The CORDEX data archive at cordex.dmi.dk

Ole Bøssing Christensen
DKC/DMI

The CORDEX Archive

- A bit of history and inspiration
- How are data going to be organised
- What is unresolved
- What are our ambitions

Our big brother: CMIP5

- The coordinated effort to collect GCM simulations for the upcoming AR5
- Expecting several petabytes (2.3) stored in a de-centralised way. ”ESG”
- Existing VERY detailed specifications, which are the result of MANY man-months. See http://cmip-pcmdi.llnl.gov/cmip5/
- CORDEX specifications are supposed to be as close as possible to CMIP5 wrt. netCDF conventions, attributes, naming…
- We have received a lot of help from CMIP5-related people!

History: PRUDENCE and ENSEMBLES

- **PRUDENCE** (2001-2004):
  - EU FP5 IP, around 3.5 M€ funding.
  - A large amount of impact studies, but centred on RCM modeling of Europe
  - Similar integration areas and a common specification of output variables
  - One canonical driving simulation (HadAM3H/HadCM3) supplemented by voluntary data sets with other drivers
  - Archive has been open to the public since 2004

History: PRUDENCE and ENSEMBLES

- **ENSEMBLES** (2004-2009):
  - EU FP6 IP, around 15M€
  - Very large and heterogeneous project
  - Two ”research themes” focused on RCM modelling for Europe plus West Africa.
  - RCM modeling:
    - Reanalysis-based plus GCM-driven simulations; ”filling the matrix”
    - IDENTICAL integration areas (mostly…)

ENSEMBLES GCM-RCM Matrix

Global model Regional inst.	METO-HC Standard	METO-HC Low sens.	METO-HC Hi sens.	MPIMET Standard	MPIMET Ens.m. 1	MPIMET Ens.m. 2	IPSL	CNRM	NERSC	MIROC	CGCM3	Total number
METO-HC	**2100**	**2100***	**2100***	2100								4
MPIMET				**2100**			2050*					2
CNRM								**2100**				1
DMI				**2100***				**2100**	**2100***			3
ETH	**2100**											1
KNMI				**2100*** **2100**	**2100***	**2100***				**2100***		1+4
ICTP				**2100**								1
SMHI		**2100***		**2100*** **2100***					**2100**			3+1
UCLM	**2050**											1
C4I			**2100***		**2050 (A2)***							2
GKSS							**2050***					1
METNO	**2050***								**2050***			1
CHMI								2050*				1
OURANOS**											**2050***	1
VMGO**	**2050***											1
Total (1951-2050)	5	2	2	6+2	1+1	0+1	2	3	3	0+1	1	25+5

Red: Online now; *: non-contractual runs; **:affiliated partners without obligations; underscore: 50km resolution; (in parantheses): Expected date. For partner acronym explanations, see the participant list. **NOTE** that all partners also did an ERA-40 driven analysis 1951(1961)-2000

Common ENSEMBLES area

Internal zone:
Models with rotated lat/lon:
Common area:
[lon1, lon2, lat1, lat2, polon, polat] = [-21.72, 15.46, -20.68, 20.90, 18.00, -39.25]
nlon x nlat = 170 x 190

Previous archives

- **PRUDENCE** EU FP5 2001-2004 http://prudence.dmi.dk
  - Europe in around 50km resolution
  - SRES A2 and B2 scenarios
  - One GCM plus a few simulations with others
  - Open to the public since 2004
  - 30y time slices for current and end-of-century conditions
  - 21 surface fields in daily, monthly and seasonal resolution
  - Around 56 time slices of 30 years
  - Data received via DVD, tape, download, disks
  - Accessible through direct download or OpenDaP/DODS
  - Totally around 500GB archive size
- **ENSEMBLES** EU FP6 2004-2009 http://ensemblesrt3.dmi.dk
  - Europe in around 25km resolution
  - SRES A1B scenario
  - Several GCM/RCM combinations
  - Open since 2008 following the definition of an ENSEMBLES data policy
  - Wish-list of around 130 daily and sub-daily fields. Not all fields on the output list are present for all simulations
  - Transient simulations 1951-2050 or 1951-2100.
  - 25 expts in 25km resolution of 100-150 years
  - Data received mostly on USB disks, also download e.g. from ECMWF
  - Email addresses are asked for and saved, but no real security involved
  - Piggyback on the DMI 100MBit/s network connection, and backup in the DMI mass storage system
  - Accessible through direct download or OpenDaP/DODS
  - Totally around 23TB archive size

Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects

Front page

Public part

News
Participants
Project summary
Description of work (.PDF)
Reports and publications

Data distribution front page

Direct download
DODS

Maintained by
Ole Bøssing Christensen

Overview of the PRUDENCE experiments

Updated 3/2/2005.

Institute/ Contact	Model	Driving data	Ens.	Experiment	Acronym	Seasonal data	Monthly data	Daily data
DMI Jens H. Christensen	HIRHAM	HadAM3H A2	1	Control	HC1	Online	Online	Online
			1	Scenario	HS1	Online	Online	Online
			2	Control	HC2	Online	Online	Online
			2	Scenario	HS2	Online	Online	Online
			3	Control	HC3	Online	Online	Online
			3	Scenario	HS3	Online	Online	Online
		HadAM3H A2 (SMHI Baltic SSTs)	1	Scenario	HS4	Online	Online	Online
		HadAM3H B2		Scenario	HB1	Online	Online	Online
		ECHAM4/OPYC (OAGCM SSTs)		Control	ecctrl	Online	Online	Online
		ECHAM4/OPYC A2 *Not the same expt. as the SMHI MPIA2!*		Scenario	ecscA2	Online	Online	Online
		ECHAM4/OPYC B2		Scenario	ecscB2	Online	Online	Online
		ECHAM5 A2		Control	ECC	Online	Online	Online
				Scenario	ECS	Online	Online	Online
	HIRHAM High res.	HadAM3H A2	1	Control	F25	Online	Online	Online
			1	Scenario	S25	Online	Online	Online
	HIRHAM Extra High res.	HadAM3H A2	1	Control	F12	Online	Online	Online
			1	Scenario	S12	Online	Online	Online

Public part

RCM data portal

Front page

RT3 participant list

Older news

List of output variables

The GCM/RCM combination matrix

The AMMA-region matrix

Fields in the ERA40 archive

The integration area common to most simulations

Plots from the quick-look analysis

The PRUDENCE project (our predecessor)

The CORDEX project (our successor)

