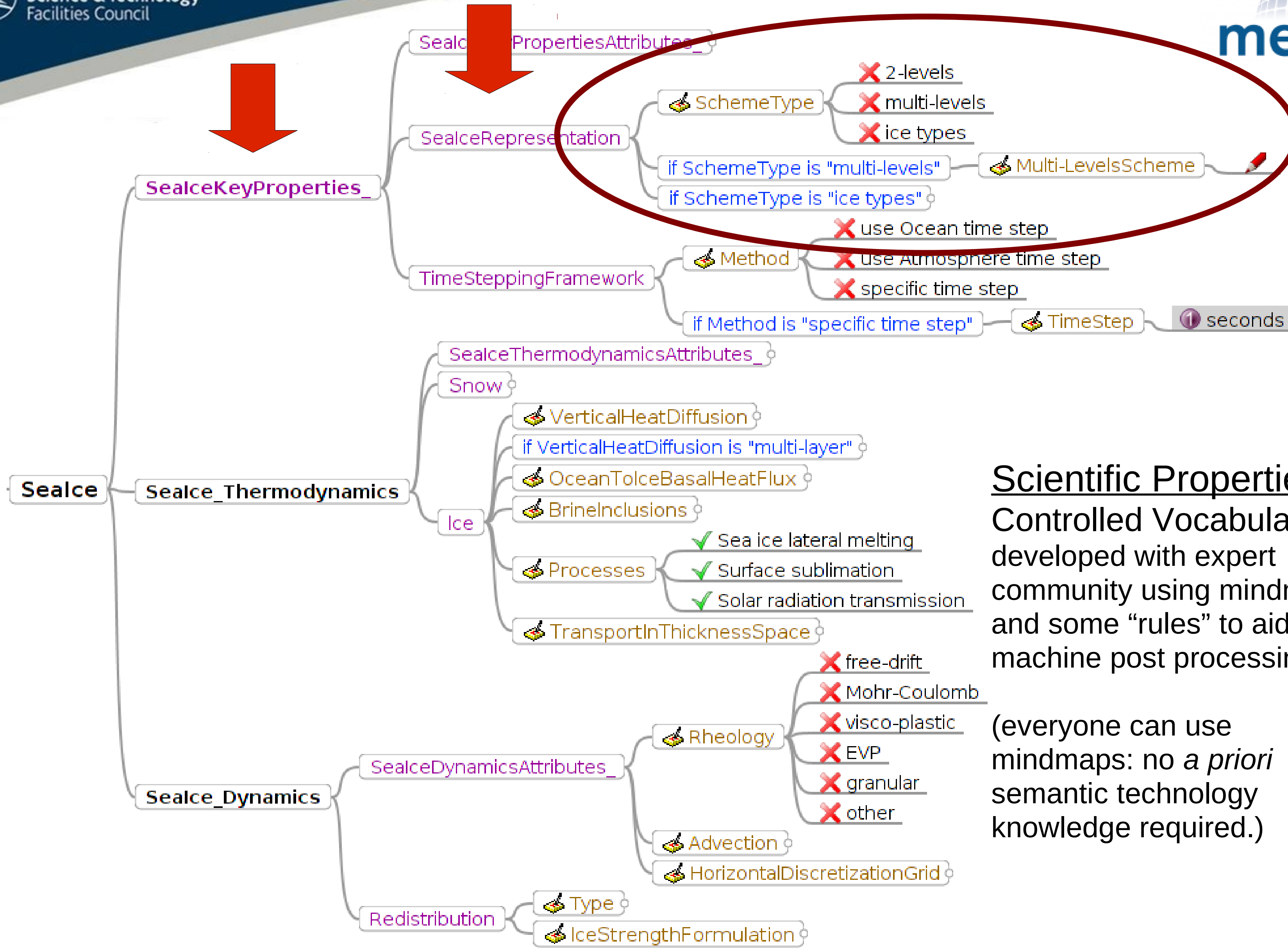
metafor



AtmChem2D-Sources
AtmChem3D-Sources
AtmChemEmissionAndConc
AtmChemKeyProperties
AtmChemSpaceConfig
AtmChemTransport
AtmGasPhaseChemistry
AtmHeterogeneousChemistry
AtmosAdvection
AtmosCloudScheme
AtmosConvectTurbulCloud
AtmosDynamicalCore
AtmosHorizontalDomain
AtmosKeyProperties
AtmosOrographyAndWaves
Atmosphere



(Yes we know we shouldn't have this sort of detail in the UML, and it wont be … shortly)



Scientific Properties:

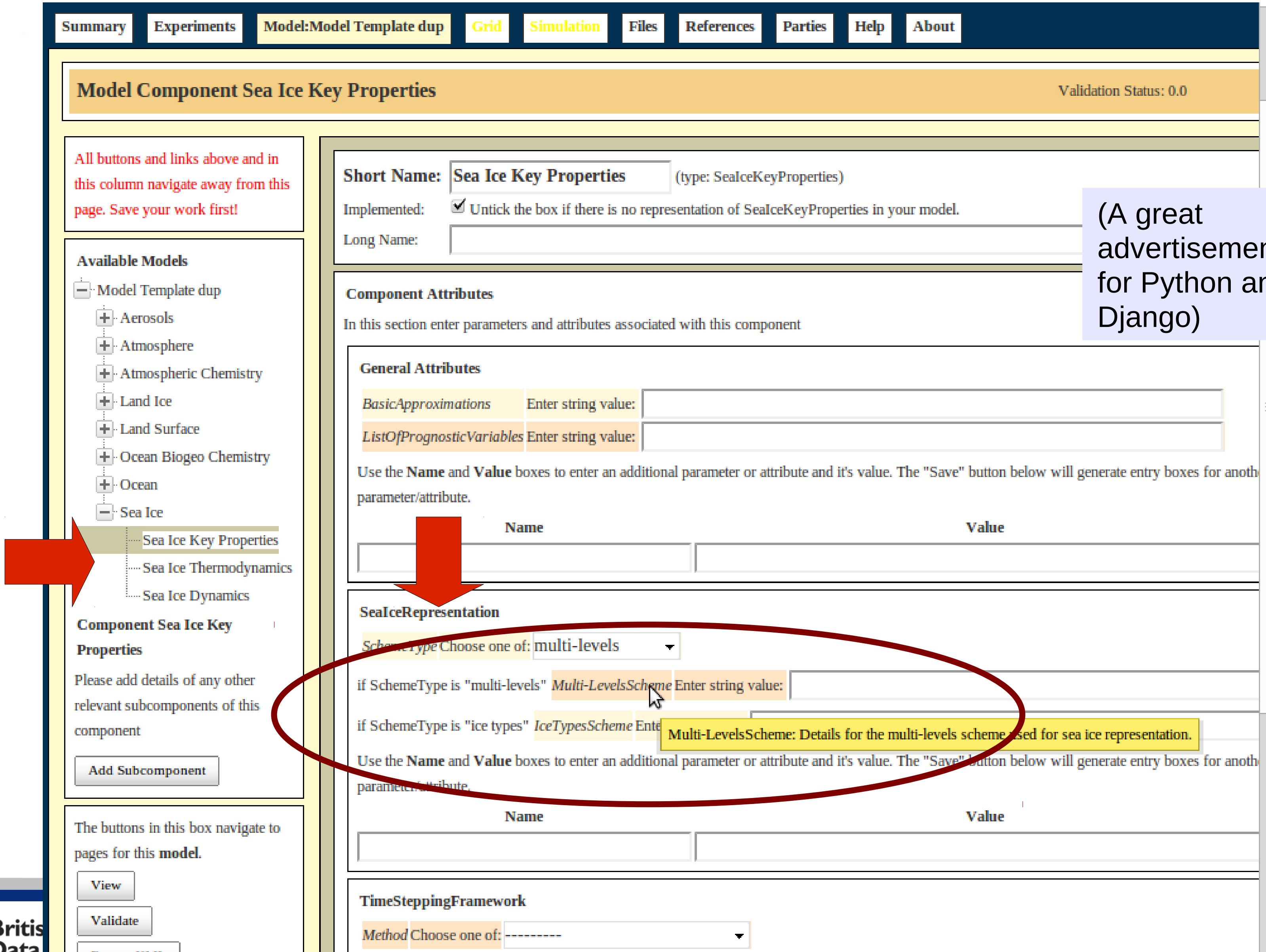
Controlled Vocabularies developed with expert community using mindmaps and some “rules” to aid machine post processing …

(everyone can use mindmaps: no *a priori* semantic technology knowledge required.)

```
-<node COLOR="#0033ff" CREATED="1233752067356" ID="Freemind_Link_884094832"
 MODIFIED="1255963916698" TEXT="if SchemeType is "multi-levels"">
   <font NAME="SansSerif" SIZE="14"/>
  -<node BACKGROUND_COLOR="#ffffff" COLOR="#996600" CREATED="1233752067356"
   ID="Freemind_Link_1138073599" MODIFIED="1257330309017" TEXT="Multi-LevelsScheme">
     <font NAME="SansSerif" SIZE="14"/>
    -<hook NAME="accessories/plugins/NodeNote.properties">
      -<text>
         [definition]Details for the multi-levels scheme used for sea ice representation.[/definition]
       </text>
     </hook>
    -<node CREATED="1256041603108" ID="Freemind_Link_1643703382" MODIFIED="1257328758858"
     STYLE="fork" TEXT="">
       <font NAME="SansSerif" SIZE="14"/>
       <icon BUILTIN="pencil"/>
     </node>
   </node>
 </node>
```


A piece of the mindmap XML …

… vocabulary driven content in web based "human entry tool"



(A great advertisement for Python and Django)

CIM Tools

Production

CIM Applications	HTML+AJAX	HTML+AJAX	HTML+AJAX
Architecture Stack	Pylons	Plone	Django
Web Service Interfaces	REST	REST	REST
CIM Tools	XML Difference	Faceted Browse	Creation
Query Interfaces	Xquery	Sparql	Django queryset
(CIM Document Model)	XML-Schema	RDF-S or OWL	Django ORM
CIM Persistence	eXist XML db	triplestore	Relational DB

