TECHNICAL MEMORANDUM

466

Report on the seventeenth meeting of Computing Representatives 19–20 May 2005

P. Prior (Compiler)

Operations Department

December 2005

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Contents

Part I: ECMWF Staff contributions

Part II: Member State and Co-operating State contributions

Preface

The seventeenth meeting of Computing Representatives took place on 19–20 May 2005 at ECMWF. Twenty two Member States and Co-operating States, plus the CTBTO, were represented. The list of attendees is given in annex 1.

The Head of the Computer Division (Isabella Weger) opened the meeting and welcomed representatives. She gave a presentation on the current status of ECMWF's computer service and plans for its development. Each Computing Representative then gave a short presentation on their service and the use their staff make of ECMWF's computer facilities. Participants were also invited to report on their Disaster Recovery Systems, if any, and experience with tape libraries. There were also presentations from ECMWF staff members on various specific developments in the ECMWF systems. The full programme is given in Annex 2.

This report summarises each presentation. Part I contains ECMWF's contributions and general discussions. Part II contains Member States' and Co-operating States' contributions; all the reports were provided by the representatives themselves.

Part I

ECMWF Staff contributions and general discussions

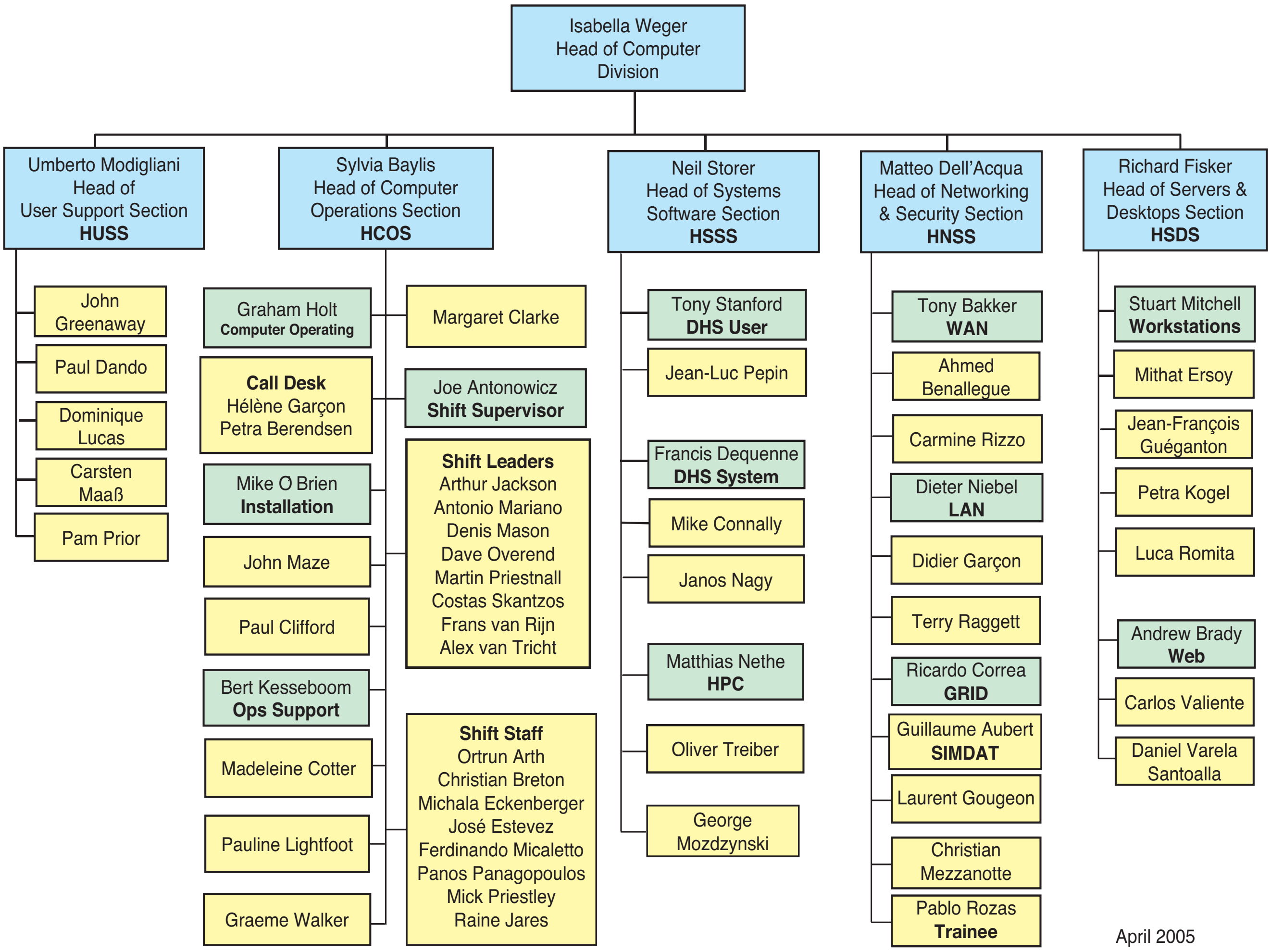
ECMWF Computing Service: Status and Plans — *Isabella Weger, Head of Computer Division*

Major activities over the past 12 months

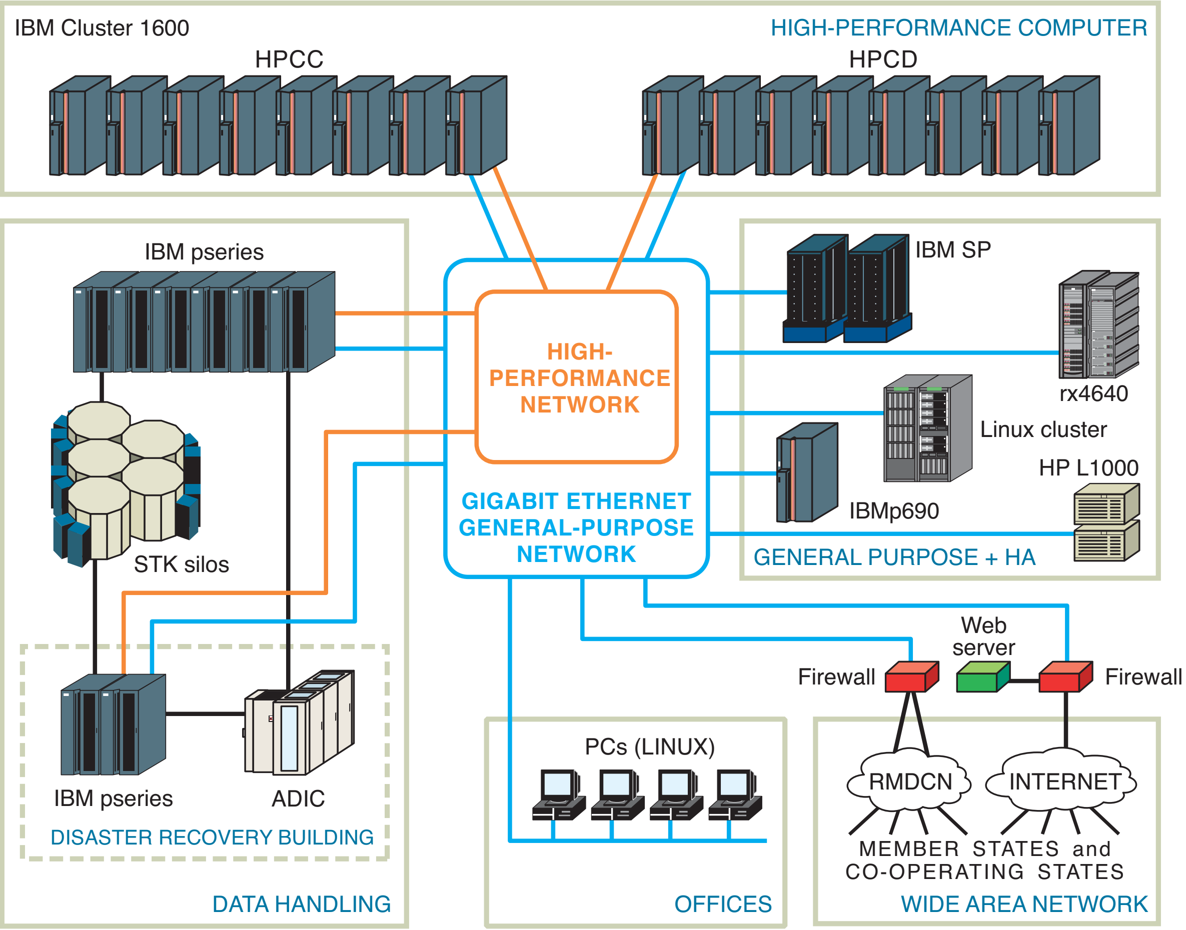
- The migration from Phase 1 to Phase 3 of the IBM HPCF was completed in November 2004.
- Phase 3 of the IBM HPCF continues to provide an excellent service at a high level of availability, although we are experiencing a higher level of Multi-Chip Module failures than other sites. This is under investigation.
- More improvements were made to job scheduling on the IBM HPCF, not only to take into account the increase in the number of CPUs per node (from 8 in Phase 1 to 32 in Phase 3) but also to allow the reservation of nodes for the forecast suite while maximising system utilisation (running Member States' workload on the same cluster as the Operational Suite).
- The migration from ecgate1 (SGI Origin) to the new IBM server ecgate was successfully completed in September 2004.