Members' part

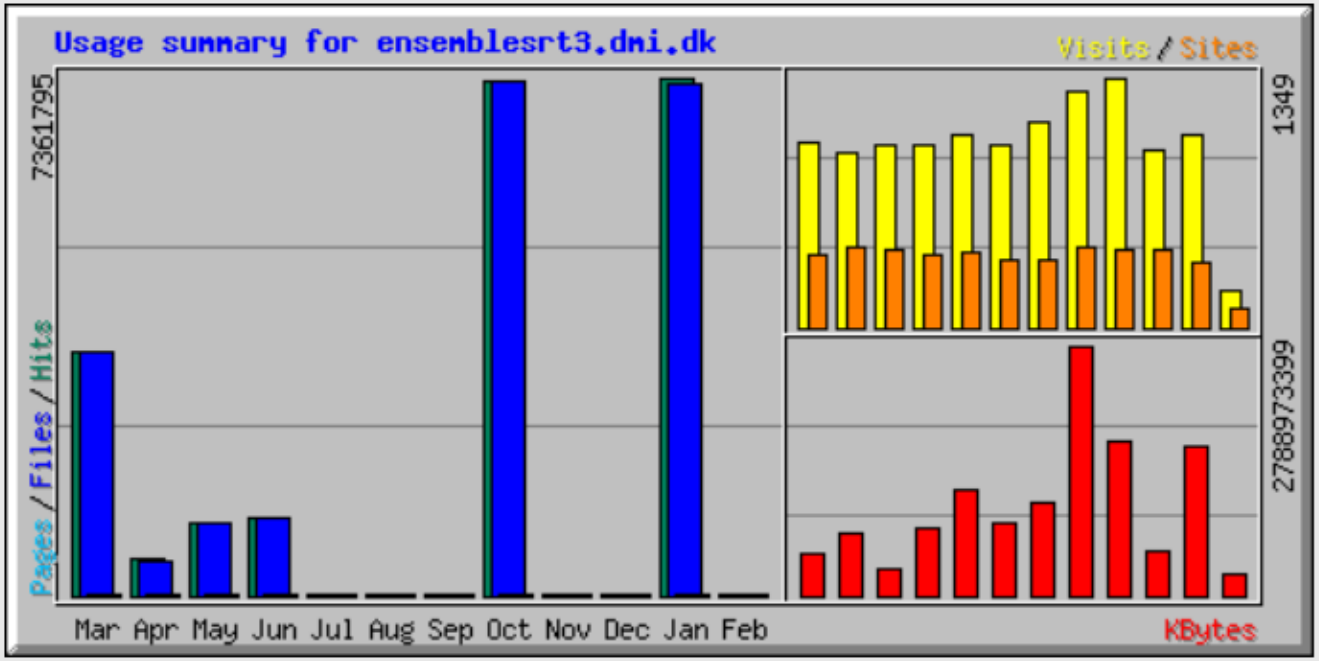
Ensembles RT3 mailing list

Plots from C4I's validation against HOAPS

Daniela Jacob							

RT2B: Transient experiments 1951-2050 or 1951-2100 driven by global experiments according to this plan

Institute/ Contact	Scenario	Driving GCM	Model	Resolution	Acronym	DODS/OpenDAP access	Direct download
C4I Ray McGrath	A2	ECHAM5	RCA3	25km	C4IRCA3	Online	Online
CNRM Michel Déqué	A1B	ARPEGE	Aladin	25km	CNRM-RM4.5	Online	Online
	A1B	ARPEGE_RM5.1 New ens.mb. to 2100	Aladin	25km	CNRM-RM5.1	Online	Online
KNMI Erik van Meijgaard	A1B	ECHAM5-r3	RACMO	25km	KNMI-RACMO2	Online	Online
	A1B	ECHAM5-r1	RACMO	50km	KNMI-RACMO2	Online	Online
	A1B	ECHAM5-r2	RACMO	50km	KNMI-RACMO2	Online	Online
	A1B	ECHAM5-r3	RACMO	50km	KNMI-RACMO2	Online	Online
	A1B	MIROC	RACMO	50km	KNMI-RACMO2	Online	Online
OURANOS Dominique Paquin	A1B	CGCM3	CRCM	25km	OURANOSMRCC4.2.1	Online	Online
SMHI Erik Kjellström	A1B	ECHAM5-r3	RCA	50km	SMHIRCA	Online	Online
	A1B	BCM	RCA	25km	SMHIRCA	Online	Online
	A1B	ECHAM5-r3	RCA	25km	SMHIRCA	Online	Online
	A1B	HadCM3Q3	RCA	25km	SMHIRCA	Online	Online
MPI Daniela Jacob	A1B	ECHAM5-r3	REMO	25km	MPI-M-REMO	Online	Online