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

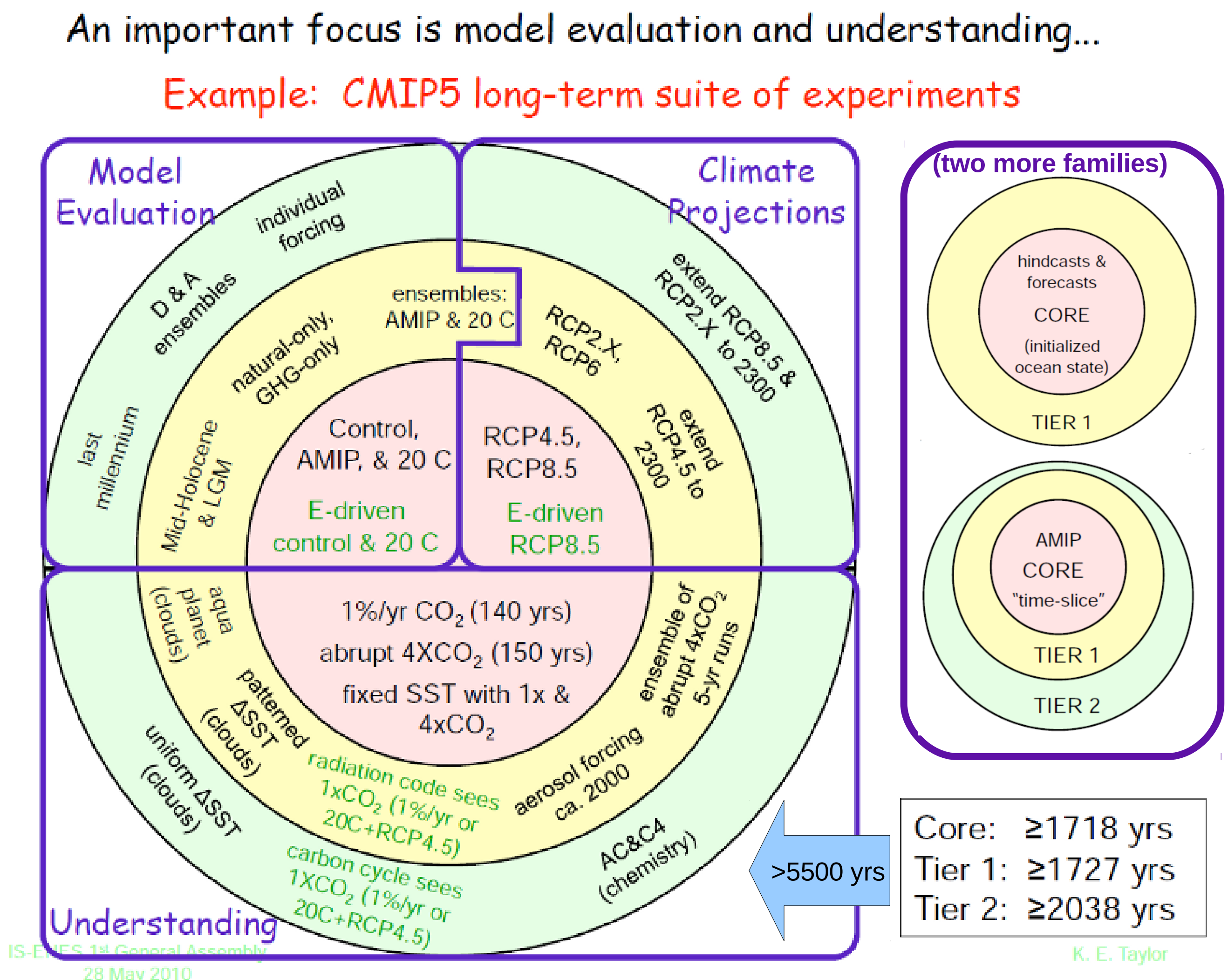
<u>CMIP5: Fifth Coupled Model Intercomparison Project</u>

- 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



Take home points here:

Many distinct experiments,

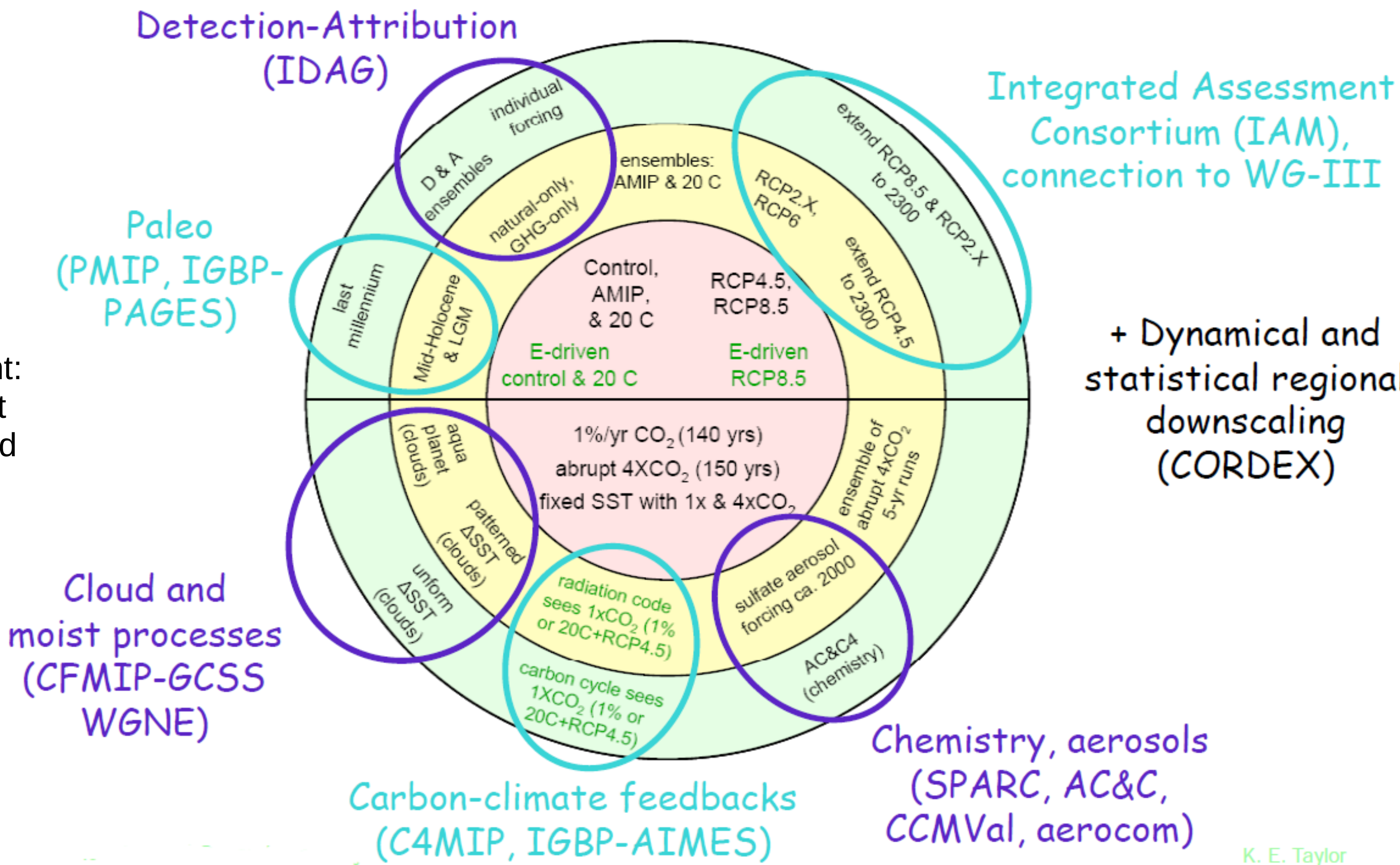
with very different characteristics,

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

Take home point:
- many different communities and projects



(from Karl Taylor)

CMIP5 in numbers

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

SI prefix	Name	Power of 10 or 2		Status
k kilo	thousand	10^{3}	2^{10}	Count on fingers
M mega	million	10^{6}	2^{20}	Trivial
G giga	billion	10^{9}	2^{30}	Small
T tera	trillion	10^{12}	2^{40}	Real
P peta	quadrillion	10^{15}	2^{50}	Challenging
E exa	quintillion	10^{18}	2^{60}	Aspirational
Z zetta	sextillion	10^{21}	2^{70}	Wacko
Y yotta	septillion	10^{24}	2^{80}	Science fiction

Stuart Feldman, Google

Handling the data!

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

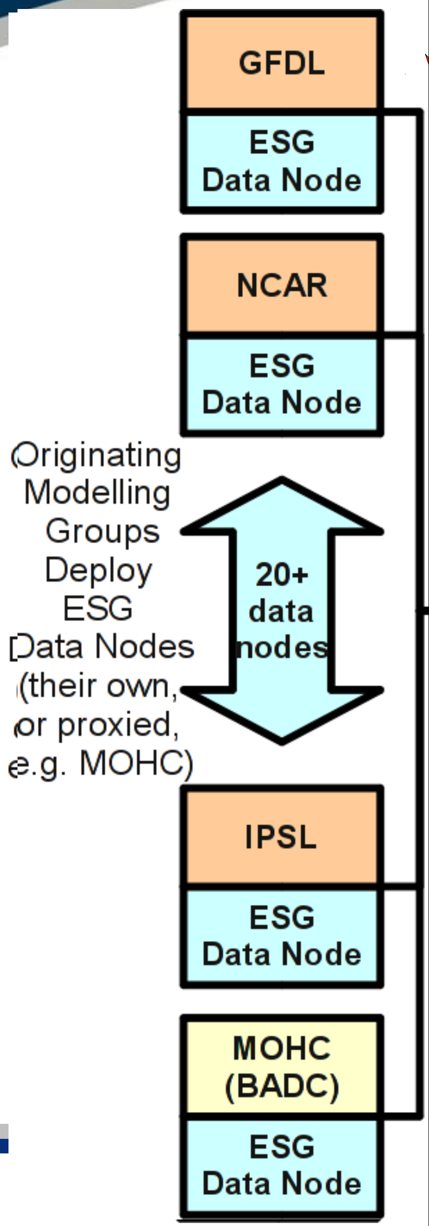
ESG Publisher inc CDAT

Bulk Data Mover (BDM)

