NCAR

The IPCC AR4: data management and lessons learned

CISL 2005 User Forum

Gary Strand, NCAR/CGD

strandwg@ucar.edu

National Center for Atmospheric Research

Friday, May 15, 2009

What is the IPCC AR4?

NCAR

"The 4th Assessment Report of the Intergovernmental Panel on Climate Change"

The first large-scale coordination of climate modeling efforts, data analysis, data management and data dissemination by the global climate modeling community: 20 global coupled climate models from 13 modeling centers located around the world

A predefined set of 12 experiments

Controls (to assess internal model variability) [2]

CO_2 increase experiments (simplest climate change) [2]

Simulations of 20th century climate [1]

Simulations of future climate based on IPCC "scenarios" [4]

Smaller additional experiments (AMIP, 2XCO2, slab ocean) [3]

Dissemination of data to the climate science community via the Earth System Grid

Unprecedented in scale and scope

The NCAR models

PCM1

- Fully-coupled global climate model
- Atmosphere: “T42” resolution (~310 km), 18 levels
- Ocean: 384x288, 32 levels, curvilinear grid
- Sea ice: 27km polar stereographic grid

34 experiments submitted to the IPCC

CCSM3

- Fully-coupled global climate model
- Atmosphere: “T85” resolution (~155 km), 26 levels
- Ocean: 320x384, 40 levels, curvilinear grid
- Sea ice: 320x384, curvilinear grid

70+ experiments submitted to the IPCC

Current data management

super

data

super's disk

data

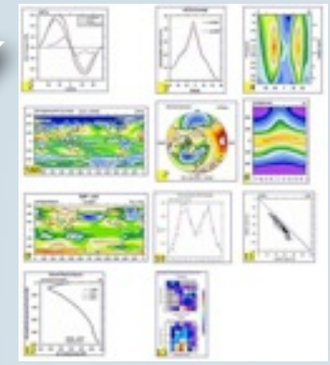
Archival storage

Later... (much later?)

Archival storage

Local disk(s)

Random notes



Diagnostics

Web portal

Model data arrangement

t_0	t_1	t_1	...	t_m
f_1	f_1	f_1		f_1
f_2	f_2	f_2		f_2
...	...	...		...
f_n	f_n	f_n		f_n

IPCC-required arrangement

f_1:	t_0	t_1	t_2	...	t_m
f_2:	t_0	t_1	t_2	...	t_m
...					
f_n:	t_0	t_1	t_2	...	t_m

The NCAR contribution

Site	PCM1	CCSM3	Total
NCAR	11 TB	54 TB	65 TB
ESL	--	33 TB	33 TB
ORNL	3 TB	14 TB	17 TB
NERSC	1 TB	7 TB	8 TB
Total	15 TB	108 TB	123 TB

IPCC requirements

- Specific model fields, unchanged as well as derived
- Data from atmosphere, land surface, ocean and sea ice
- Monthly averages, daily and sub-daily (atm only), annual averages
- Single model field per netCDF file, all time samples
- File sizes must be ~2 GB (as practical)
- Considerable amount of metadata required
- Defined horizontal and vertical resolutions
- Stringent conventions

Tools

NCAR

Hardware

NCAR: “tempest” for more CPU-intensive jobs
“mineral” - CGD-hosted Linux box with 5 TB disk storage
ORNL: “eagle” and “cheetah” - IBM supers
NERSC: “seaborg” - IBM super

Software

Combination of scripts, netCDF operators, NCL and Fortran codes

Verification

Format via PCMDI-supplied checker
Data by scientists

Lessons Learned

Requirements from data producers and data consumers must be unambiguous
Expect software and hardware problems
Devote sufficient resources (people, hardware and especially time)
Employ careful checking
Monitor and track progress

Hardware resources have been inadequate

IPCC AR5 (in ~7 years) **will** force these issues to be resolved

The Earth System Grid

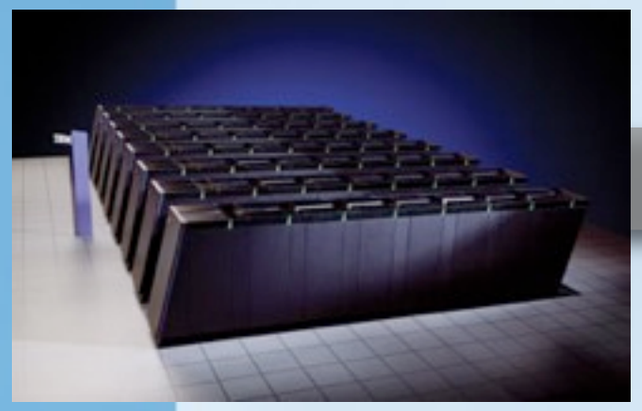
NCAR

- Two complimentary portals, located at PCMDI and NCAR
- PCMDI portal hosting all submissions
  - Approximately 20TB on disk
  - Only IPCC requested subset of data
- NCAR portal hosting only CCSM and PCM data
  - Approximately 20TB published
  - All fields, including IPCC required data
  - Uninterpolated data, for detailed analysis

Hundreds of registered accounts so far

A climate model with good data management

NCAR



Big iron

data

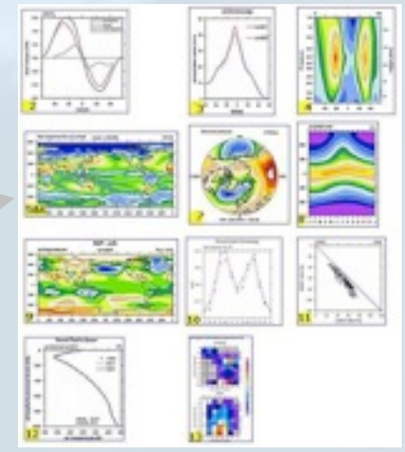
Big disk (50TB+)

data

Data processing engine

Archival storage

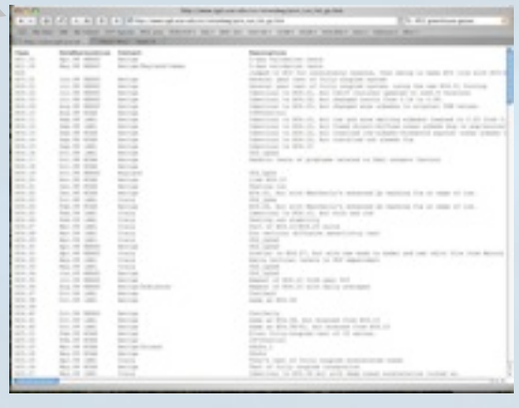
Database



Diagnostics

Local disk

Web portal



Web page(s)

The end (until AR5)