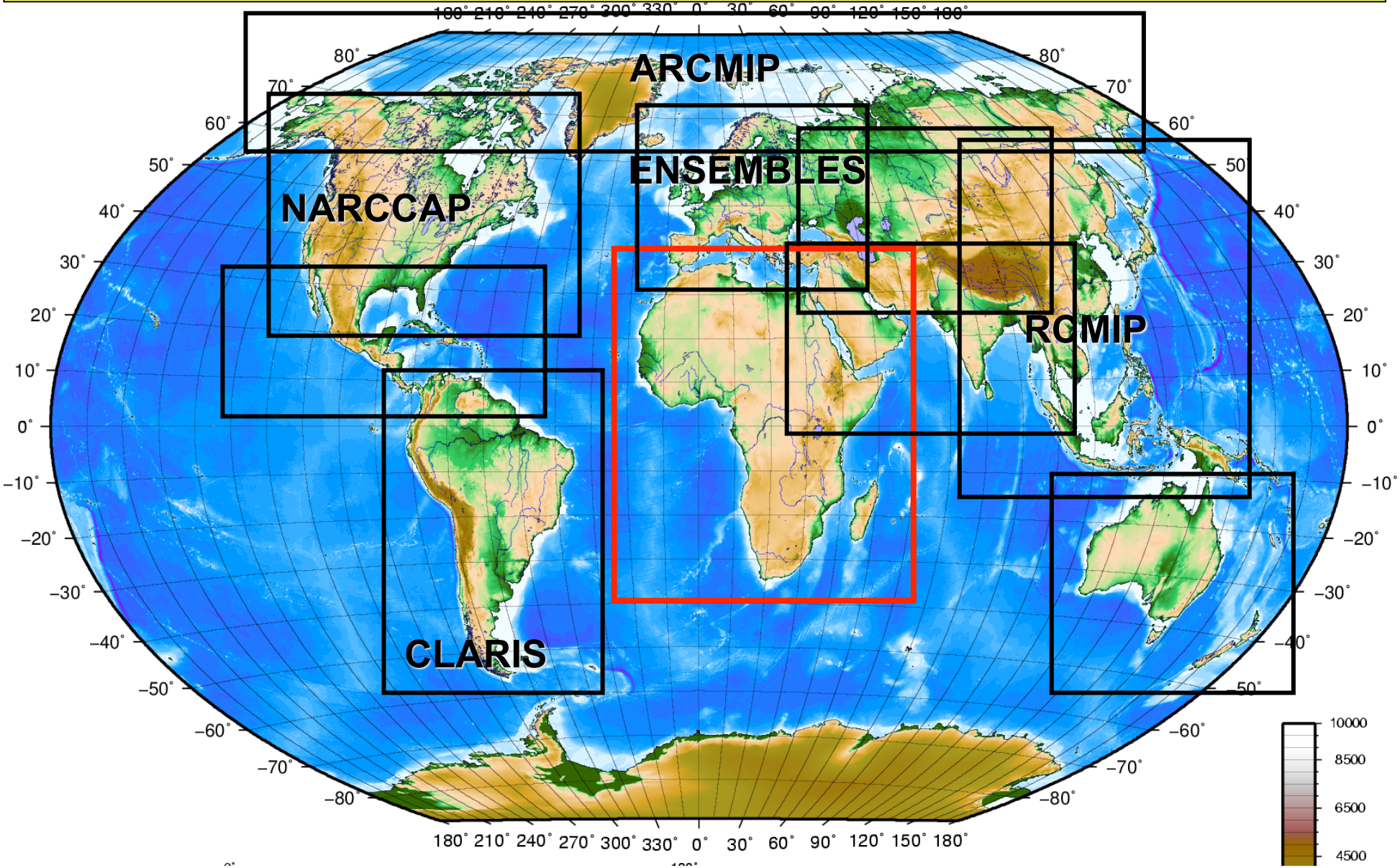


Summary by Month

Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Feb 2009	1350	919	395	48	98	234332345	194	1582	3676	5400
Jan 2009	237477	233992	245	33	345	1668234809	1045	7625	7253775	7361795
Dec 2008	579	398	228	30	413	508824151	954	7086	12346	17956
Nov 2008	1128	837	431	44	423	1728716595	1349	12932	25127	33856
Oct 2008	236179	235234	364	41	435	2788973399	1274	11310		
Sep 2008	1018	494	460	36	359	1037260176	1107	13824		
Aug 2008	768	636	206	31	371	823956173	982	6389		
Jul 2008	602	462	228	33	402	1190587004	1033	7083		
Jun 2008	36796	36652	276	32	386	749414979	988	8284		