Databases: Thredds Catalog, Postgres & Myproxy

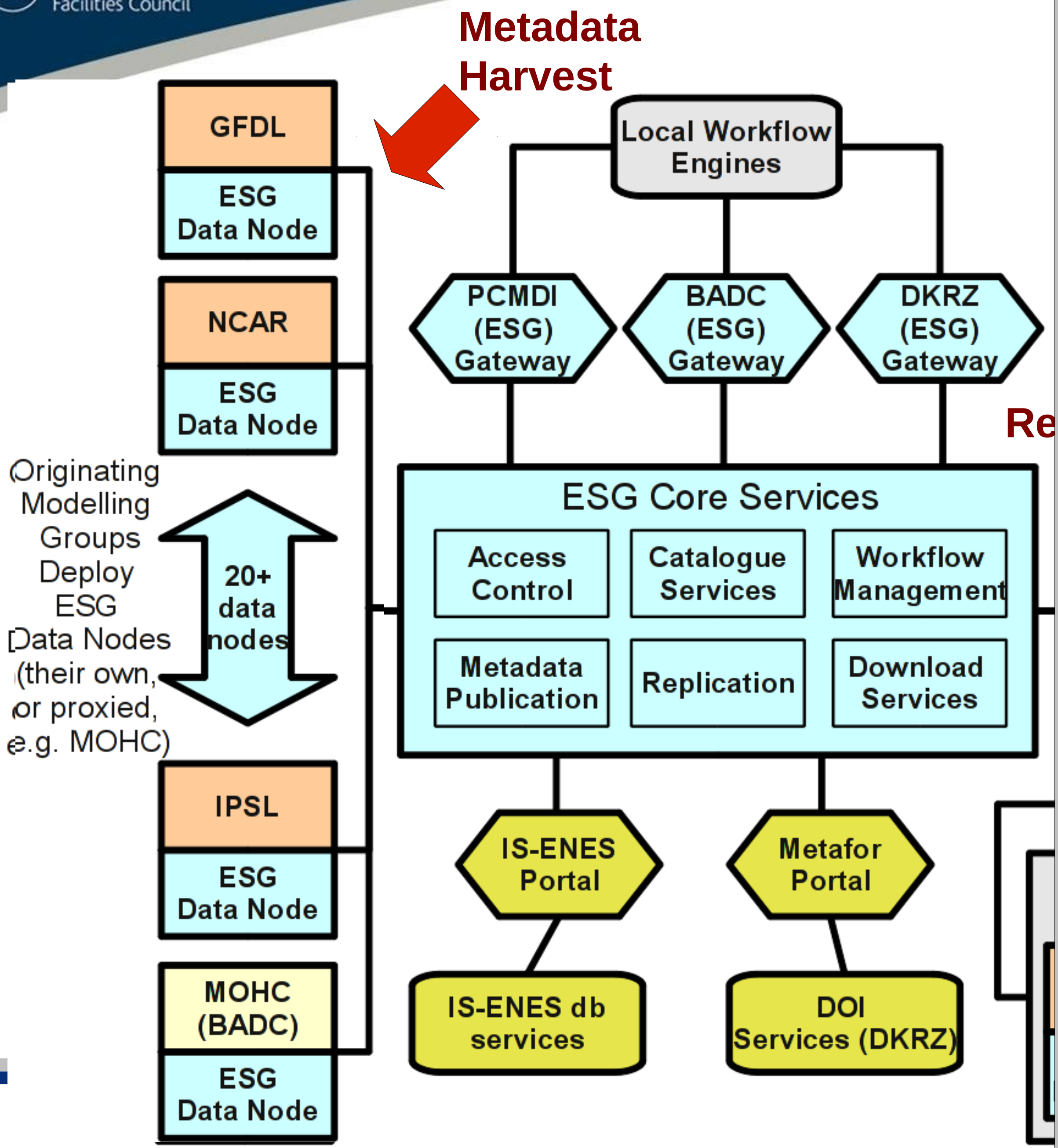
Filesystem: preferably with a DRS layout

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



ESGF

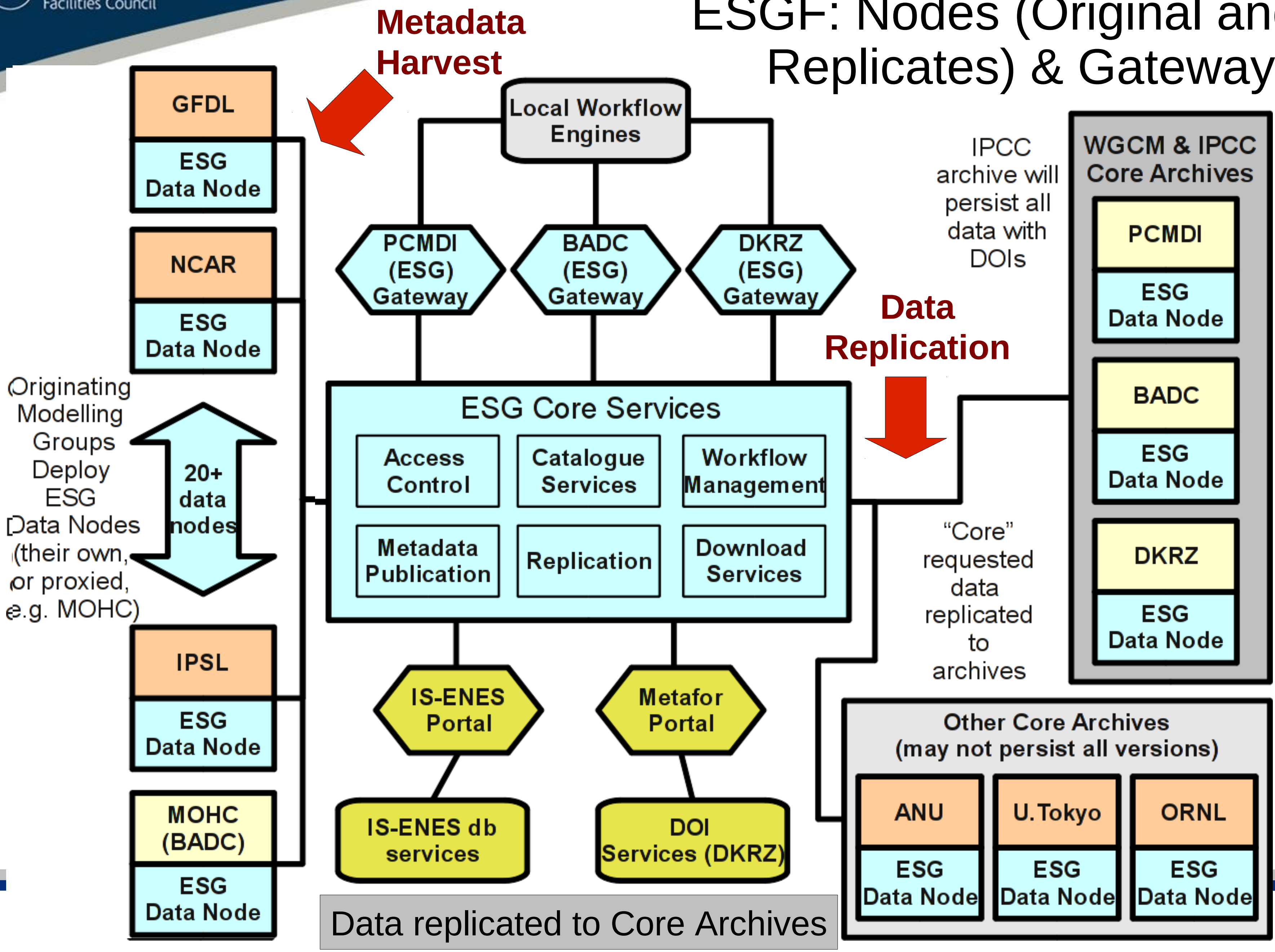
Data Nodes
With
Gateways

Data nodes
publish
to
Gateways

Gateways
replicate
metadata

(all data visible
on all gateways)

ESGF: Nodes (Original and Replicates) & Gateways



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”.

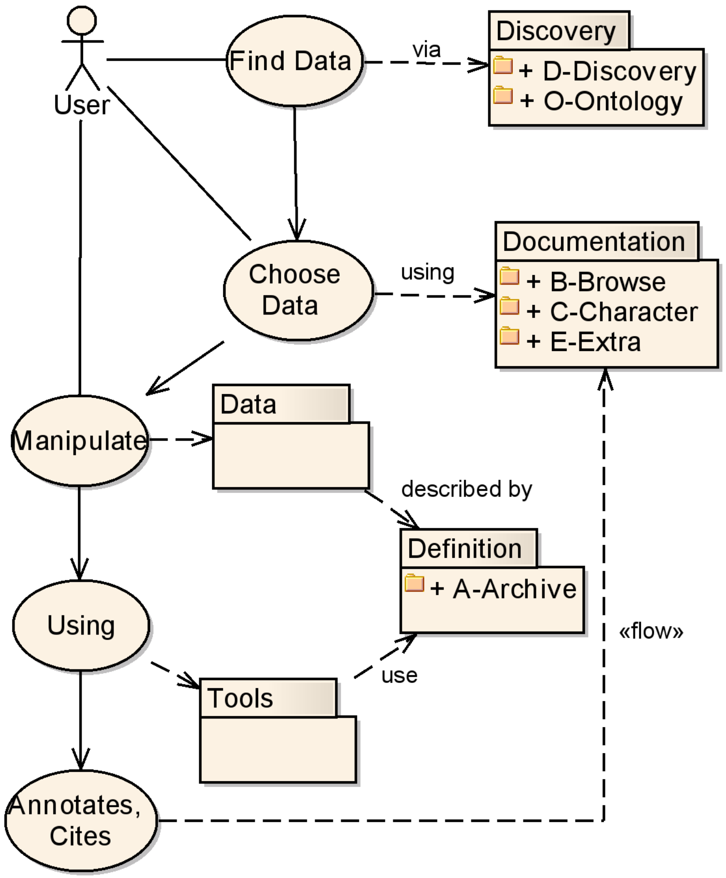
compliance enforced by CMIP5 quality control systems, leading to