  It is providing a very stable service to Member State and Co-operating State users.
- A Gaseous Fire Suppression System was installed in the main computer hall and tape library.
- A third Uninterruptible Power Supply machine was installed.
- The Computer Building extension was started and is expected to be completed in the summer.
- A survey of external users with interactive access to the ECMWF computing facilities was conducted in February 2005

Computer Division Organigram

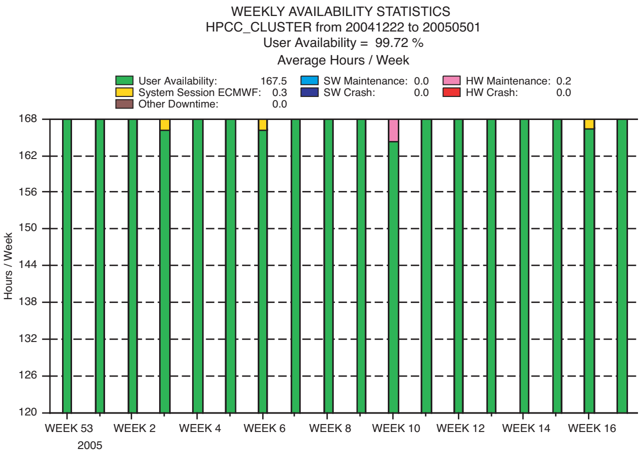

Isabella Weger Head of Computer Division
Umberto Modigliani Head of User Support Section HUSS
John Greenaway
Paul Dando
Dominique Lucas
Carsten Maaß
Pam Prior
Sylvia Baylis Head of Computer Operations Section HCOS
Graham Holt Computer Operating
Call Desk Hélène Garçon Petra Berendsen
Mike O Brien Installation
John Maze
Paul Clifford
Bert Kesseboom Ops Support
Madeleine Cotter
Pauline Lightfoot
Graeme Walker
Margaret Clarke
Joe Antonowicz Shift Supervisor
Shift Leaders Arthur Jackson Antonio Mariano Denis Mason Dave Overend Martin Priestnall Costas Skantzos Frans van Rijn Alex van Tricht
Shift Staff Ortrun Arth Christian Breton Michala Eckenberger José Estevez Ferdinando Micaletto Panos Panagopoulos Mick Priestley Raine Jares
Neil Storer Head of Systems Software Section HSSS
Tony Stanford DHS User
Jean-Luc Pepin
Francis Dequenne DHS System
Mike Connally
Janos Nagy
Matthias Nethe HPC
Oliver Treiber
George Mozdzynski
Matteo Dell'Acqua Head of Networking & Security Section HNSS
Tony Bakker WAN
Ahmed Benallegue
Carmine Rizzo
Dieter Niebel LAN
Didier Garçon
Terry Raggett
Ricardo Correa GRID
Guillaume Aubert SIMDAT
Laurent Gougeon
Christian Mezzanotte
Pablo Rozas Trainee
Richard Fisker Head of Servers & Desktops Section HSDS
Stuart Mitchell Workstations
Mithat Ersoy
Jean-François Guéganton
Petra Kogel
Luca Romita
Andrew Brady Web
Carlos Valiente
Daniel Varela Santoalla
April 2005

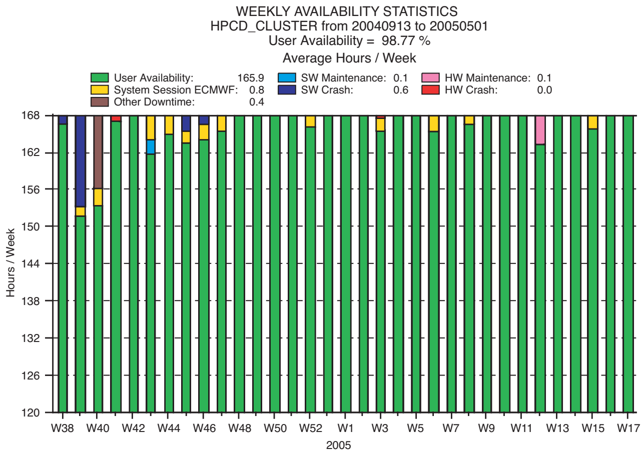
ECMWF Computer Environment

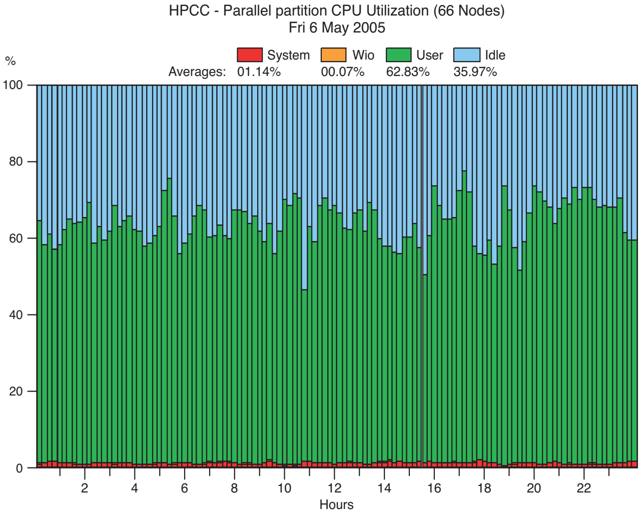


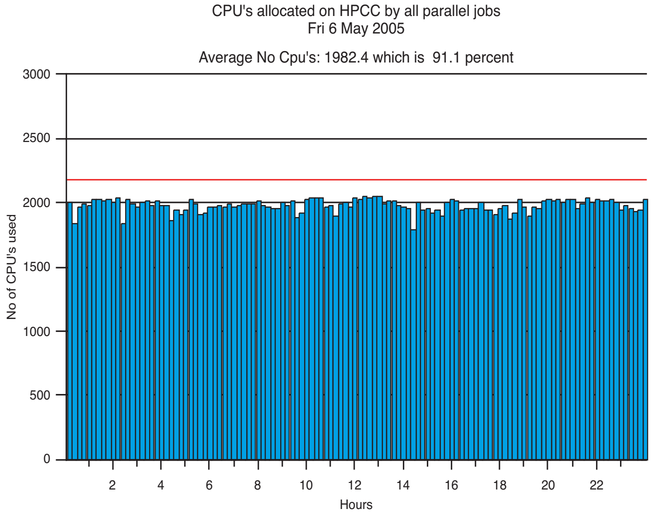
IBM HPCF - Phase 3

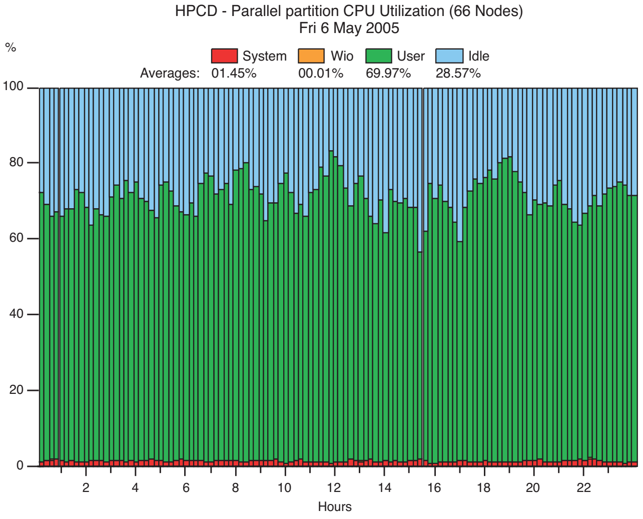
- 2 identical clusters: HPCC and HPCD
- Overall performance: 2.5 Tflops sustained
- HPCC
  - Available from Dec.2004
  - Usage profile: ECMWF operational suite & ECMWF research
- HPCD
  - Available from Sept.2004
  - Usage profile: Member States' applications and research & ECMWF research

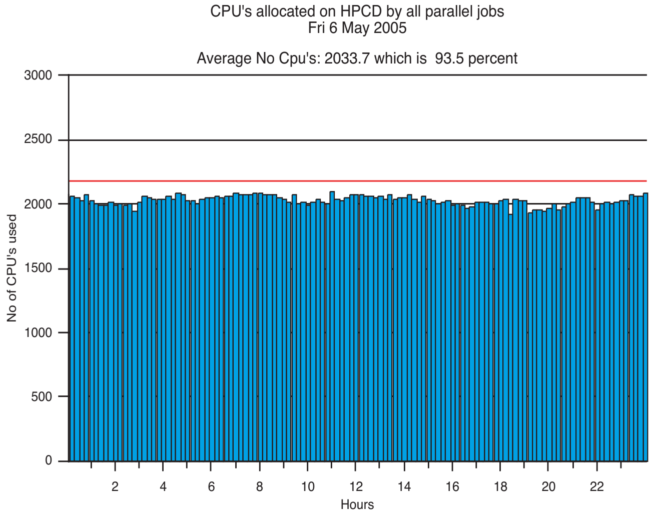

WEEKLY AVAILABILITY STATISTICS
HPCC_CLUSTER from 20041222 to 20050501
User Availability = 99.72 %
Average Hours / Week
User Availability: 167.5
System Session ECMWF: 0.3
Other Downtime: 0.0
SW Maintenance: 0.0
SW Crash: 0.0
HW Maintenance: 0.2
HW Crash: 0.0
Hours / Week
168
162
156
150
144
138
132
126
120
WEEK 53
2005
WEEK 2
WEEK 4
WEEK 6
WEEK 8
WEEK 10
WEEK 12
WEEK 14
WEEK 16


WEEKLY AVAILABILITY STATISTICS
HPCD_CLUSTER from 20040913 to 20050501
User Availability = 98.77 %
Average Hours / Week
User Availability: 165.9
System Session ECMWF: 0.8
Other Downtime: 0.4
SW Maintenance: 0.1
SW Crash: 0.6
HW Maintenance: 0.1
HW Crash: 0.0
Hours / Week
168
162
156
150
144
138
132
126
120
W38
W40
W42
W44
W46
W48
W50
W52
W1
2005
W3
W5
W7
W9
W11
W13
W15
W17


HPCC - Parallel partition CPU Utilization (66 Nodes)
Fri 6 May 2005
%
System
Wio
User
Idle
Averages: 01.14%
00.07%
62.83%
35.97%
100
80
60
40
20
0
2
4
6
8
10
12
14
16
18
20
22
Hours