Our record then:
3TB/month
Actual peak:
30TB in one month

CORDEX domains



CORDEX expected data volume

- African domain: Around 40.000 points. All 0.44-deg. domains: around 250.000 points
- Simulation length:
  - Reanalysis-driven: 56y
  - One historical 1951-2004 plus two RCP scenarios 2005-2100, 246y
  - Possibly AMIP: 40y
  - Near-future 3x10y=30y
- Number of fields:
  - Core: 42 monthly, some seasonal, corresponding to roughly 1.5 field per day
  - Tier 1: 57 daily fields.
  - Tier 2: 15 3-hourly plus 35 6-hourly, or 260 fields per day
- Number of simulation sets:
  - Say, 10 institutions doing 3 GCMs each, or 30 simulations
- Also, a high-res European area, 160.000 points, 10 simulations…
- Total: 4 bytes/number x 250.000 points x 372y x 365.25 days/y x (1.5,57,260) fields x 30 GCM/RCM combinations, or:
  - Core: 5.6 TB + 1.2 TB for EUR-11
  - Tier 1: 211 TB + 45 TB for EUR-11
  - Tier 2: 964 TB + 206 TB for EUR-11 (total, distributed)

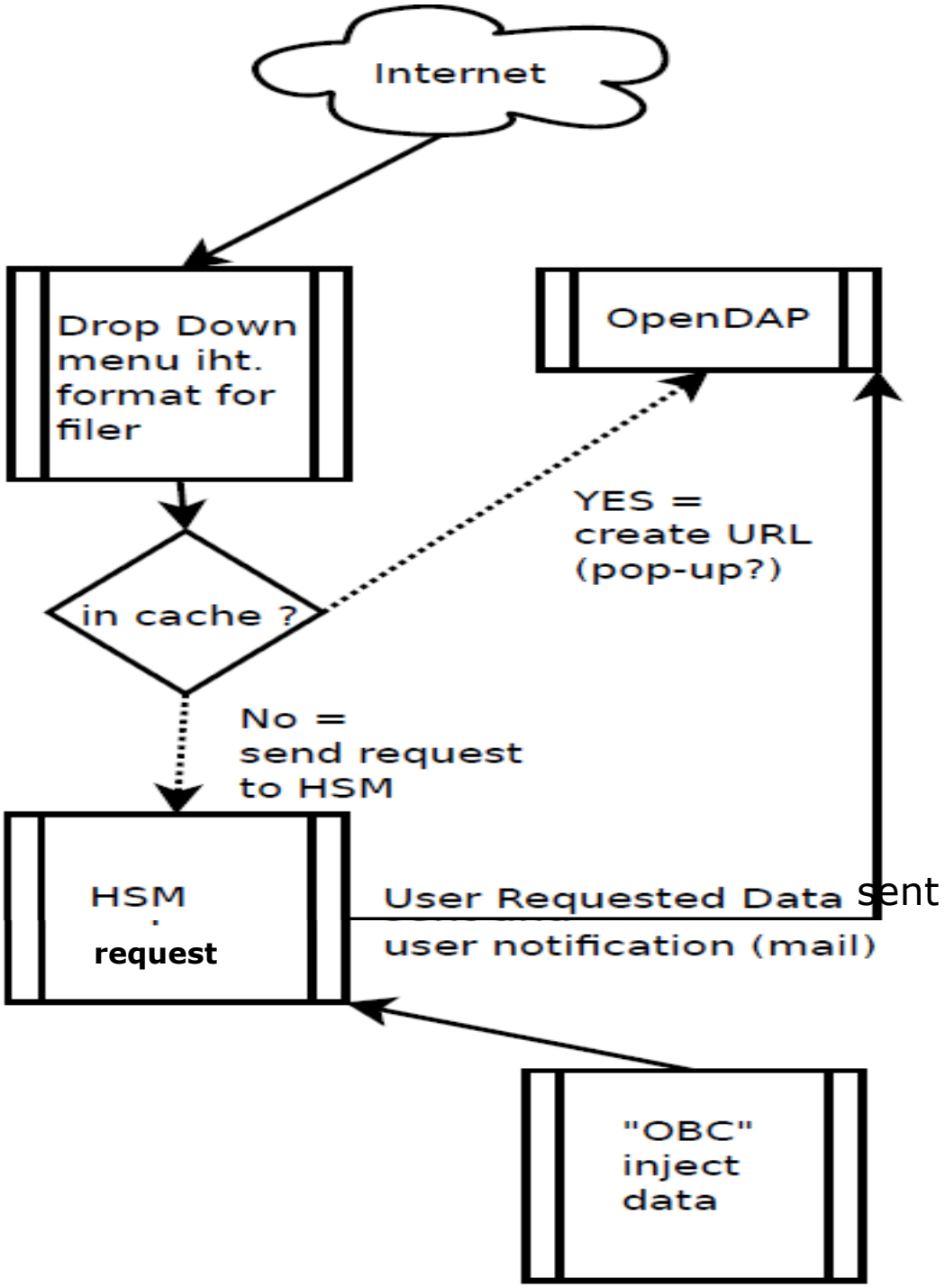
Is this a lot of data??