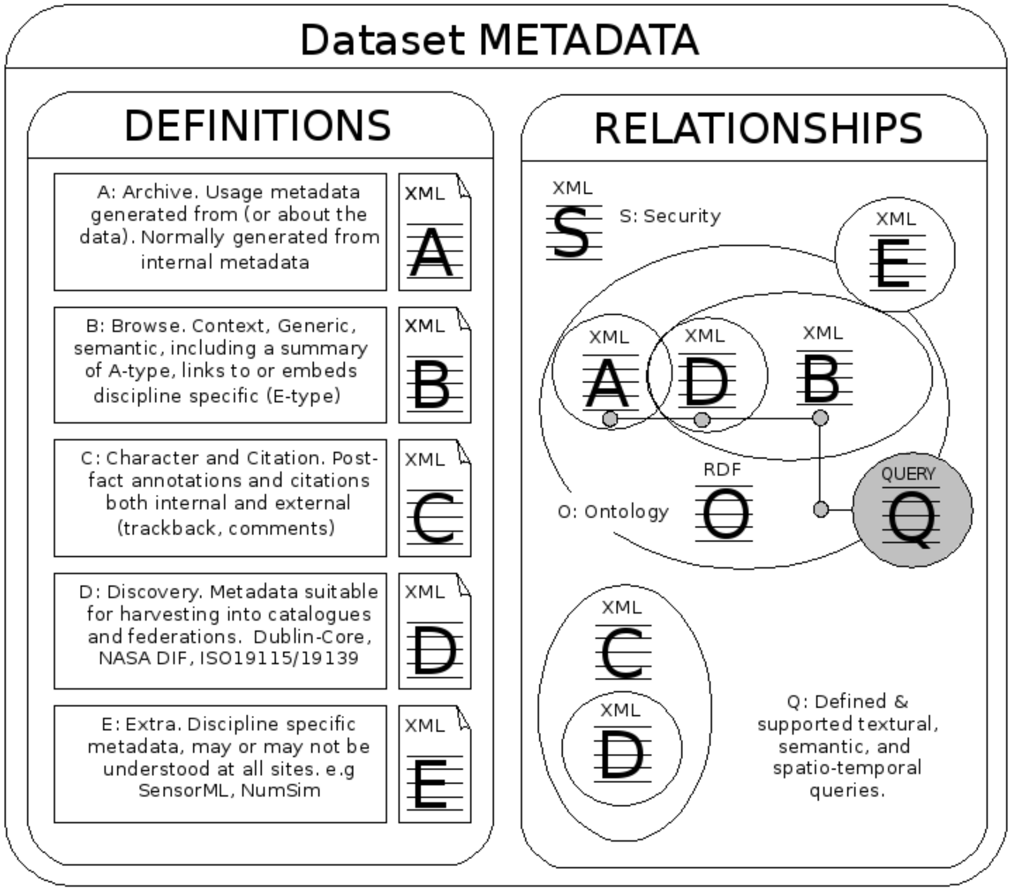
C: **Character** Metadata

- 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)

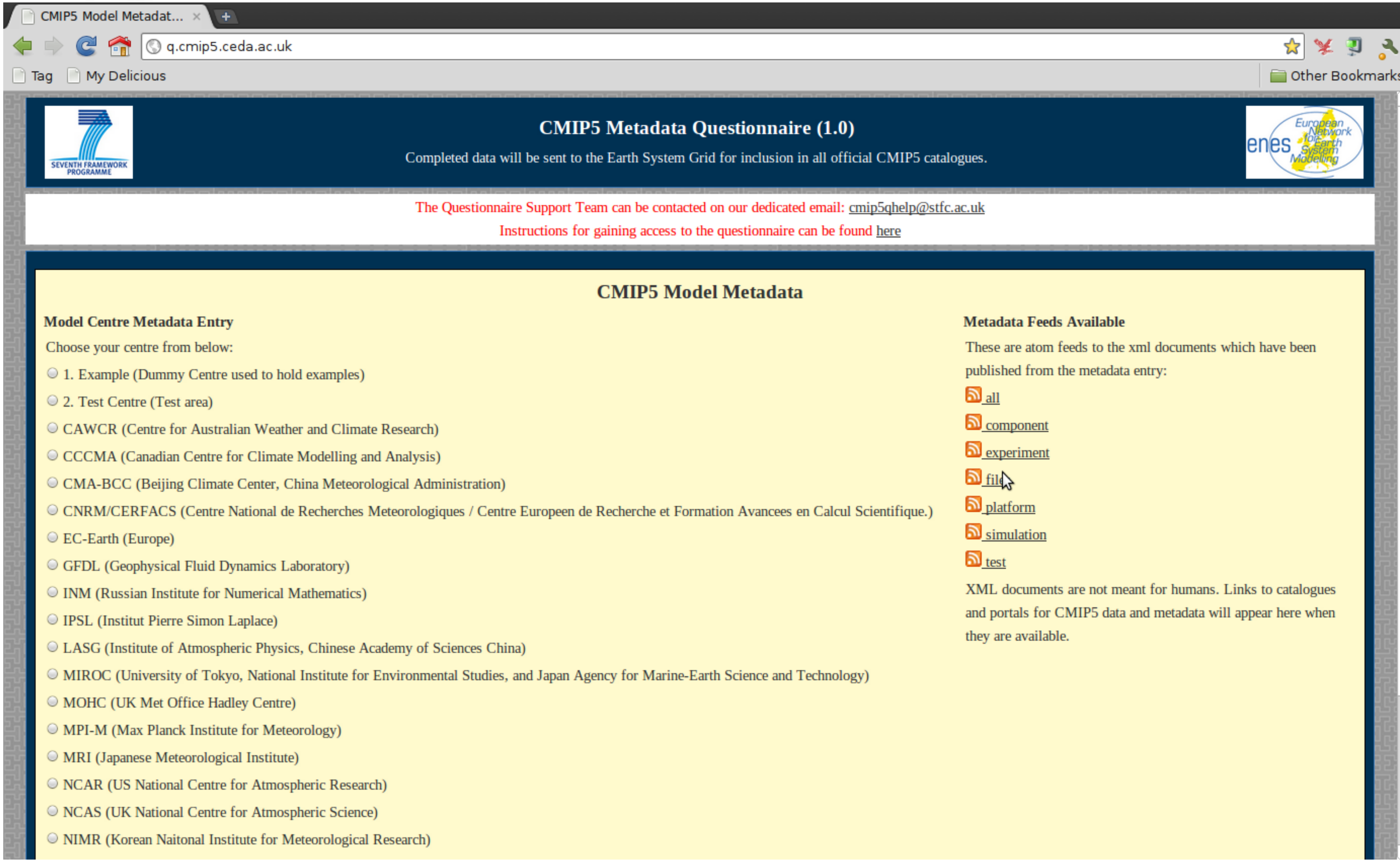
1) Metadata Taxonomy: Discovery, Documentation, Definition





In CMIP5 we haven't really addressed formal D (ISO19115 class) metadata yet

2) Metafor Questionnaire



CMIP5 Model Metadat...

q.cmip5.ceda.ac.uk

Tag | My Delicious | Other Bookmarks

CMIP5 Metadata Questionnaire (1.0)

Completed data will be sent to the Earth System Grid for inclusion in all official CMIP5 catalogues.

The Questionnaire Support Team can be contacted on our dedicated email: cmip5qhelp@stfc.ac.uk

Instructions for gaining access to the questionnaire can be found here

CMIP5 Model Metadata

Model Centre Metadata Entry

Choose your centre from below:

- 1. Example (Dummy Centre used to hold examples)
- 2. Test Centre (Test area)
- CAWCR (Centre for Australian Weather and Climate Research)
- CCCMA (Canadian Centre for Climate Modelling and Analysis)
- CMA-BCC (Beijing Climate Center, China Meteorological Administration)
- CNRM/CERFACS (Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique.)
- EC-Earth (Europe)
- GFDL (Geophysical Fluid Dynamics Laboratory)
- INM (Russian Institute for Numerical Mathematics)
- IPSL (Institut Pierre Simon Laplace)
- LASG (Institute of Atmospheric Physics, Chinese Academy of Sciences China)
- MIROC (University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology)
- MOHC (UK Met Office Hadley Centre)
- MPI-M (Max Planck Institute for Meteorology)
- MRI (Japanese Meteorological Institute)
- NCAR (US National Centre for Atmospheric Research)
- NCAS (UK National Centre for Atmospheric Science)
- NIMR (Korean Naitonal Institute for Meteorological Research)