CPU's allocated on HPCC by all parallel jobs
Fri 6 May 2005
Average No Cpu's: 1982.4 which is 91.1 percent
No of CPU's used
3000
2500
2000
1500
1000
500
0
2
4
6
8
10
12
14
16
18
20
22
Hours


HPCD - Parallel partition CPU Utilization (66 Nodes)
Fri 6 May 2005
%
System
Wio
User
Idle
Averages: 01.45%
00.01%
69.97%
28.57%
100
80
60
40
20
0
2
4
6
8
10
12
14
16
18
20
22
Hours


CPU's allocated on HPCD by all parallel jobs
Fri 6 May 2005
Average No Cpu's: 2033.7 which is 93.5 percent
No of CPU's used
3000
2500
2000
1500
1000
500
0
2
4
6
8
10
12
14
16
18
20
22
Hours

Framework for MS time-critical applications

- The framework was discussed at last year's TAC and approved by Council.
- There are 3 options:

  1) Simple job submission monitored by the Centre:
     - Enhancement of the "job submission under SMS control" facility
     - Based on the ECaccess framework
     - Service available to all registered users.
  2) Member State SMS suites monitored by the Centre:
     - Suitable for more complex applications with several tasks with interdependencies amongst them
     - SMS suites developed according to technical guidelines to be provided by the Centre
     - To be requested by the TAC representative of the relevant Member State.
  3) Member State SMS suites managed by the Centre:
     - Further enhancement of the previous option
     - Application developed, tested and maintained by the MS
     - It must be possible to test the application using ECMWF e-suite data
     - MS suite handed over to ECMWF
     - MS responsible for the migration of the application, ECMWF will monitor this suite
     - ECMWF could provide first-level on-call support, while second-level support would be provided by the MS
     - To be requested by the TAC representative of the relevant Member State.
- Current MS activities
  - The NORLAMEPS system, which requires a "Targeted" version of ECMWF EPS to initialise their LAM, has been implemented as "option 3" and has been running at ECMWF since February 2005.
  - Recently, Italy asked the Centre to support the COSMO-LEPS suite and the IFS-EuroHRM-EuroLM suite as "option 2". The process of implementing them has started
  - Finland has informally asked about the possibility of running a back-up version of their operational HIRLAM model at ECMWF.
- Technical guidelines to advise on the development of such suites are being written.

IBM HPCF - Phase 4

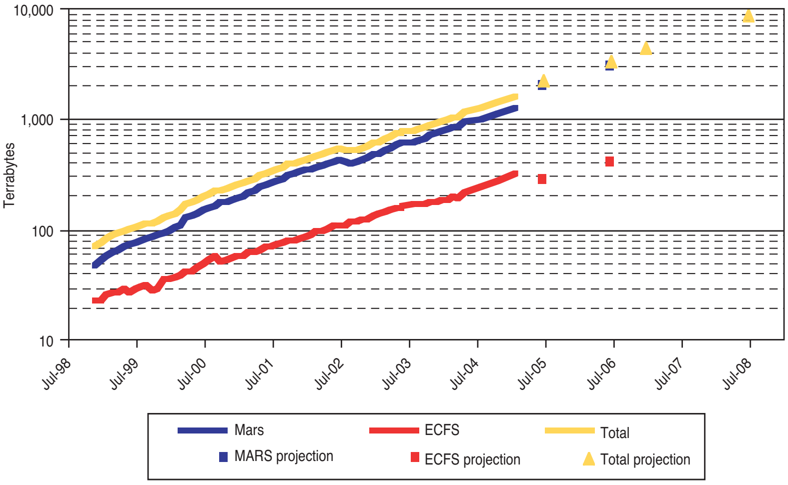
- The IBM contract will be extended to March 2009.
  - Council decision, 61st session (December 2004).
- Two new "Phase 4 clusters" will replace the existing Phase 3 clusters in 2006 and deliver about twice their performance.
- Overall performance of about 4.5 Tflops sustained
  - 2 identical clusters, consisting of p5-575+ SMP servers, connected by the pSeries High Performance Switch (the exact number of nodes is not yet determined, as this is dependent on the results of the performance test)
  - about 50 TB of disk space per cluster.
- Each p5-575+ server will have:
  - 16 POWER5+ CPUs (8 dual-core chips)
  - 32 GB of memory (a few will have 128 GB)
  - The CPUs incorporate simultaneous multi-threading technology.

DHS

- The HPSS-based system continues to perform very well.
- All the Phase 3 equipment has been installed. Some of this equipment was installed in the Disaster Recovery System building.

- The system currently consists of:
  - STK tape silos,
  - IBM p-Series p650 and p660 servers,
  - FAStT fibre-channel disks,
  - IBM 3592 tape drives for primary data storage and
  - LTO-2 tape drives for secondary (backup) data storage
- Phase 4 equipment will be installed later this year.
- The ECFS migration started at the beginning of last year and was completed in November.
- 165 TB of data in 10 million files residing on over 5000 tape cartridges were "back-archived" (i.e. transferred from the old system to the new one).
  - The back-archiving is described in more detail in the latest edition of the ECMWF newsletter.
- Backup of ECFS data — please note:
  - **by default, no secondary (backup) copy is made of ECFS data (unlike on the old ECFS system).**
  - **The user has to specify the "-b" option on the "ecp" command to request that a secondary copy be made.**
- HPSS upgrade to version 6 is likely later this year
  - This is a major change that dispenses with the need to use DCE (Distributed Computing Environment).
  - As usual, we will perform the upgrade as transparently as possible, without any major downtime of the DHS service.

Volume of data stored in the archive

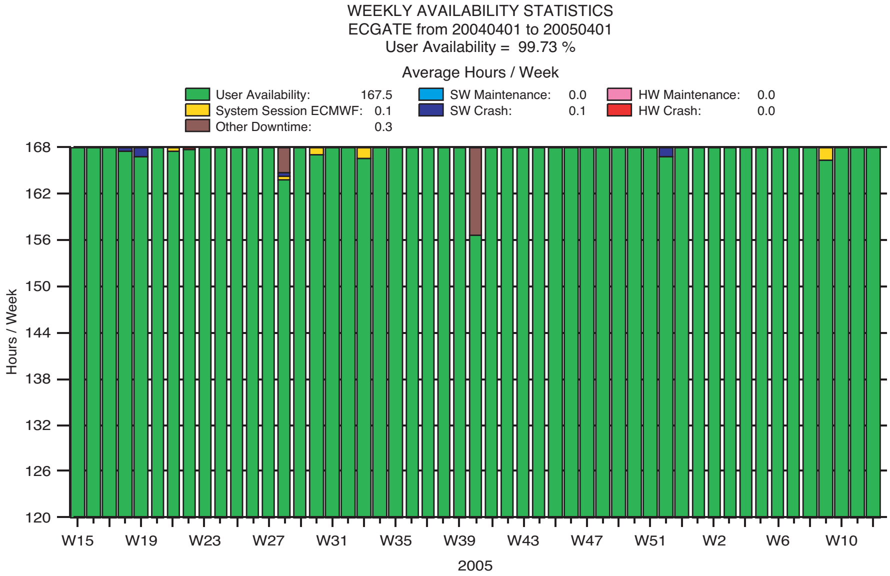


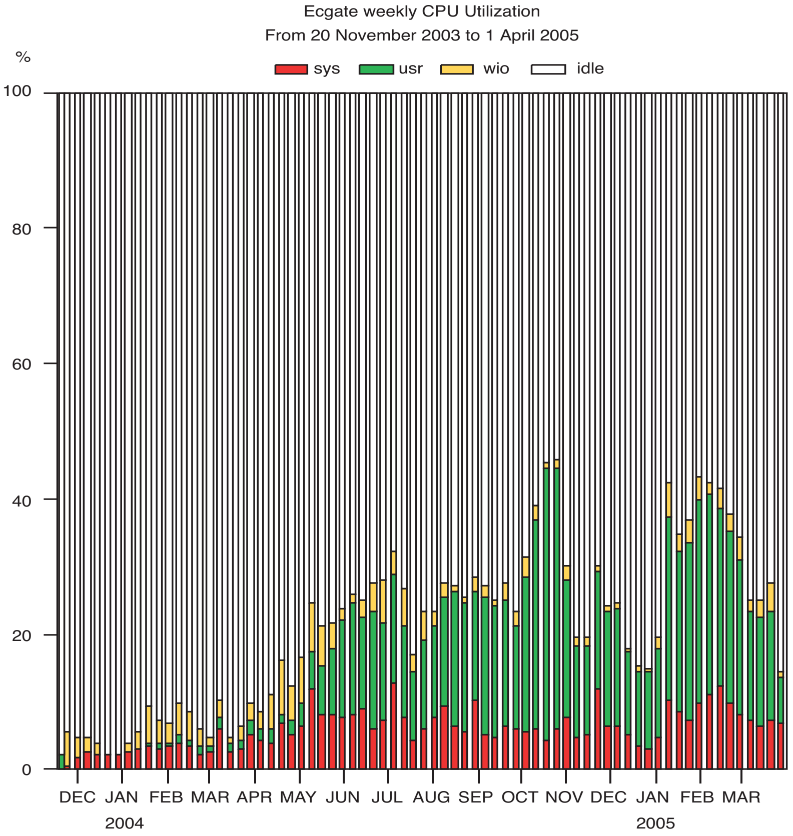
These values do not include the secondary (backup) copy of the most critical data.

Servers and Desktops

- The desktop Linux systems are being upgraded to newer versions of the various system components (SUSE 9.1, KDE 3.2, VMware 4, Windows XP SP2, Office 2003, ...)
- All SGI Origin Servers have been decommissioned.
- Following an ITT, a replacement Highly Available System for data acquisition, pre-processing and dissemination was installed in 4Q2004:
  - 4 HP Integrity Servers, each with 4 1.5 GHz Itanium2 CPUs, 4 GB memory
  - 1 HP Integrity Server with 2 1.5 GHz Itanium2 CPUs, 8 GB memory (development system)
  - an EVA5000 Fibre Channel Disk Subsystem with ~3 TB usable disk space
  - runs HP-UX 11 and HP Serviceguard to provide High Availability.