- Data amounts
  - PRUDENCE 0.5TB
  - ENSEMBLES 23TB
  - CORDEX Core+Tier 1, roughly 263TB
  - All CORDEX data 1.400TB
  - CMIP5 2.3 PB (Dr. Asrar yesterday) 2.300TB
- So: Still an order of magnitude less than CMIP5, but a full order of magnitude more than ENSEMBLES
- An A4 sheet of paper holds 2kB of printed characters (bytes). Printer paper is roughly 0.1mm thick, *i.e.*, 20MB/m.
  - Prudence printed is a stack 25km high
  - ENSEMBLES is 1.100 km
  - CORDEX Core+Tier 1 may be 13.000 km
  - CORDEX total around 70.000 km
  - CMIP5 : 115.000 km
- Seriously: Storing is cheap, but bandwidth is limited. Time to download at 100MBits:
  - PRUDENCE: 12 hours
  - ENSEMBLES: 23 days
  - *etc.*

How do we plan to do it

- Low budget and many data. Uncertain total amount of data
  - Therefore we will use the existing DMI tape archive plus a cache of about 15TB.
  - Users specify a wish list of files; some may be in cache, others on disk.
  - If on disk, accessible through download or OpenDaP
  - If not, the user receives a mail when the files have been retrieved
- Data delivery: We can download and we can accept external disks in the mail. No incoming area planned.
- Note: Total bandwidth out of DMI could easily be the bottleneck, at most 1TB/day (100Mbps)
- We are investigating how to implement ESG software in this non-standard case.
  - This would also facilitate SHARED data delivery


DMI


Internet
Drop Down menu iht. format for filer
OpenDAP
YES = create URL (pop-up?)
in cache ?
No = send request to HSM
HSM
request
User Requested Data sent
user notification (mail)
"OBC" inject data

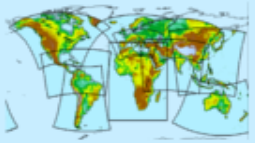
Status as of today

- The page http://cordex.dmi.dk/ exists, with specifications posted
- A server has been acquired and installed
- WE ARE READY FOR DATA
- No data yet
- Plans to set up a file checker program, offered by DKRZ
  - Can be run at DMI with sample data
  - Can be implemented and run at home
- Currently data access in the old, proven way
- Talks with CMIP5 people about possible implementation of ESG software. Not that easy: ESG currently requires accessible data (on disk).
- Talks with IS-ENES and the Swedish NSC about data sharing
- There is a need for more user-friendliness than the predecessors. Hopefully ESG can help.
  - Pressing need for texts, information and guidance
  - For documentation
  - For visualisation tools
  - We cannot do this at the DMI