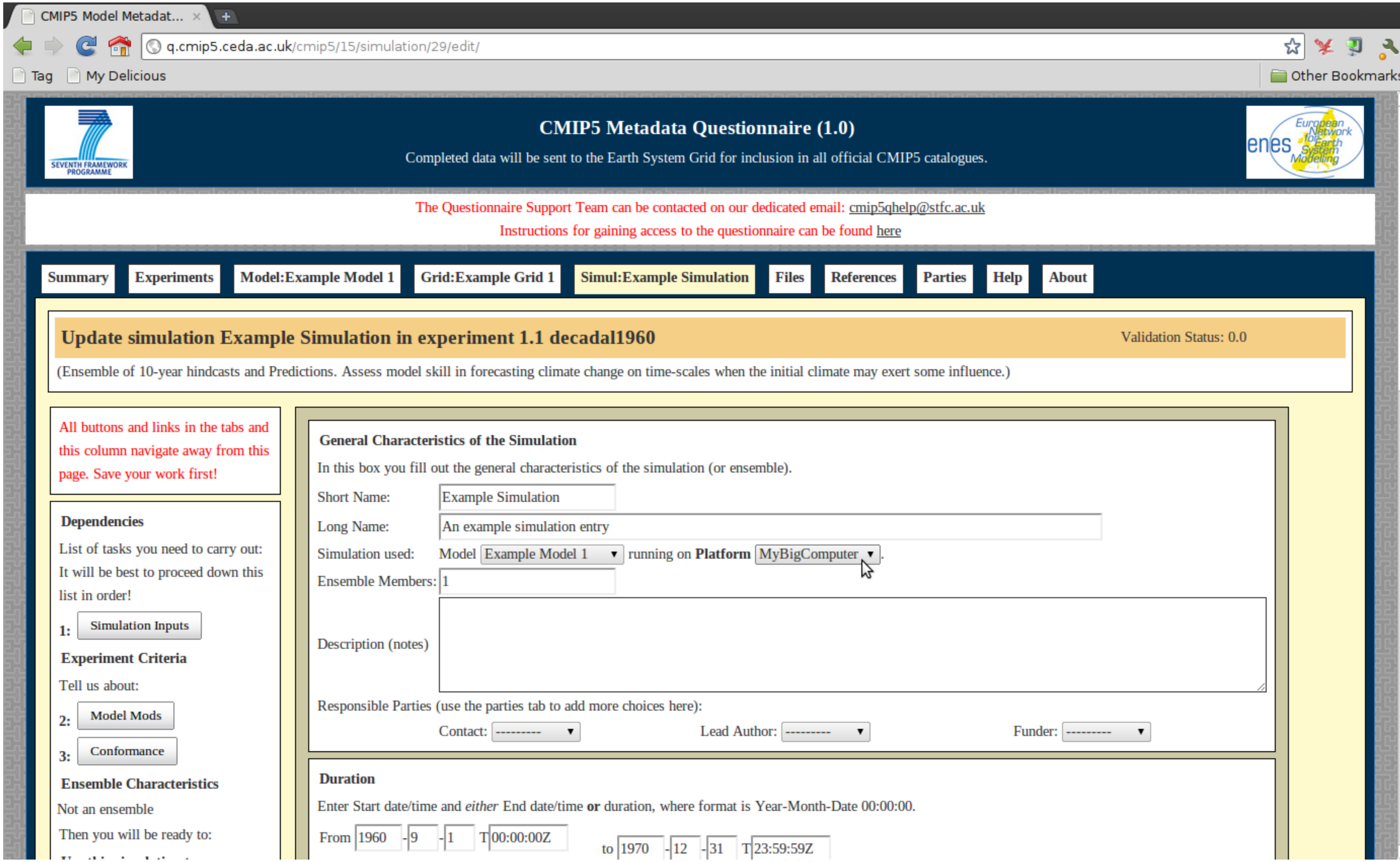
Metadata Feeds Available

These are atom feeds to the xml documents which have been published from the metadata entry:

- all
- component
- experiment
- file
- platform
- simulation
- test

XML documents are not meant for humans. Links to catalogues and portals for CMIP5 data and metadata will appear here when they are available.

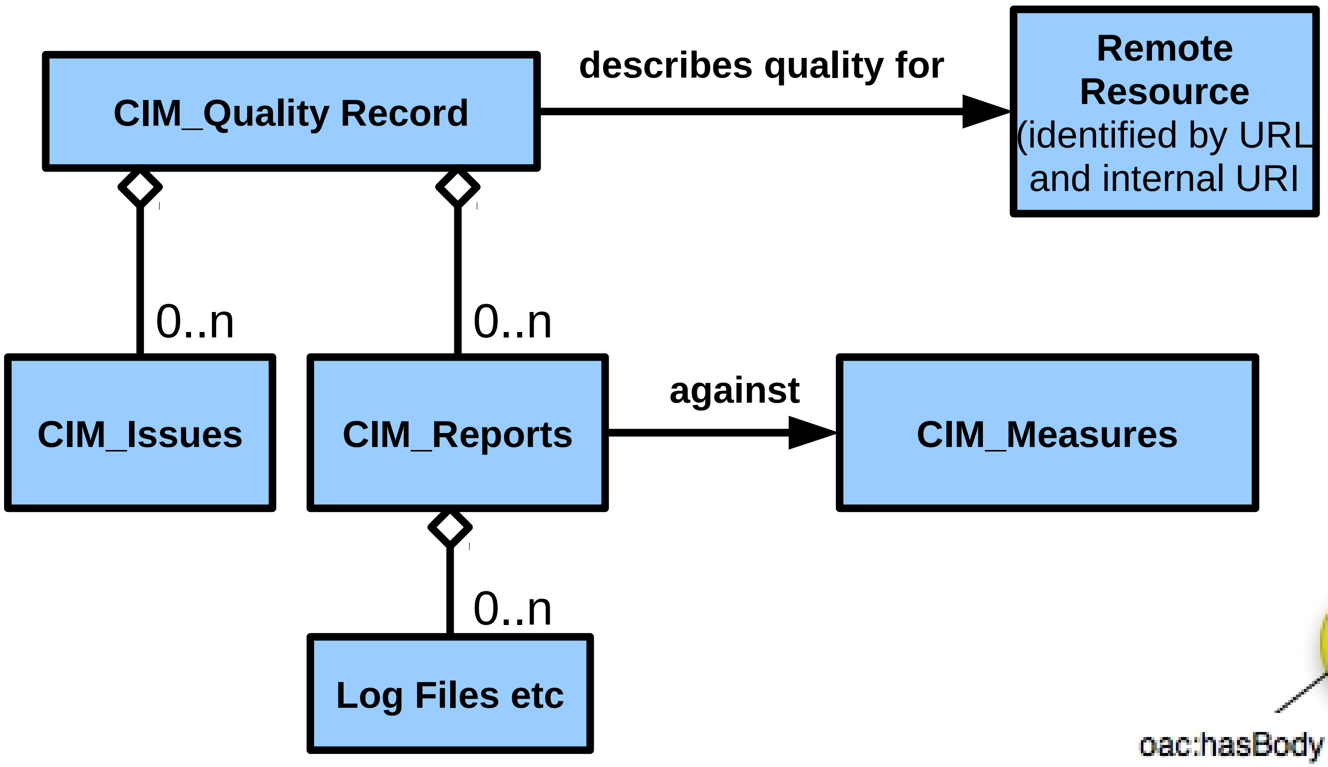
Metafor questionnaire: many parts ...



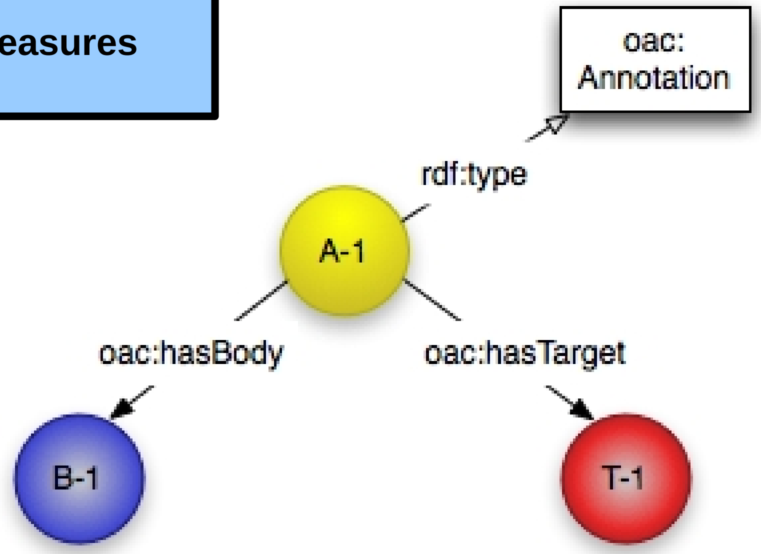
Metafor Quality Control

Specialises ISO19115 DQ package

Extrinsic!
Not intrinsic!



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



CMIP5 qctool (courtesy of Metafor)

Denmark Norway Japan United Kingdom Italy Russia S. Korea Germany France The Netherlands China Canada USA Australia

Home | Resources | Issues | Reports | Measures | Users | Admin | Atom:

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.

qctool v0.21 produced by CEDA for metafor

Hosted at the British Atmospheric Data Centre

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

(Another great advertisement for Python and Django)

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 signficant amount of automated quality control, coupled with scientific “spot checks”.

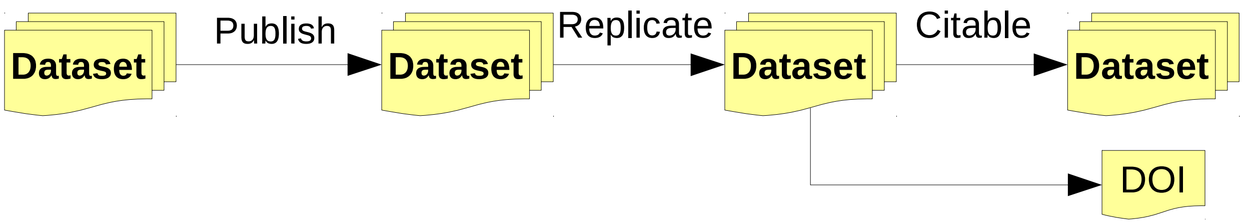
- The ESGF quality control will be according to a set of defined “qc levels”

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)