- The Linux Cluster is being gradually introduced into service.
  - It is currently used to produce plots for the web and for printing.
  - It will be used for verification jobs soon.
- ecgate has continued to provide a stable service:
  - overall availability exceeds 99.7%
  - cpu usage is roughly 35% of capacity.

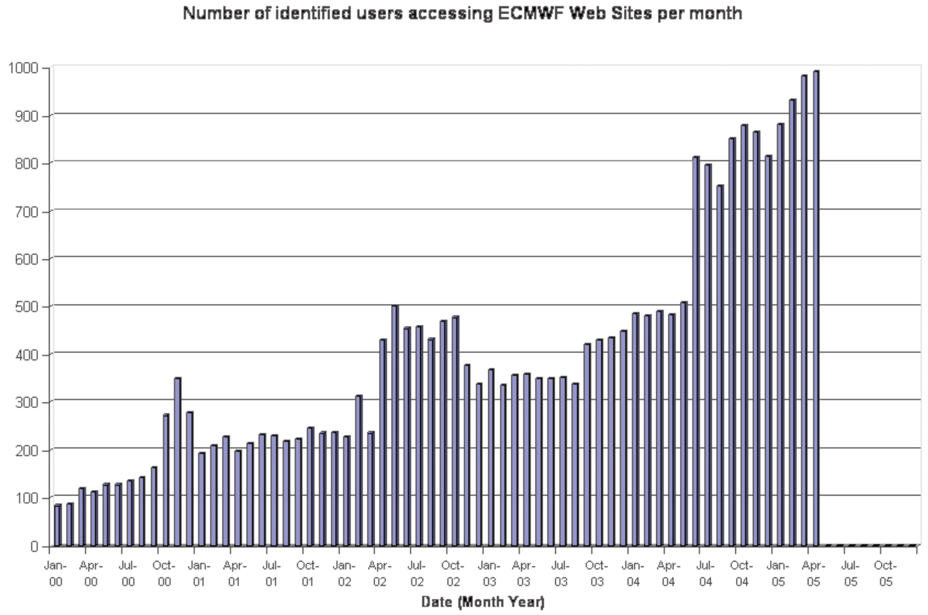




Web Service

- The ECMWF web servers continue to provide a stable and reliable service. New content includes:
  - Monthly Forecast charts
  - WMO EPS Meteograms
  - Web based Content Management System for News and Press Releases
  - The addition of the interface to the Entity Management System to allow Computing Representatives to register users
- The use of the ECMWF web site continues to increase.
- The ratio of identified to anonymous users shows a significant increase, due to the addition of the web-only self-registration for domains, since the introduction of the new web login last June.

Web Service — No. of identified users



Web Service — Statistics

	2001	2002	2003	2004
Total number of page accesses by all users (millions of pages/year)	4.08	8.09	10.9	13.6
Change compared with previous year (% increase)	11.8	98.0	35.0	25.2
Total number of page accesses by identified users (millions of pages/year)	0.58	0.95	1.56	2.02
Change compared with previous year (% increase)	134.4	64.2	68.7	26.5
Average time between page accesses (seconds)	7.7	3.9	2.89	2.31
Ratio of total users to identified users	7.1	8.5	6.8	6.8

- A strategic project to develop web service interfaces to main ECMWF tools has been started under the "Plots-on-Demand" project. This will expose MARS, ODB, Verification and Magics through a common Web Service API and enable the development of a new application for delivering plots on demand.
- A JetStor disk array (6.5TB) has been evaluated and will be used (with a suitable IBM xSeries server) to provide a cost-effective enhancement to the ECMWF Data Server for the ENSEMBLES EU project.

Entity Management System

- The Entity Management System has been used by the Call Desk and User Support to register both internal and Member State users.
- The system has been extended to enable Computing Representatives to carry out certain registration tasks directly via a browser interface.
  - The interface has been tested by User Support since summer 2004.
  - More recently, the interface for Computing Representatives has been tested by KNMI and UKMO.
- The web registration interface is available for MS use.

LAN

- Phase 2 of the High Performance Network was delivered in September 2004.
  - Core of the network is based on two Force10 E600 switches interconnected by 4x10GE.
- ITT for the replacement of the General Purpose LAN was issued early February 2005.
  - Responses are under evaluation.
- Investigate options for the introduction of IP telephony.
- Extend the wireless LAN into all ECMWF office areas.

RMDCN

- 45 sites are connected to the RMDCN.
- New members since last year's meeting:
  - India's connection to the RMDCN was accepted on October 2004.
  - Serbia and Montenegro's connection to the RMDCN was accepted on November 2004.
  - Saudi Arabia has been connected to the RMDCN and is in the process of acceptance.
- Migration of transport technology from Frame Relay to MPLS (Multi-Protocol Label Switching) is planned.
  - Proposal was supported by ECMWF Council and by WMO region VI.
  - The migration would result in doubling the access capacity for all current RMDCN members.
  - Supplement to the RMDCN contract is being discussed with Equant.
- The new standard package for each Member State would be:
  - 768 kbps access line
  - 768 kbps IP Gold port
  - Enhanced backup at 384 kbps.
- Migration to MPLS for the first RMDCN sites should start later this year.
- Co-ordinate Phase 2 of IPSec tests between RMDCN members to investigate the use of Internet-based Virtual Private Networks in an operational environment.
  - Final results will be presented during the next ROC meeting
- The Centre's Internet was upgraded to 70 Mbps in early March 2005.

ECPDS

- New software, ECPDS, has been developed to support the foreseen increase in the dissemination requirement.
- ECPDS offers different transport mechanisms (FTP, SFTP) and the possibility of using the ECaccess network to securely disseminate products over the Internet.
- Migration to ECPDS started on 11 April 2005 and almost all destinations receive now products via ECPDS.
- Monitoring interface is available through the RMDCN and the Internet.

Infrastructure work

- A new 2MVA Uninterruptible Power Supply system was installed and integrated with the two original UPS systems:
  - to provide increased UPS capacity to restore N+1 resilience
  - to replace one of the old standby generators.
- A Gaseous Fire Suppression System which would utilise an inert gas to extinguish any fire in the computer hall or tape library was installed.

Other activities - GRID

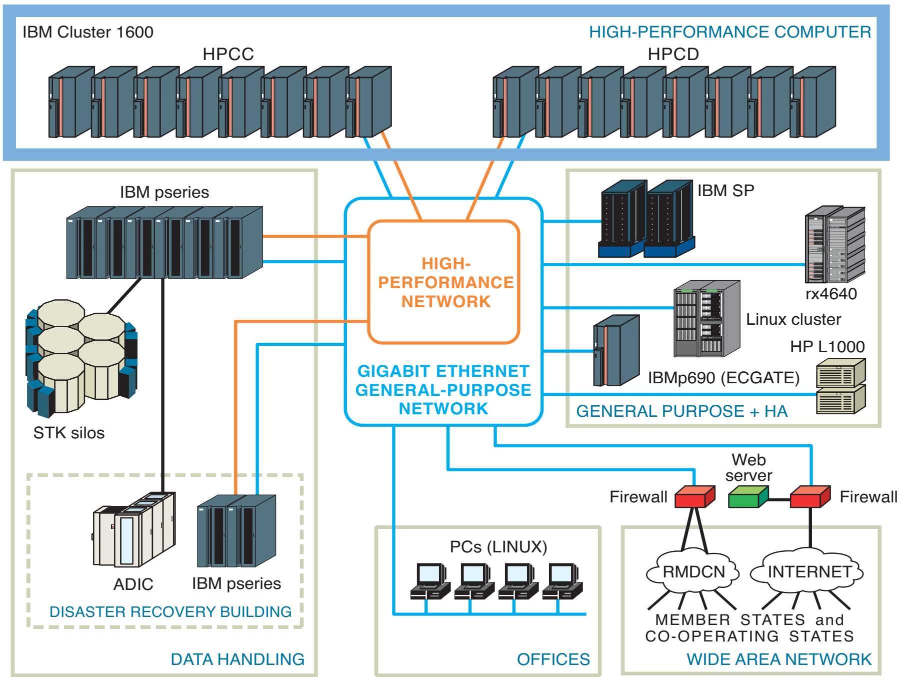
- DEISA
  - Continue to actively participate and so obtain a better understanding of GRID middleware, multi-cluster GPFS and multi-cluster LoadLeveler.
  - Security model that fits well with ECMWF's security policy has been proposed and development will start soon.
- SIMDAT
  - Co-ordinate the meteorological activity of the project.
  - Capture of the requirements of the V-GISC (Virtual Global Information System Centre) has been completed.
  - Technical design of the V-GISC demonstrator has been finalised and development has started.

Major ongoing/planned activities

- Start tests on multi-cluster GPFS for the HPCF clusters
- Continue to implement DHS - Phase 4
- Update the DHS to HPSS version 6
- Complete ITT for the replacement of the General Purpose LAN
- Complete the implementation of the application monitoring system based on HP OpenView and Big Sister
- Organize and co-ordinate the migration of the RMDCN transport technology from Frame Relay to MPLS
- Deploy a unified ECMWF Certificate Authority and Registration Authorities for X509 certificates to ECaccess, VPN services, IPSEC routers, web users, DEISA and SIMDAT
- Enable ECaccess to be used as part of the framework for submitting and monitoring time-critical Member State applications and investigate options for a high-availability service.
- Implement the V-GISC demonstrator, by deploying a Grid infrastructure between the partners that offers transparent and secure access to distributed data
- Implement "plots-on-demand" based on web services
- Install a 4th UPS machine
- Install an additional chiller to provide more chilled water capacity
- Complete the installation of the water mist fire suppression system
- Complete the work on the extension of the Computer Hall.