http://cordex.dmi.dk/

CORDEX climate data archive

Home



Main Menu

- Home
- Specifications
- All news
- Relevant Links

Login Form

Username

Password

Remember Me

Login

- Forgot your password?
- Forgot your username?
- Create an account

CORDEX: A COordinated Regional climate Downscaling EXperiment

First status report from Trieste

Written by Ole Bøssing Christensen

Monday, 21 March 2011 09:49

Dear colleagues,

The first important message from the CORDEX meeting in Trieste this week is the following: It seems that modellers have been interpreting the original area specification document different from me. So, in the original document wrt. rotated lat/lon areas, the "corner" coordinates are centre coordinates of the corner point. In other words, the definition in my table in the specifications is one half grid point off. It will obviously be corrected asap. Hopefully all of you have been taking area information from the original document at IPSL and interpreting them as everybody else...

By the way: I have now added a fixed link on the left to facilitate retrieval of the most recent final specification document.

Last Updated on Monday, 21 March 2011 19:55

Doubt about domain specs

Written by Ole Bøssing Christensen

Wednesday, 16 March 2011 13:55

Dear all,

I just heard that there may be an error in the integration domain specifications in the original document, and therefore also in the specification table located here, which was based on that document! Probably, the definitions in the area document were NOT corners, as specified in the document, but centers of the corner points. This makes differences of 0.22 degrees most of the time. We will clear this up in Trieste next week, but you better wait a while.

Another error: I claimed that I had corrected the link to IPSL. Apparently I had not. Now it is finally fixed. Sorry!

Archive specifications

Written by Ole Bøssing Christensen

Monday, 17 January 2011 08:38

Further changes

Written by Ole Bøssing Christensen

Tuesday, 15 March 2011 11:42

Dear colleagues,

After a meeting with several people knowledgeable about CMIP5, we have made further harmonisations and clarifications. Here is the document. You should, however, be aware of the fact that there are two very relevant meetings next week, one CORDEX conference in Trieste, plus an IS-ENES meeting in Paris. Particularly the CORDEX meeting could very well lead to further changes in the document. This is annoying for everyone, but it is probably even more annoying if necessary changes are postponed until later.

Best, Ole

Last Updated on Tuesday, 15 March 2011 13:02

Starting up the CORDEX archive

Specifications

- MANY thanks to Stephanie Legutke, Grigory Nikulin, Martin Juckes, and several others!!!
- Please conform to specs for
  - File names
  - Compulsory netCDF attributes
  - netCDF conventions (CF1.6), including variable-description attributes, projection information, units, names *etc*.
- We do require core indices. Not all Tier-1 data possible (but try!). And remember that *you* must store Tier 2 *in format*!
- Many ”final” specifications have already been posted; many inconsistencies and errors have been corrected. Hopefully we are almost there now