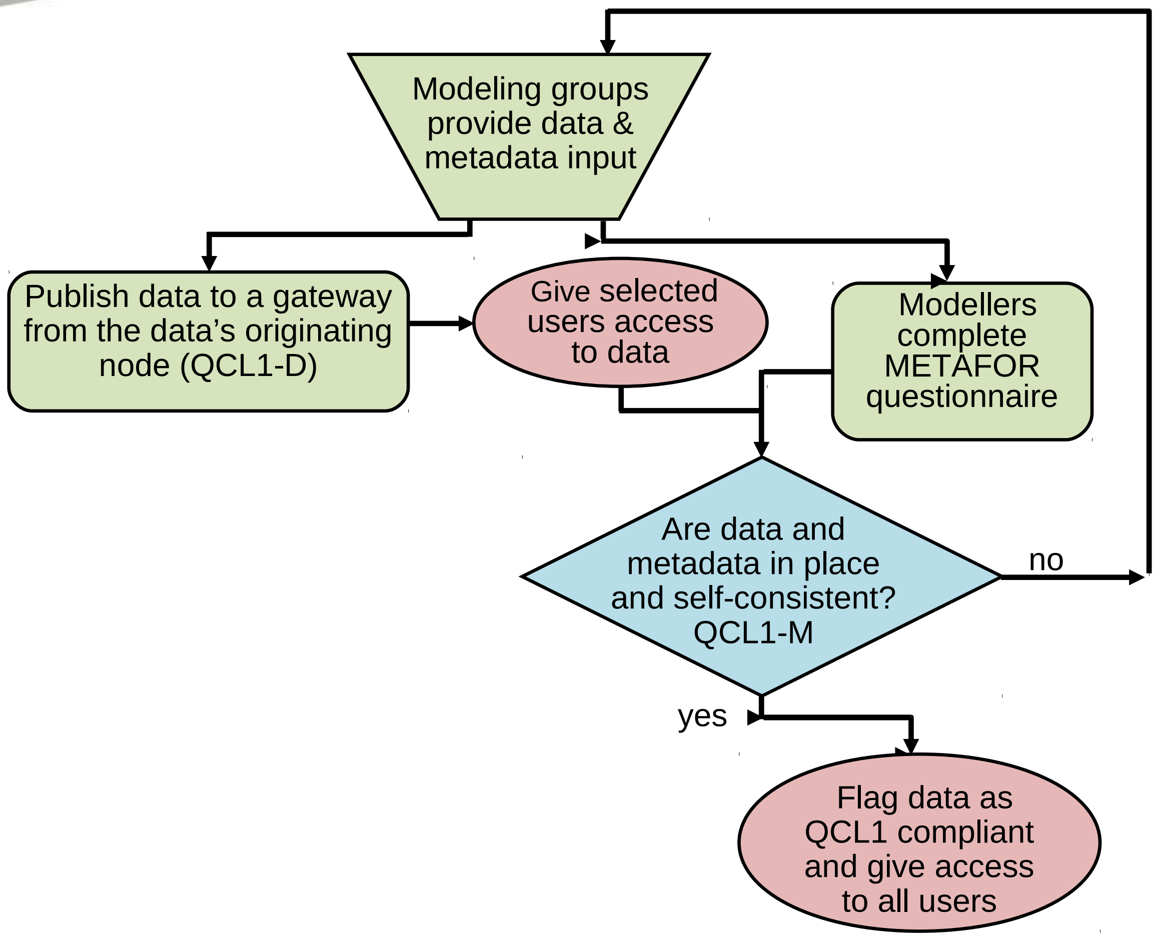
CMIP5 Quality Control



Label	Data	Metadata
qc-1d	ESG publisher enforces some data checking	CF compliance
qc-1m		Questionnaire enforces some constraints and vocabularies, requires XSD validation.
qc-2m		Subjective examination by metafor team.
qc-2d	Automated examination with subjective spot checks: carried out at PCMDI, DKRZ and BADC.	Provisional DOI granted.
qc-3	Further subjective tests at DKRZ, author approval of all metadata and output. Final DOI granted.	
Scientific Metrics	CMIP5 requires no scientific validation, but qc system will support data annotatoin against specific metrics of scientific interest.	

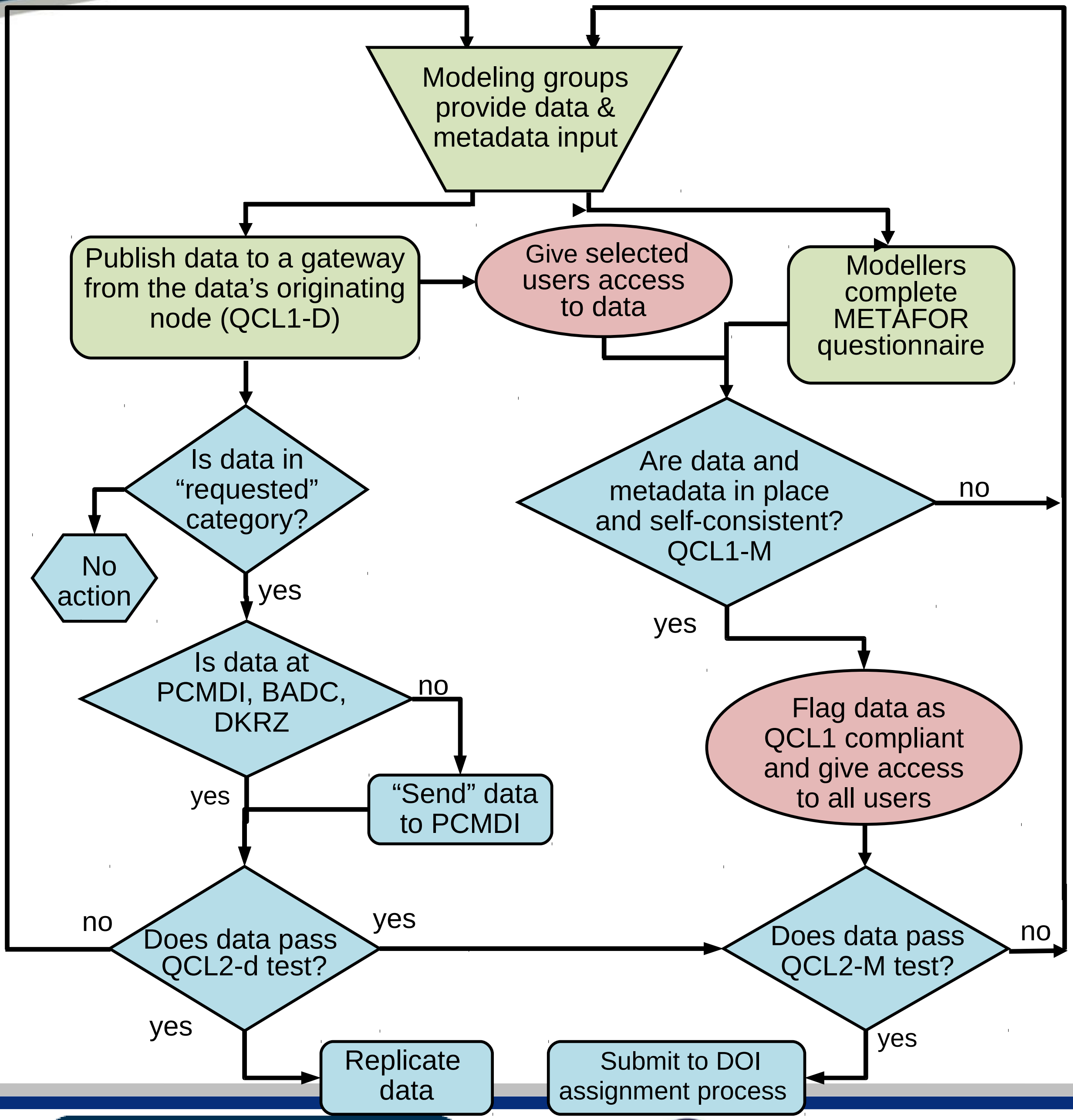
CMIP5 Quality Control as a gatekeeper to global data flow and access:

- fail qc-1d:
  data not published
- fail qc-1m:
  no data access

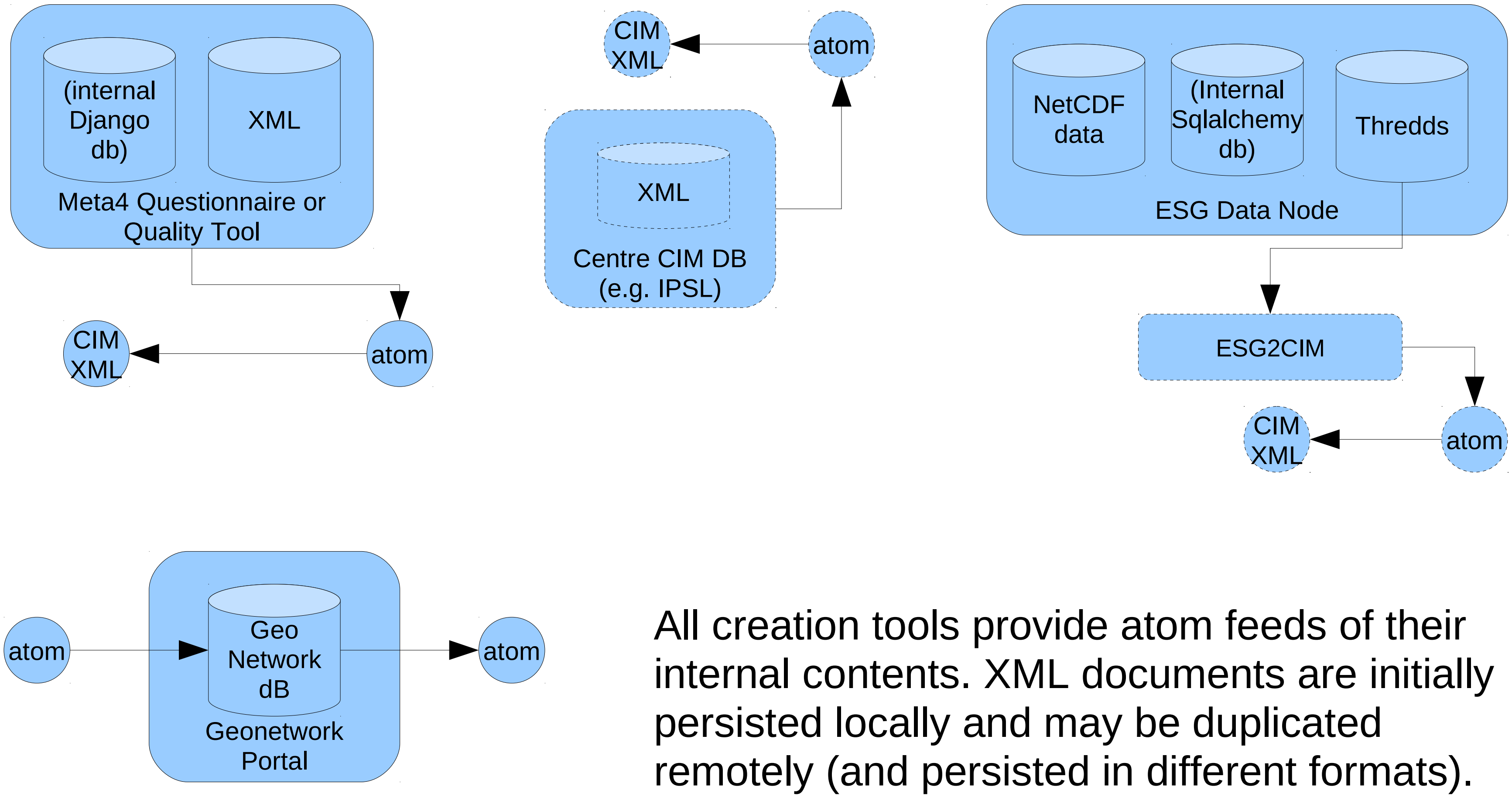


<u>CMIP5 Quality Control</u> as a gatekeeper to global data flow and access:

- 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



CIM content: Creation and Editing



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

Atom Feed

FEED DESCRIPTION

ENTRY DESCRIPTION – points to XML payload


```xml
<?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>
    ...
```

Bringing it all together for CMIP5

The players:

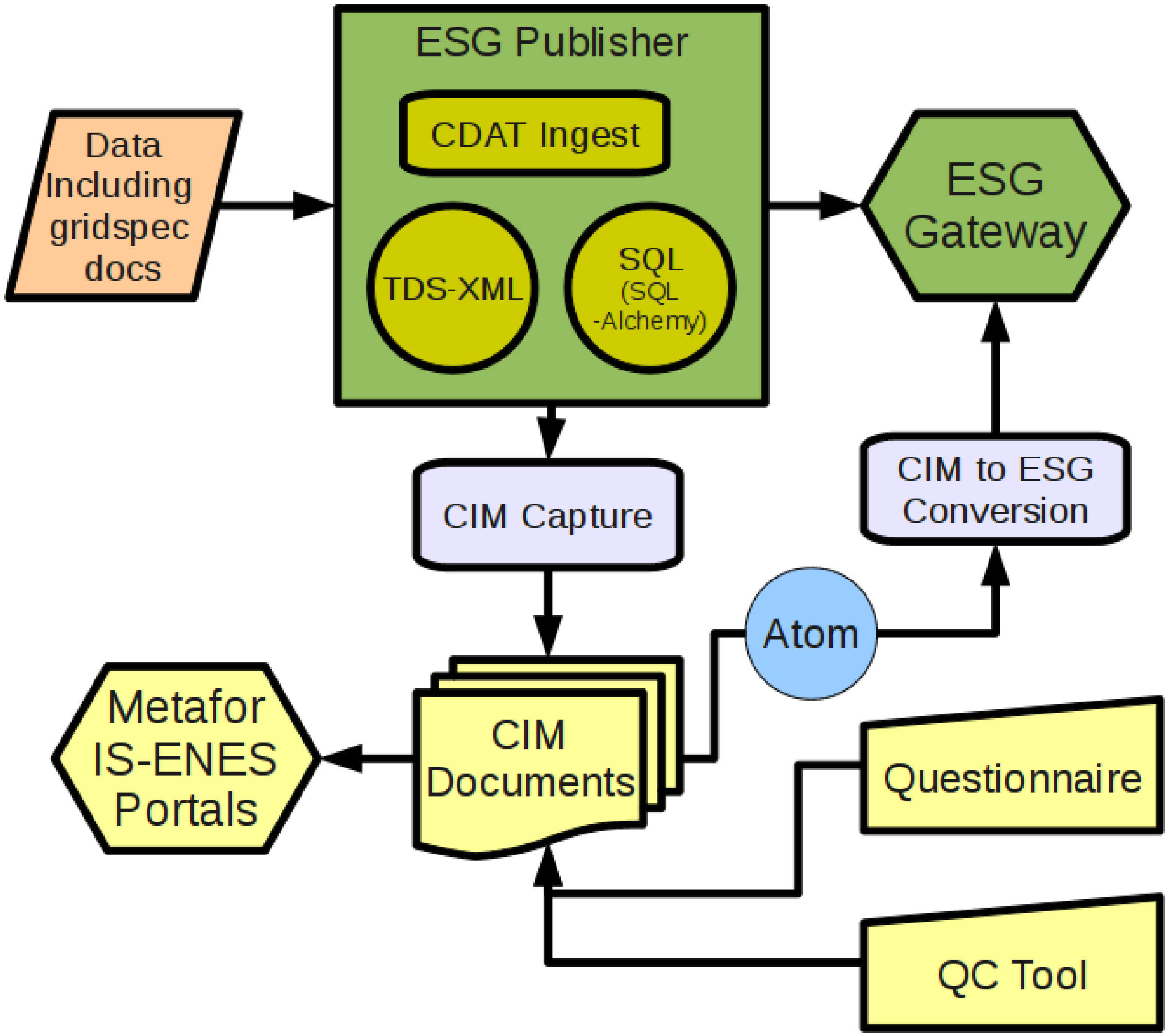
1) NetCDF
CF Conventions +
CMIP5 extensions
(orange)

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

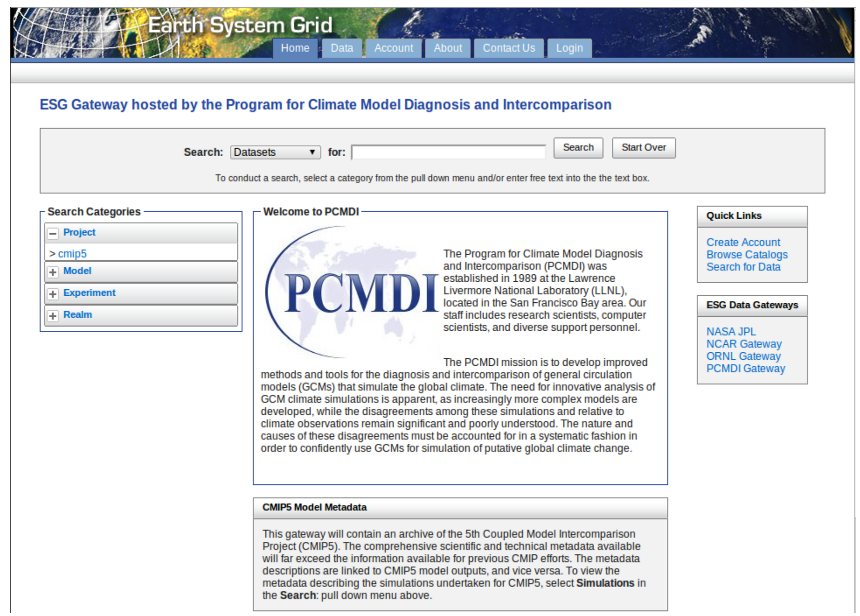
(essentially the gateway harvests the information in their TDS catalog)

and

provide a search interface both to the harvested data,

and to metadata harvested from the metafor questionnaire

NCAR



PCMDI

Science & Technology Facilities Council

ESG Gateways

Earth System Grid

Home | Data | Account | About | Contact Us | Login

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

ESG Gateway hosted by the British Atmospheric Data Centre

Search: Datasets **for:** [Search] [Start Over]

To conduct a search, select a category from the pull down menu and/or enter free text into the the text box.

Search Categories

- **Project**
  - > CMIP3
- **Experiment**
- **Realm**
- **Variable**

Welcome to British Atmospheric Data Centre's ESG gateway

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

This gateway is provided by the BADC on on behalf of the 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.

Quick Links

- Create Account
- Browse Catalogs
- Search for Data

ESG Federation

- PCMDI Gateway
- BADC Gateway
- DKRZ Gateway

- ANU Gateway
- NASA JPL Gateway
- NCAR Gateway
- ORNL Gateway

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Science & Technology Facilities Council

Earth System Grid

Home | Data | Account | About | Contact Us | Login

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Advanced Search

Search: Datasets **for:** [Search] [Start Over]

To conduct a search, select a category from the pull down menu and/or enter free text into the the text box.

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.sresa1b.3hr.atmos
   Authorization: Guest Users

Files Download

*.nc

[Download Files]

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*).

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Science & Technology Facilities Council

Earth System Grid

Home | Data | Account | About | Contact Us | Logout

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Download Data

Sub Select File Results

File Name:

*.nc

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

Sub-Select

Variables:

- [ ] Latent Heat Flux (hfls)
- [ ] Sensible Heat Flux (hfss)
- [ ] Specific Humidity (hus)
- [ ] Accumulated Total Precipitation (pr)
- [ ] Surface Pressure (ps)
- [ ] Mean Sea Level Pressure (psl)
- [ ] Surface Downward Longwave Radiation (rlds)
- [ ] Surface Upward Longwave Radiation (rlus)
- [ ] Top-of-Atmosphere upward Longwave Radiation (rlut)
- [ ] Surface Downward Shortwave Radiation (rsds)
- [ ] Surface Upward Shortwave Radiation (rsus)
- [x] Air Temperature (ta)
- [x] Temperature 2m (tas)
- [x] 2m max temperature (tasmax)
- [x] 2m min temperature (tasmin)
- [ ] Zonal Wind Component (ua)

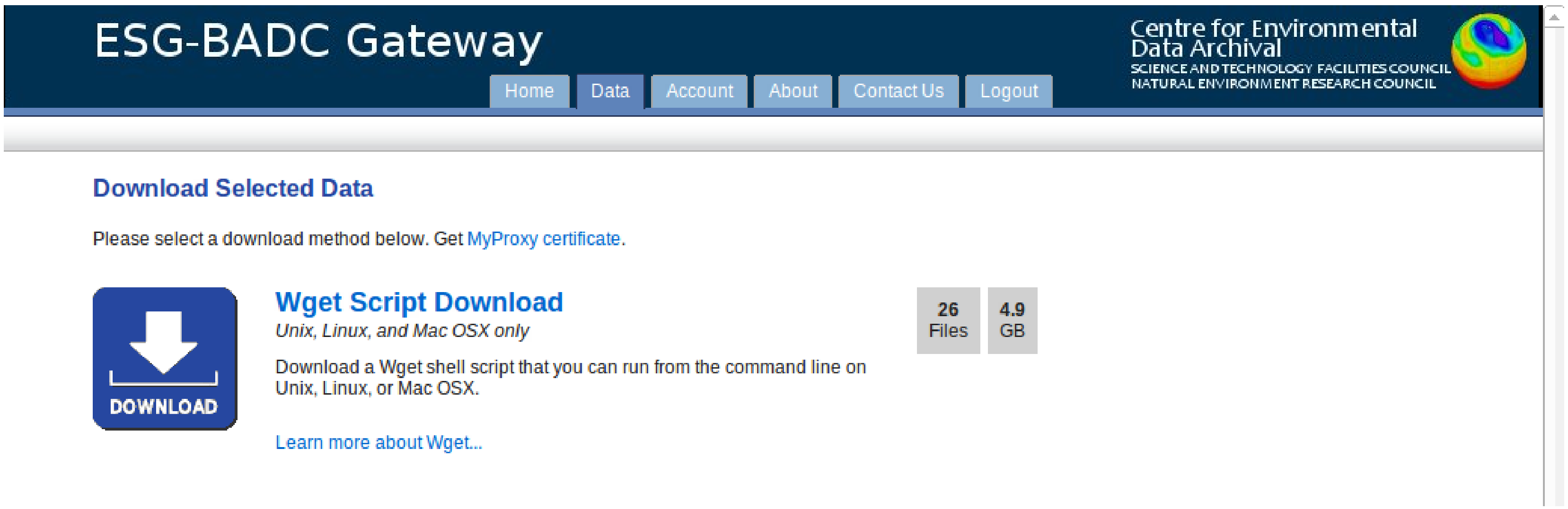
File Download Selection

cmip3_drs.output.BCCR.BCM2.sresa1b.day.atmos
288 File(s)