HPCF & DHS Update — *Neil Storer*

HPCF



Phase 3 timetable

- HPCD was installed over summer, “Ready For Trial” mid-Aug.
- MS jobs started running on the HPCD in September.
- The Operational Suite moved to the HPCD in October.
- HPCC “Ready For Trial” in mid-Dec.
- Some changes were made to the job scheduling system to give MS jobs better turnaround and to help alleviate problems seen when we first started running mixed OS and MS workloads on HPCD.
- The Operational Suite moved to HPCC in April.
- The users are exceedingly pleased with service provided by the Phase 3 systems.

HPC paging problems

- We have seen various instances of “paging problems”.
- The interactive service in particular has suffered several occasions when users ran applications that used larger amounts of memory than they expected.
- When paging gets really bad, the system starts to kill processes, not necessarily the ones causing the paging. Sometimes the system “hangs”.
- We plan to change the interactive “soft limits” for:
  - data 1 GB
  - stack 512 MB

  It is possible for the user to override these values.
- For batch jobs paging is often catastrophic. A feature in the next release of the system will kill jobs that page excessively, rather than letting them continue to run hundreds of times slower, as they would otherwise do.

ECMWF often has requests for more memory but this is not generally practical. Memory usage can be reduced by using a combination of OpenMP and MPI. In jobs using MPI uniquely, much memory is taken up by content replicated over all processors which is only used by individual processors. The number of MPI tasks should be cut down and processing split within MPI tasks, using OpenMP. This will save considerable amounts of memory.

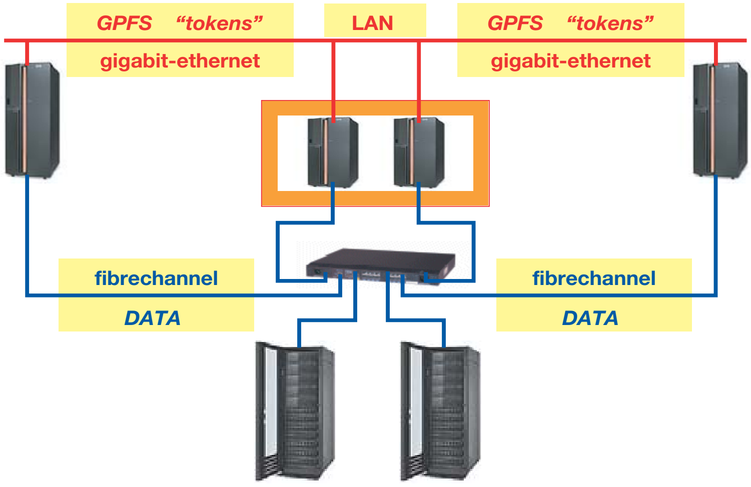
Member State file systems (HPCD)

- ms_home
  - quota-protected (80 MB per user), same as ECMWF "home"
  - fully backed-up: weekly full + daily incremental dumps.
- ms_temp (6 TB - 60% full today)
  - increased from 2TB to 6TB in April
  - not backed-up
  - no run of select-delete since the increase
  - previously select-delete runs caused mainly by "rogue" jobs.
- ms_perm (250 GB - 10% full today)
  - not backed-up (by ECMWF)
  - not controlled by select-delete
  - "administered" by User Support.

Multi-cluster GPFS (MC-GPFS) pilot study

- The latest version of GPFS enables "native" access to data from multiple clusters concurrently at much higher data rates than are possible using NFS.
- Currently various data are replicated on both clusters, effectively reducing the amount of usable disk space. MC-GPFS removes the need to replicate the data.
- Currently, data is transferred between clusters over the LAN, either using FTP-like applications or via ECFS. MC-GPFS enables each cluster to access the data efficiently, directly over a fibre-channel storage area network.
- MC-GPFS should help with resiliency.
- MC-GPFS removes synchronisation problems (e.g. out-of-date copies) since there is only 1 version of the data. MC-GPFS helps with data management.

Multi-cluster GPFS configuration



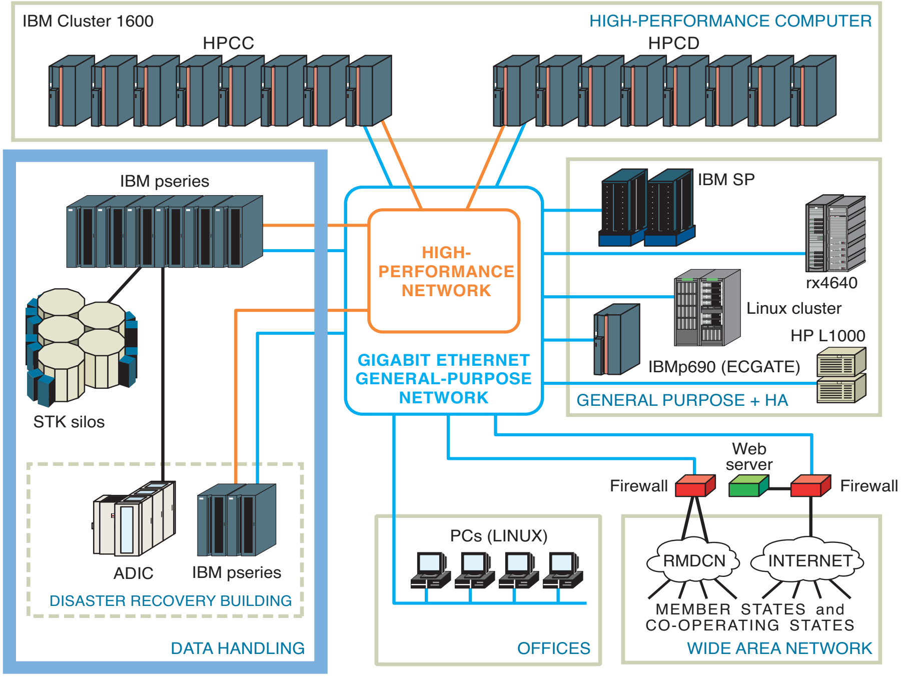
HPCF plans

- The contract extension (until 1Q09) includes:
  - Replacement of both clusters in 1H 06 with two new clusters:
    - 16-way Power5+ nodes
    - 32 GB memory per node (4 nodes per cluster with 128 GB)
    - 8-way Power5+ I/O & network nodes
    - 65 TB of (raw) disk space per cluster
    - Multiple (probably 8) nodes per cabinet.
- Performance commitments are based on our three main applications (deterministic forecast, 4D-VAR, EPS);
- The sustained performance will increase from ~2.5 TF to ~4.5 TF;
- IBM expects a much better percentage of peak performance with the Phase 4 system, due to simultaneous multi-threading (SMT) and better memory bandwidth.
- We plan to issue an ITT for a replacement HPCF in 2007.

Simultaneous Multi-Threading

- Extra hardware in each of the CPUs (or "cores") enables them to execute 2 threads of instructions simultaneously. Certain registers are duplicated, functional units are not. This is different from having 2 distinct CPUs on a chip.
- To the operating system it appears as if there are twice the number of CPUs. A 16-way SMT system would appear to have 32 CPUs. So to use it effectively you would run at least 32 threads, either as 32 separate single-threaded processes or a parallel job using 32 threads (MPI, OpenMP or a hybrid of the two).
- It is difficult to estimate the performance gain that programs can expect by utilising SMT; in some instances there could actually be an overall loss of performance.

DHS



ECFS migration

- The ECFS service has been migrated completely to the new HPSS-based system. The TSM system was terminated at the end of last year.
- The migration was done in such a way that it was totally transparent to the users.
- Over 9 months the ECFS team ran 18,000 "back-archiving" tasks, using ECMWF's SMS batch scheduler. These tasks used an SQL database to control and keep track of the progress of the "back-archive" and this helped considerably to simplify and streamline the process.
- This "back-archive" process transferred 165 TB of data in 10 million files that resided on over 5000 tape cartridges in the TSM-based system, without any loss of data.

ECFS

- The ECFS file size limit has been increased from 2GB to 6GB. We have actually successfully tested 32 GB files, but have chosen the 6 GB limit because of the way HPSS performs file allocation. Be aware that certain Unix systems cannot handle files over 2GB in size.
- An "emv" command is available to rename a file in ECFS. Currently this only works if the source and target files are in the same directory. The command is being modified to allow the file to be moved into a different directory. Eventually "emv" will work with directories, not just files, to enable users to rename their files in ECFS.
- At present it is not possible to use the recursive option ("-R") on commands such as "els" and "erm". This will be addressed at a later date.

ECFS back-up copies

- Please take note that (unlike the old TSM-based ECFS system) in the new HPSS-based ECFS system, by default, **no** secondary (backup) copy is made of ECFS data.
- The user has to specify the "-b" option on the "ecp" command to request that a secondary copy be made of data that cannot easily be reconstructed, should the primary copy be destroyed.

DHS plans

- It is planned to rewrite the ECFS client software. The current user API (Application Programming Interface) is a set of Perl scripts. This design does not lend itself to functional and recoverability enhancements.
- Last week a single user job accessed over 10,000 files in ECFS. This is over 30% of the daily total number of accesses. We plan to develop an ECFS scheduler, to manage and control the ECFS workload.
- We plan to upgrade to HPSS version 6 later this year.
- The robotic tape libraries in the main computer hall and the DRS building are no longer manufactured. Maintenance cost for these is starting to increase (in one instance will cease by the end of the decade). We are investigating options for replacing the tape libraries over the next few years.

M. Pithon asked when the new system release with the feature to kill excessively paging jobs was planned to be available. N. Storer replied that the AIX software already allows users to specify the amount of real memory they require and any requirements beyond this amount will result in the job being aborted, rather than paging. However, the current LoadLeveler does not support this feature; the next version of LoadLeveler, which it is hoped to test soon, will have hooks to enable its use. Testing will include trying to find a way of implementing the feature without having a major impact on users' work. A particular problem to be taken into account is that previous jobs may have left shared memory segments on nodes and this should not cause current jobs to abort.