Specifications

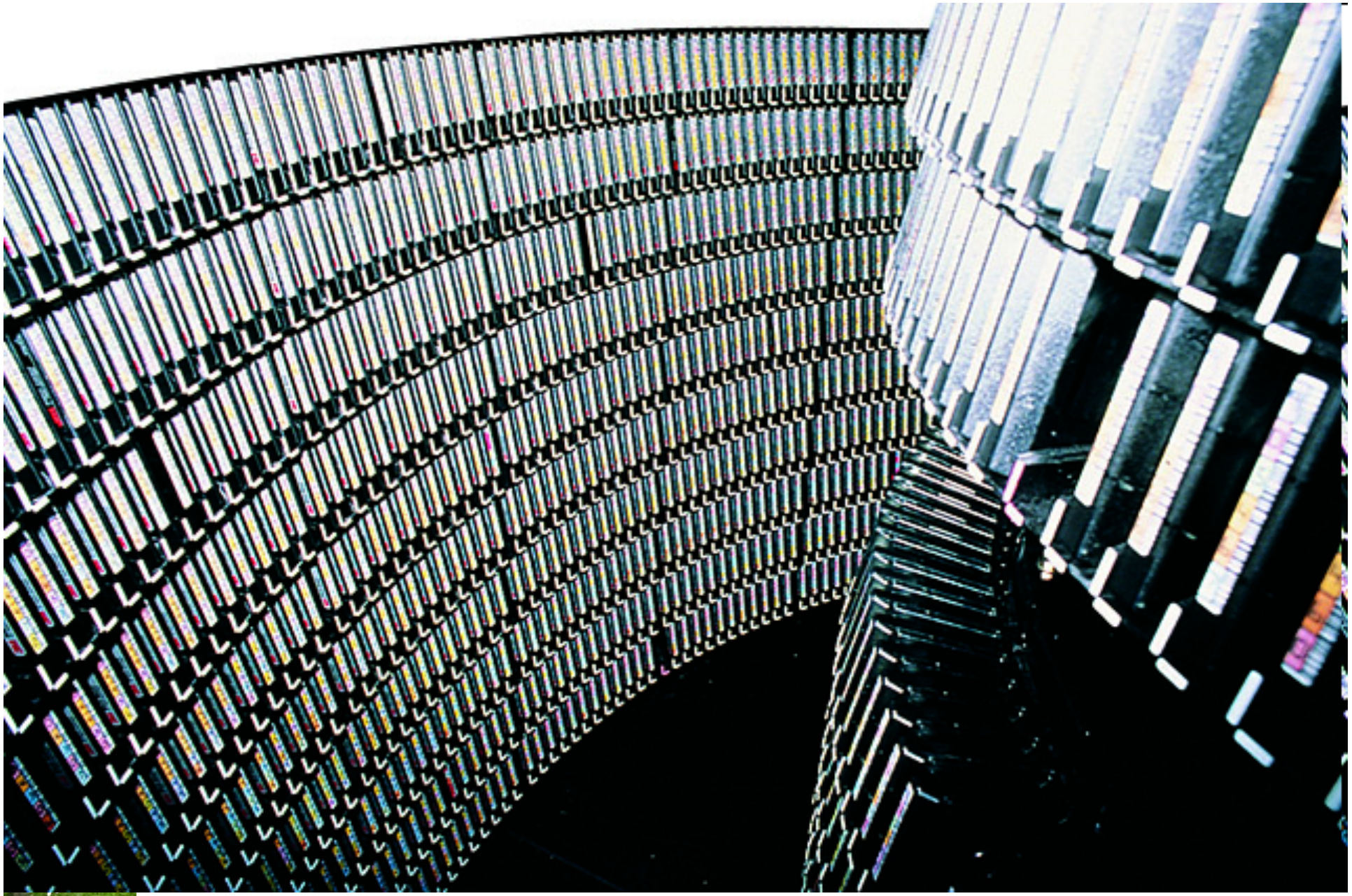
- Current urgent issues:
  - Are the areas correctly defined?
  - Any attributes missing?
  - Any ambiguous variable specifications
  - Are we happy about periods in file names (instantaneous vs. averaged files…)
  - Has anybody noticed anything missing?

DMI

CORDEX

How can you help?

- Data delivery: Possibly less experienced users, so PLEASE stick to specifications to facilitate useability
- Contributions to web site: We need
  - Tools to handle and preview data; hopefully CMIP5 can help
  - Documents about the data, dos and don’ts (recall the discussion yesterday)
  - Logo, suggestions about organisation… Cannot be my personal blog like ensemblesrt3, due to the expected large audience



(NCAR 1986: 1.6 TB!!)

Organisation

- Debian stable and apache 2
- Gzip'ed netCDF files following the CF1.0 convention for
  - Direct download
  - OpenDaP access
    - Sub-windows in space and sub-periods in time
    - Strides
    - Direct integration in "OpenDaP-aware" applications
- Daily, sub-daily and monthly-mean files on native grid
- Common regular interpolated grid for monthly means as well
- Not very user friendly! No online previews or analyses (yet)