Download ALL Selected File(s)

☑	File	Size	Format	Location	Direct Download
☑	hfls_A2_BCM2_sresa1b_r1_2056-2065.nc	114.25 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2056-2065.nc	114.25 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2056-2065.nc	114.25 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2056-2065.nc	114.25 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2081-2090.nc	114.19 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2081-2090.nc	114.19 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2081-2090.nc	114.19 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2081-2090.nc	114.19 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2091-2100.nc	102.49 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_r1_2091-2100.nc	102.49 MB	NetCDF	DISK	download
☑	hfls_A2_BCM2_sresa1b_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 certficates

Access Control and Delivery: (2) Direct from TDS on Data Node

THREDDS Data Server

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

Dataset: cmip3_drs.output.UKMO.HADCM3.1pctto2x.mon.seaIce/sic_O1_HADCM3_1pctto2x_r1_1859_De

- *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
- *RestrictAccess:* 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)

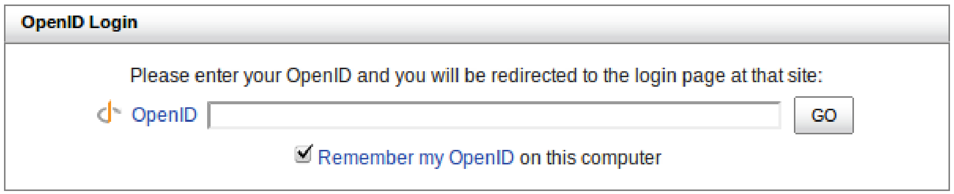
Properties:

- 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 = "e6a90b1eb5291c30c9c40d52f1828cef"
- checksum_type = "MD5"

Access Control and Delivery: (2) Direct from TDS on Data Node

Data Access Login

1

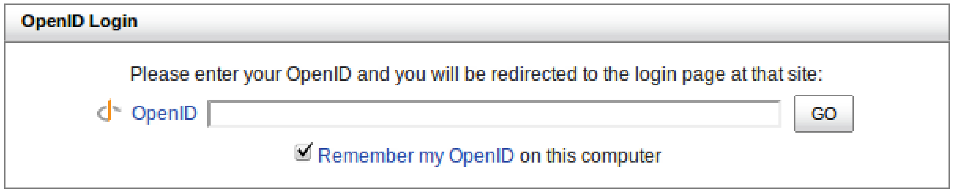


Science & Technology Facilities Council

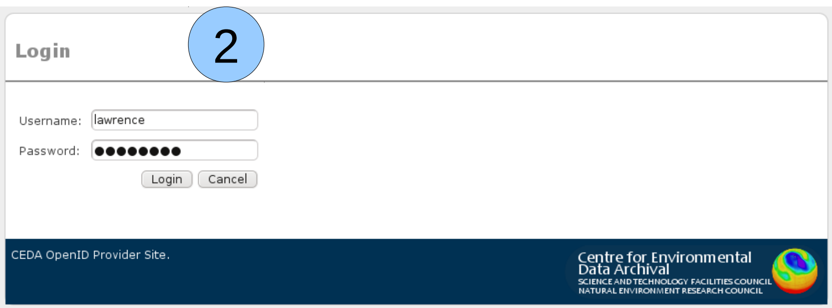
Access Control and Delivery: (2) Direct from TDS on Data Node

Data Access Login

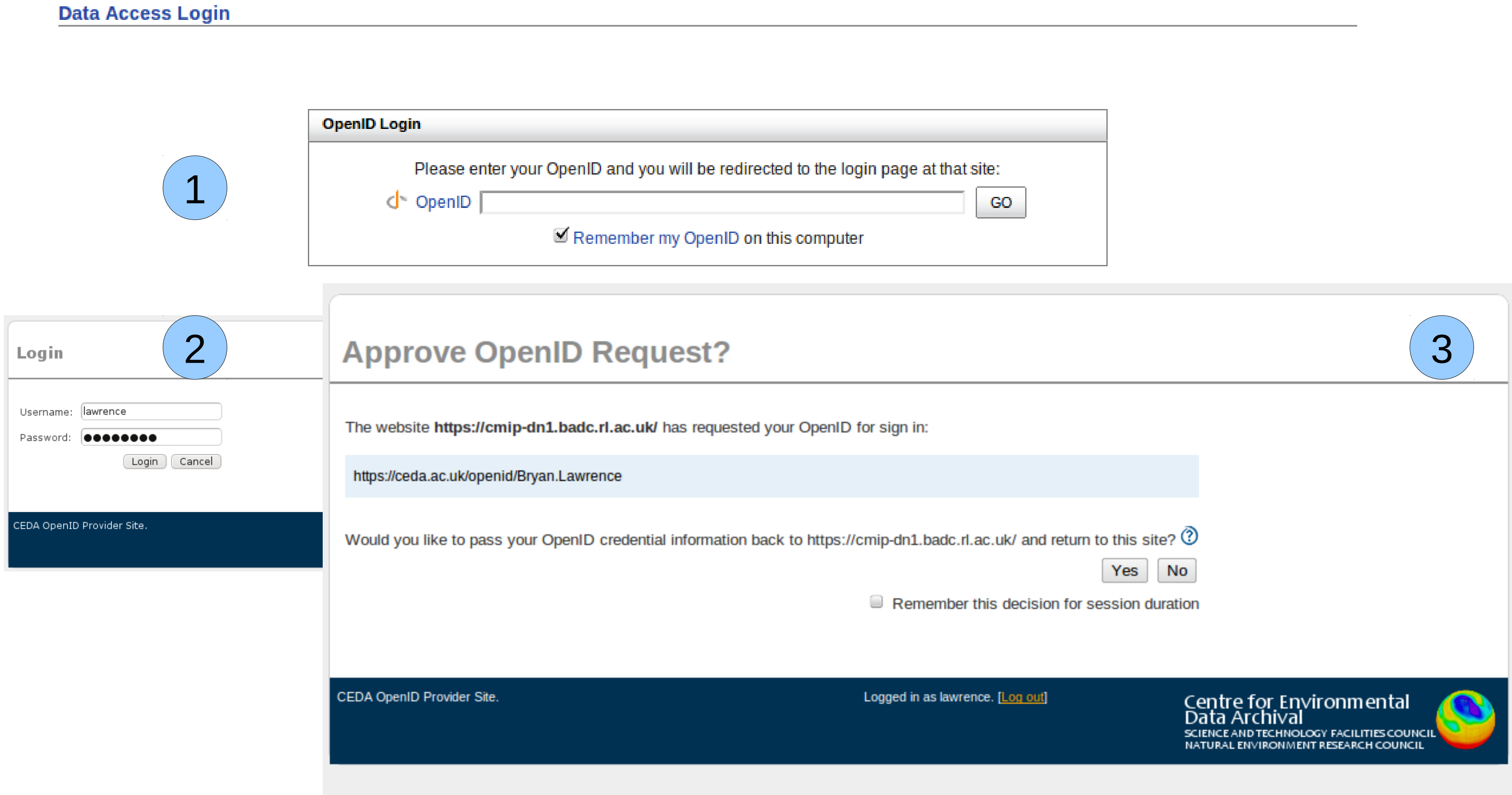
1



2

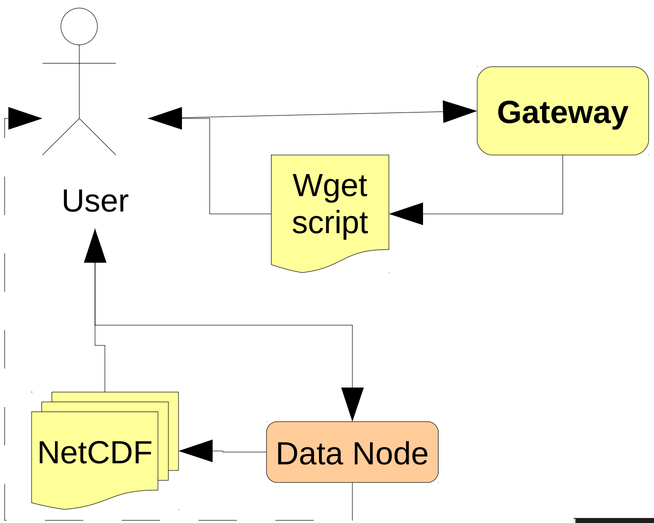


Access Control and Delivery:
(2) Direct from TDS on Data Node



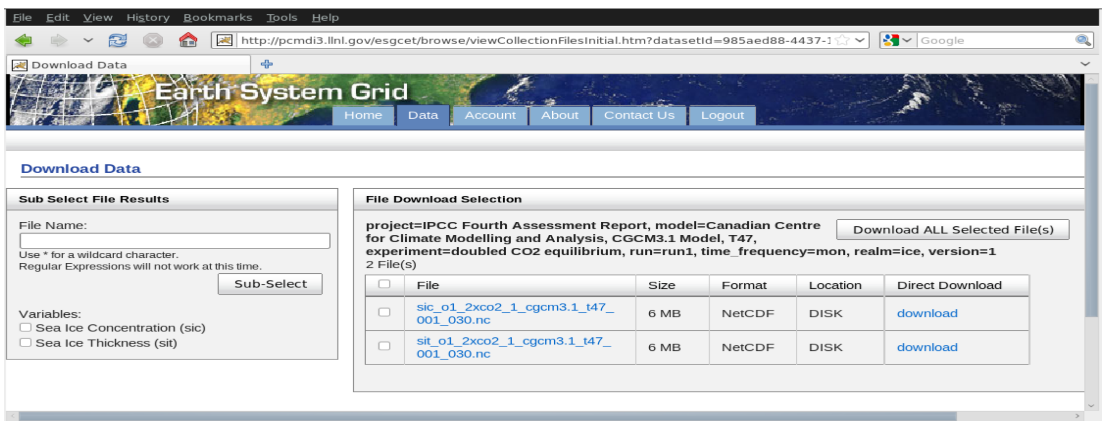
ESG (and CEDA) have comprehensive access control middleware suitable for use in browsers and command line – federated globally!

(Back to the data):User Perspective



New Version Notification

Data nodes will also deploy other tools: secure opendap coming soon (it's done, with modifcations 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)

Peroration

The Climate problem is one that integrates much of e-research, and in particular, the neccessity 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 actitvities.