E. Krenzien asked when the rewrite of the ECFS client software was planned. N. Storer replied that a design had not yet been decided upon. The ECFS server has only just been rewritten. The client software was unlikely to be rewritten before early 2006.

SIMDAT and DEISA projects — *Matteo Dell'Acqua*

DEISA

- Distributed European Infrastructure for Supercomputing Applications
- 5 year infrastructure project partially funded by the EC
  - Contract with EC was signed on 1 May 2004.
- Objective of DEISA is to deploy a production quality HPC infrastructure.
- DEISA consortium includes

  IDRIS - CNRS, France (coordinator)
  FZJ - Juelich, Germany
  RZG - Garching, Max Planck Society, Germany
  CINECA, Italy
  EPCC, Edinburgh, UK
  ECMWF
  SARA, Amsterdam, The Netherlands
  CSC, Helsinki, Finland
  LRZ, Munchen, Germany
  BSC, Barcelona
  HLRS, Stuttgart, Germany

ECMWF involvement in DEISA Activities

- Five service activities and one Grid R&D activity have been defined to support the operation of DEISA Supercomputing Grid Infrastructure.
- SA2, Data Management with Global File Systems: Deployment and operation of a global distributed file system, based mainly on GPFS
  - Project has been set-up to test multi-cluster GPFS.
- SA3, Resource Management Deployment and operation of global scheduling services based mainly on Multi-cluster LoadLeveler and Unicore.
  - Currently ECMWF does not plan to use multi-cluster Loadleveler internally. We have reviewed the design document and made suggestions to improve the usability of the current version.
- Both SA2 and SA3 would greatly benefit from obtaining a network connection between ECMWF and the core sites (CINECA, FZJ, IDRIS, and RZJ).
- SA5 Security: Provides administration, authorization and authentication for DEISA, with special emphasis on single sign-on:
  - Enhance UNICORE to support strong authentication for the submission of jobs to DEISA infrastructure
  - Propose a security model supporting strong authentication and fine-grain authorisation.
- JRA7 Access to Resources in an heterogeneous environment: Development of Grid middleware based on Web Services standards with the objective of using OGSA standards in the near future
  - Participation in the design and tests.

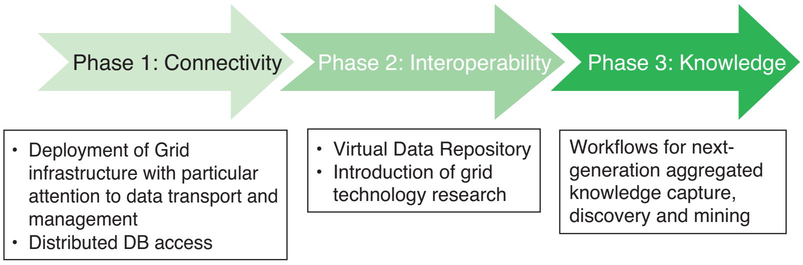
SIMDAT

- Data Grids for process and product development using numerical simulation and knowledge discovery.
- 4 year project funded by the EC
  - Contract with EC was signed on 1 September 2004.
- SIMDAT focuses on 4 applications:
  - Product design in automotive and aerospace
  - Process design in life science
  - Service provision in meteorology.
- Objective of SIMDAT is to use data grid technology to resolve a complex problem for each of the 4 applications

SIMDAT Strategy

- 7 Grid-technology areas have been identified for achieving SIMDAT objectives:
  - Integrated Grid infrastructure offering basic services to applications
  - Access to data distributed on Grid sites
  - Management of Virtual Organisation
  - Ontology
  - Integration of analysis services
  - Workflows
  - Knowledge Services

Meteorological application



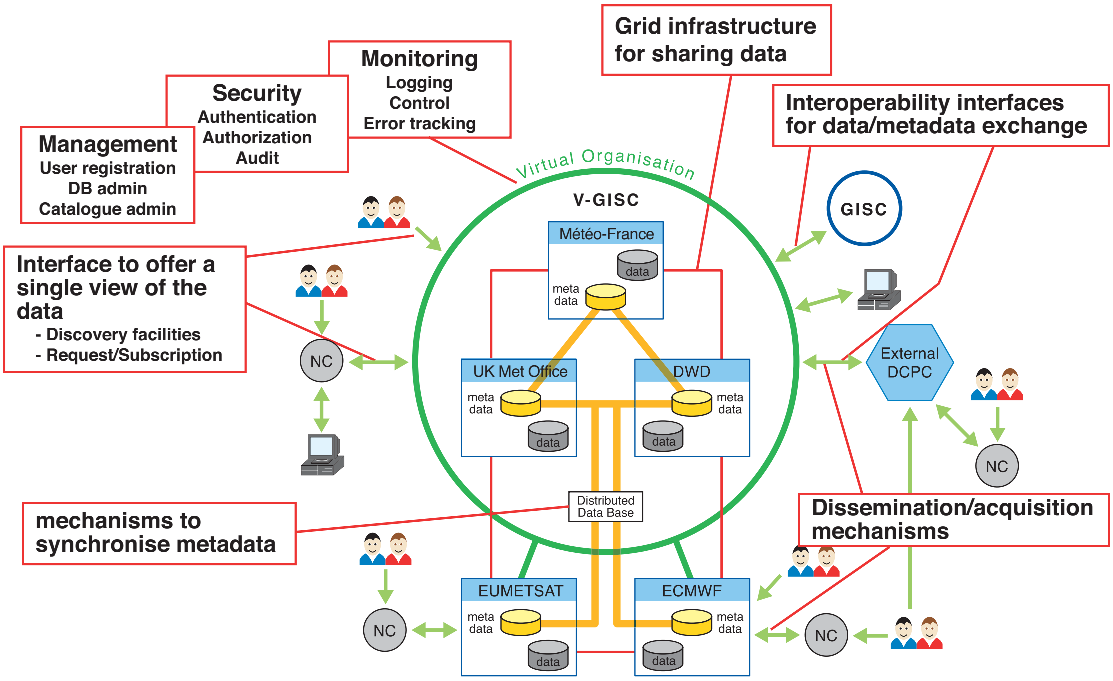
- Deployment of Grid infrastructure with particular attention to data transport and management
- Distributed DB access

- Virtual Data Repository
- Introduction of grid technology research

Workflows for next-generation aggregated knowledge capture, discovery and mining

- 5 partners: DWD, Meteo-France, UK Met Office, EUMETSAT and ECMWF
- A complex problem: To build a Virtual GISC, an integrated and scalable framework for the collection and sharing of distributed data that will offer:
  - A single view of meteorological information which is distributed amongst the 5 partners
  - Discovery facilities and standardised retrieval mechanisms
  - Standardised mechanism for routine dissemination of data
  - Standardised mechanism for collection of data
  - Quality of service, efficiency, reliability and security
  - Processing services and shared data manipulation facilities.
- Grid technology will be used:
  - To connect the diverse data sources and create a Virtual Database
  - To enable flexible, secure collaboration through virtual organisation.

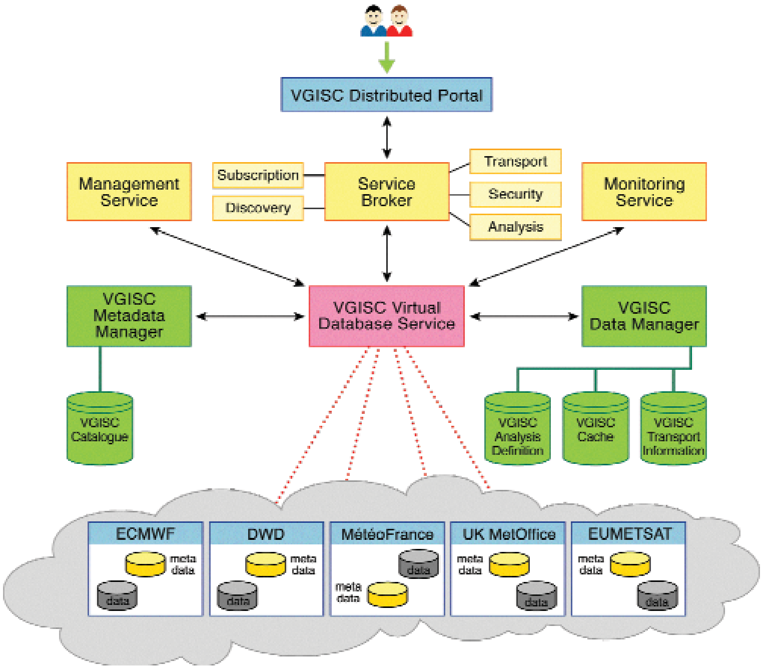
V-GISC infrastructure



V-GISC Conceptual view

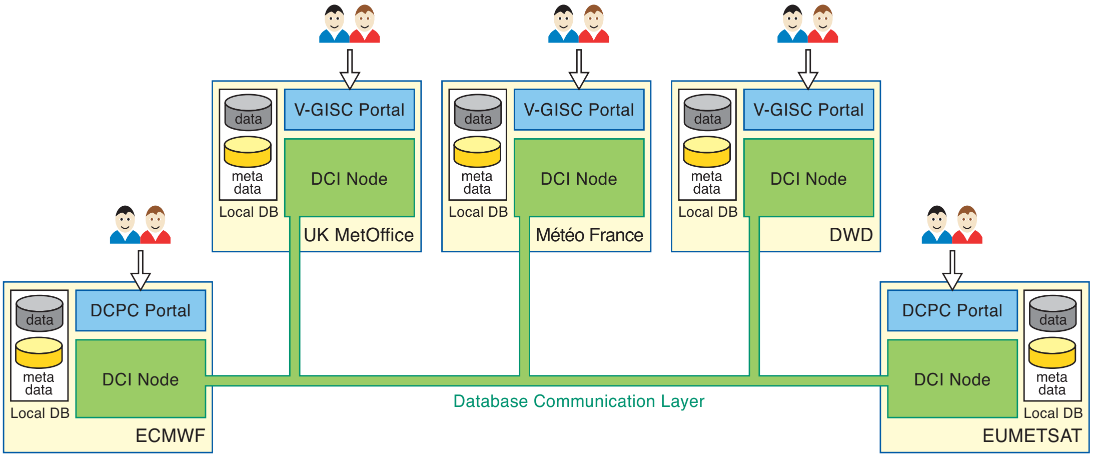
- Virtual Database
  - Provide a unified view of all the shared datasets through a distributed catalogue
  - Maintain the distributed catalogue amongst the partners using synchronization mechanisms
  - Provide interfaces with the legacy databases
  - Implement data replication mechanisms
  - Preserve the integrity of the data.
- Data access and distribution Services
  - Collection & dissemination services that support secure, efficient and reliable transport mechanisms
  - Quality of Service (QoS): traffic prioritization, queuing mechanisms, scheduling
  - Discovery service by browsing the catalogue or using a keyword search engine
  - Interactive and batch interfaces.
- Virtual Organisation
  - Security Services (CA, AuthN, AuthZ, Audit,...)
  - User management
  - Data policy management
  - Monitoring and control.

V-GISC Conceptual view



- Through the Distributed Portal user searches for and retrieves data and subscribes to services, subject to authentication and authorization
- The Virtual Database Service provides a single view of partners' databases

VGISC Distributed Architecture



Introduction to ECPDS (ECMWF Product Dissemination System) — *Laurent Gougeon*

Project Overview

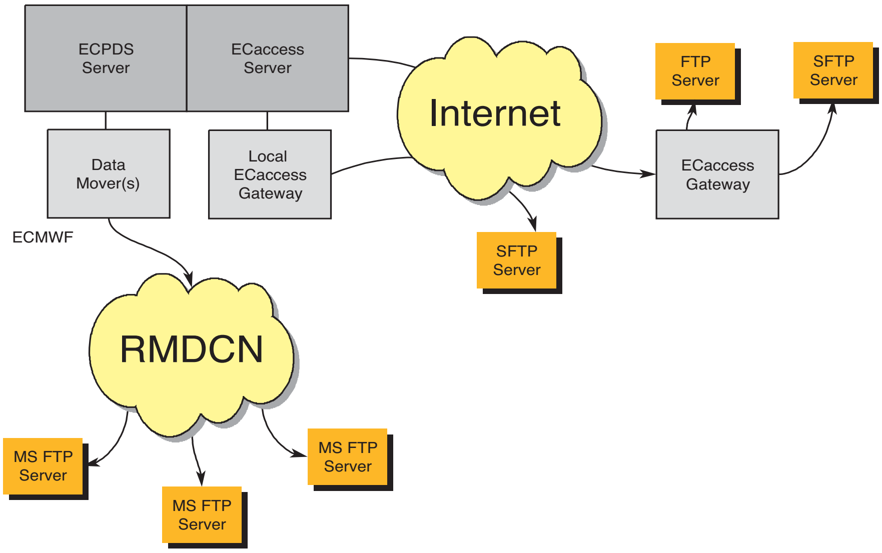
- QFTD was used to disseminate ECMWF products
  – Could not cope with the increasing requirements.
- Goals and objectives of ECPDS
  - General purpose data transmission system
  - Allow Member and Co-operating States to specify which data to deliver, on which target systems, using which networks (RMDCN or Internet)
- Scope of project
  – Provide reliable and secure transfer mechanisms
- FTP (RMDCN), SFTP (Internet without Remote Gateway), ECaccess (Internet with Remote Gateway)
  – Provide Management & Monitoring capabilities
- ECMWF administrators & analysts
- Member and Co-operating States
  – Provide Alarms and real-time Displays

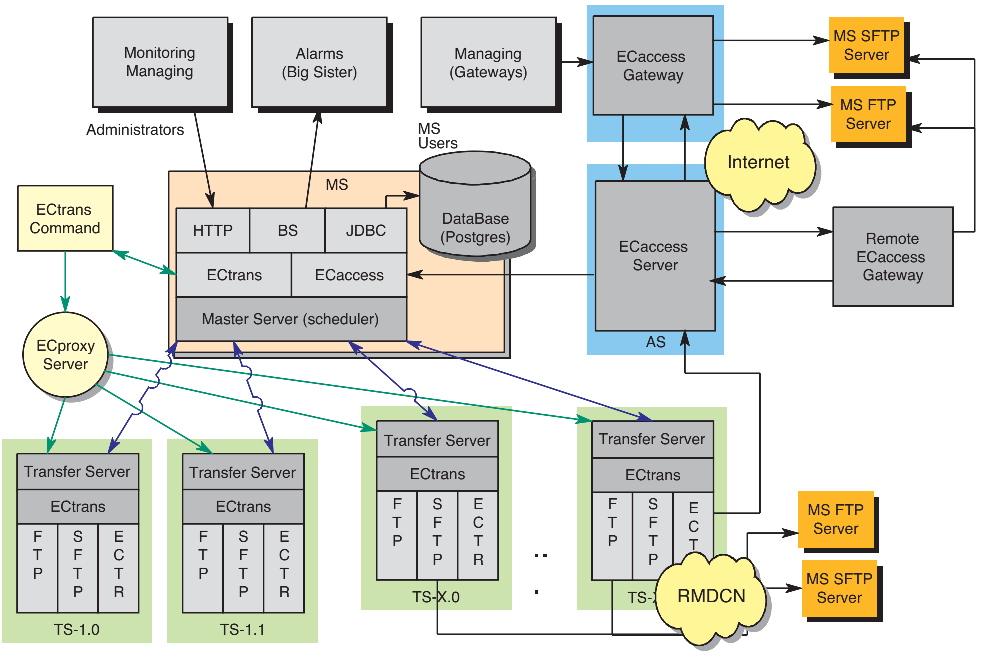
ECPDS vs. QFTD

- Context, platform and architecture independent
  – based on Java Technology
  – persistence implemented via any SQL Database
- Highly configurable
  – scalable across different hardware
  – dynamic system behaviour
- Additional features
  – transfer scheduler
  – host check scheduler
  – destination aliases
  – transfer modules
  – keep alive feature
  – mail notifications
  – access control

ECPDS architecture

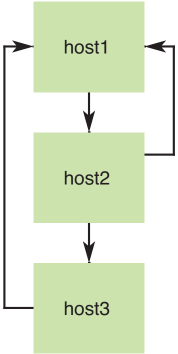
- Main components
  – Master Server, Data Mover(s), monitoring server(s), ECproxy server(s) and ECpds command
- Master Server
  – Transfer scheduler
  – Database access
- Data Mover(s)
  – Data file storage
  – Transfer protocols
- Monitoring server(s)
  – Management
  – Monitoring
- ECproxy Server(s)
- ECpds Command





ECPDS scheduler

- Policy on destinations
  - On host failure, max connections, retry count, retry frequency, max start, start frequency, reset frequency
- Policy on hosts
  - max connections, retry count, retry frequency, check frequency, timeout, target directory, transfer modules



ECPDS transfer modules

- All modules

```
ectrans.buffSize=”65536”
ectrans.closeAsynchronous=”no”
ectrans.closeTimeOut=”30000”
ectrans.connectTimeOut=”30000”
ectrans.delTimeOut=”60000”
ectrans.doFlush=”yes”
ectrans.getTimeOut=”0”
ectrans.listTimeOut=”30000”
ectrans.mkdirTimeOut=”30000”
ectrans.mov eTimeOut=”30000”
ectrans.putTimeOut=”0”
ectrans.readFully=”no”
ectrans.retryCount=”1”
ectrans.retryFrequency=”1000”
ectrans.rmdirTimeOut=”30000”
ectrans.sizeTimeOut=”30000”
```

- Ftp module

```
ftp.commTimeOut=”60000”
ftp.dataTimeOut=”60000”
ftp.ignoreCheck=”yes”
ftp.ignoreDelete=”yes”
ftp.keepAlive=”0”
ftp.lowPort=”no”
ftp.mkdirs=”yes”
ftp.passive=”no”
ftp.portTimeOut=”60000”
ftp.postConnectCmd=””
ftp.preCloseCmd=””
ftp.preGetCmd=””
ftp.prePutCm d=””
ftp.prefix=””
ftp.suffix=”.tmp”
ftp.usetmp=”yes”
ftp.vms=”no”
```

- SFtp module

```
sftp.mkdirs=”yes”
sftp.prefix=””
sftp.sftpConnectTimeOut=”10000”
sftp.sftpSessionTimeOut=”60000”
sftp.suffix=”.tmp”
sftp.usetmp=”yes”
```

... and other modules

– GFtp, LPR ...

- Web Access
  - https://ecaccess.ecmwf.int:9443/
  - https://msaccess.ecmwf.int:9443/

Current status

- All the Member and Co-operating States have been moved from QFTD to ECPDS
  - No major problems identified so far.
- What Next?
  - New ECaccess Gateway with ECpds support (v3.0.0).

N. Olsen noted that when a colleague had recently practised the dissemination change request procedure, he had tried with MARS and saw that the results were not the same. U. Modigliani replied that this was a known problem: in recent years there has been and continues to be much effort to harmonise them as much as possible.

Planned model resolution upgrades in operations — *Alfred Hofstadler*

Resolution Upgrades — Atmosphere

	Deterministic		EPS		MOFC	
	Current	Upgrade	Current	Upgrade	Current	Upgrade
Spectral	T511	T799	T255	T399	T159	T159
Gaussian	N256	N400	N128	N200	N80	N80
Dissemination (LL)	0.5	0.25	1.0	0.5	1.5	1.5
ML – Vertical Resolution	60	91	40	62	40	62

No increase in pressure levels planned.

Resolution Upgrades — Waves

	Deterministic		EPS		Mediterranean		MOFC	
	Current	Upgrade	Current	Upgrade	Current	Upgrade	Current	Upgrade
Lat/Lon	0.5	0.36	1.0	1.0	0.25	0.25	1.5	1.5
Dissemination /LL	0.5	0.25	1.0	1.0	0.25	0.25	1.5	1.5
Frequencies	30	30	25	30	30	?	25	25
Directions	24	24	12	24	24	?	12	12

Upgrade of Mediterranean wave model needs further scientific investigation.

Timetable for IFS cycle 30r1 — high resolution

- Mid May–mid June: RD testing
- Mid June: First operational testing
- End June: First technical test datasets for selected operational suites available in MARS
- July–September: Operational e-suite
  - Meteorological test datasets for all operational suites available in MARS
  - Parallel test dissemination for selected dates
- End September: Implementation
- December: increase in run-length for medium-range from 10 to 14 days, including VAREPS
- March 2006: linking MOFC to VAREPS

Impact on users

- Field sizes:
  - Model output (SH and GG) -> x 2.5
  - Lat/Long -> x 4
  - Extra model levels -> x 1.5
- Dissemination
  - Problem with GG/AUTOMATIC
  - Selection of nearest “new” model level
  - Nearest GRID point co-ordinates for Weather Parameter requests will change. Member States have to select new GRID point co-ordinates or rely on interpolation.
  - Line capacity
  - Production Schedule should stay the same.
- MS jobs
  - Check new disk space, memory, CPU, line bandwidth requirements.
- MS projects
  - Use test data sets to run “e-suites” and decide on new configuration
  - Review resource requirements (disk space, memory, CPU, line bandwidth)
- EMOSLIB 281
  - New Gaussian definitions
  - New automatic truncation
  - will become default version
  - MARS and Metview_new have been relinked
  - MS graphics applications (Metview and MAGICS) need to be relinked

G. Wotawa asked whether States would have an opportunity to test their jobs with the new resolution, before it became operational. F. Hofstadler replied that an e-suite model version will be available to run tasks in parallel with the current suite for some time before the operational change.

J. Greenaway asked whether the trajectory database would be upgraded in line with the increase in resolution of the model. U. Modigliani replied that, although ECMWF maintains the database for the trajectory model, KNMI maintains the model itself. F. Hofstadler added that it would not be easy to interpolate the data to a lower resolution for the database, as the model levels would also change. Some work on the model would be necessary.

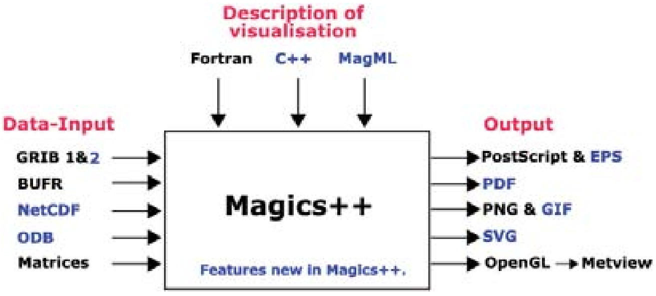
R. Sharp enquired whether the change in vertical resolution included an increase in the top level. F. Hofstadler replied that the top level would also increase in height.

R. Rudsar asked whether there were any plans for a general upgrade of the line capacity of the standard RMDCN package to enable States to take advantage of the new volumes of data. U. Modigliani reminded representatives that current plans were to double the capacity of the standard RMDCN package in 2006. This will not allow for the potential four times data volume which will be available from September this year, so States may need to use the Internet in addition. I. Weger added that the new standard RMDCN access line would be 768 kbps. As soon as contract negotiations are complete, an RMDCN Operations Committee meeting would take place to discuss migration schedules with the States and Equant. Equant has already estimated that migration will take at least six months. F. Hofstadler commented that States did not need to receive all the fields in the model resolution to benefit from the upgrade: even if they stay at their current resolution, the quality of the fields they receive will improve.

J. Greenaway asked whether any additions to the GRIB2 dissemination were planned. F. Hofstadler replied that Sea Surface Temperature anomalies from the seasonal forecasting system were currently disseminated on the GTS in GRIB2 and it was planned to augment them by probability fields from the EPS. There are no plans to disseminate to the Member States in GRIB2.

Graphics Update — *Jens Daabeck*

Magics++



Magics++ new features

- ODB data access and plotting
- NetCDF and GRIB 2 data input
- GIF and SVG output
- EPS for easier inclusion of plots in Word and Latex
- Multiple output formats from a single program
- An object-oriented C++ interface
- An XML interface (MagML)
- A new contouring package (Akima)
- A new flexible set of coastline resolutions
- Simplified legend handling
- Better support for text and graphical annotations
- Two-way interaction with Metview, allowing interactive manipulation of plots

Magics++ status

- Contouring including shading, highlights, labels and highs / lows
- Marker and hatch shading
- Line styles, eg DOT and DASH
- Three contour methods plus an automatic method (default) that chooses between them
- Automatic selection of coastline resolution for high quality at fast speed
- Grid value plotting
- Wind plotting
- Coastline plotting, including map gridlines and labels
- Cylindrical and stereographic projections
- GRIB and NetCDF data loading
- Basic ODB access
- User and automatic titles
- Layout (sub-pages, multi-page plots)
- Basic legends
- Basic XML input (MagML)

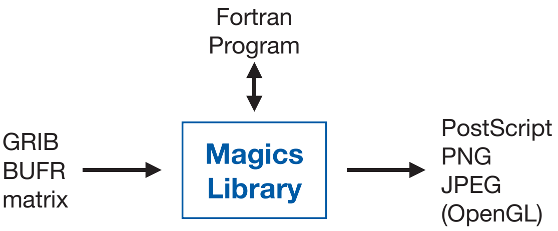
- Basic SVG and GIF/PNG output
- Multiple driver output

Magics++ plans

- Operational release 4Q2005
- 10th Meteorological Operational Systems Workshop 14 - 18 November 2005
- Export version 2006

Magics

Magics is a software library for plotting contours, satellite images, wind fields, observations, symbols, streamlines, isotachs, axes, graphs, text and legends



Magics new features

- Basic support for high resolution fields added
- Improvement in Graph Legend
- Support for scanning mode for data coded in polar-stereographic projection added
- Changes to Satellite visualisation, including improvements for Metview
- Internal performance improvements to take full advantage of the '-O2' option at compilation
- Added titles for seasonal and monthly forecasting products
- The latest internal version of Magics is 6.10 which runs at ECMWF on Linux, including cluster and AIX platforms

Magics 6.9.1 - export

- Available to the Member States
  - January 2005
- UNIX platforms

– Linux	SuSE 7.3 & 9.1 (Cluster 9.0) Portland Fortran compiler
– IBM	AIX 5.1
– SGI	IRIX 6.5
– HP	HP-UX B.11
– HP/Alpha	True64 5.1A (future support required?)
– Sun	SunOS 5.9

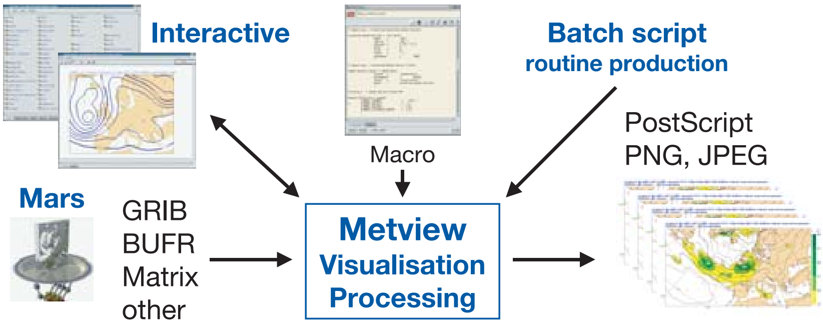
- User Guide in HTML, PDF and PostScript format

Magics plans

- Support for higher resolution forecast
  - Emoslib 281

Metview

- ECMWF’s meteorological data visualisation and processing tool
- Complete working environment for the operational and research meteorologist



Metview new features

- Support for T799 fields
- New application TimeSeries can plot time series either from GRIB data or from geopoints data
- New Tropical Cyclone Tracks plotting module
- Magics fix for end-of-leap-year date axis bug
- New Hovmøller application
- Inlined Macro Fortran functions can now also be written in Fortran 90 on all platforms
- Better handling of 10 bit satellite images and pseudo satellite images including calibrated legend, improved title, partial image (INPE), reprojection
- Latest internal Metview version is 3.7.2, based on Magics 6.10, which runs at ECMWF on Linux including cluster and AIX platforms

Metview 3.7.1 - export

- Available to the Member States
  - April 2005
- UNIX platforms

– Linux	SuSE 7.3 & 9.1 (Cluster 9.0) Portland Fortran compiler
– IBM	AIX 5.1
– SGI	IRIX 6.5
– HP	HP-UX B.11
– HP/Alpha	Not supported (future support required?)
– Sun	SunOS 5.9

- User Guide online
  - PDF and HTML format

Metview plans

- Support for higher resolution forecast
  - Emoslib 281
- Magics++ support

EPS Meteograms

- EPS Meteogram charts available via ECMWF Web pages
  - Shows EPS members' forecast distribution for a model run
- Metview user interface
- BUFR data interface
  - Dissemination files format
- EPS Meteograms also available as standalone system
- Classic Meteograms available at ECMWF via Metview
- Plans
  - Support for higher resolution forecast

J. Daabeck enquired whether any representative's service was dependent on new HP Alpha implementations of MAGICS and Metview. If not, ECMWF are keen to discontinue its support as soon as possible. M. Pithon requested time to check with her colleagues. I. Weger undertook to email States with a reminder survey.

The ECMWF Linux cluster: 1 year on... — *Petra Kogel*

History

- Beginning of 2004, it was decided to evaluate Linux cluster technologies:
  - Would a Linux cluster be suitable as ECMWF general purpose server?
  - Would it be suitable for ECMWF HPC?

Questions

- Do the very new technology components work?
  - Infiniband interconnect
  - Lustre shared filesystem
- Is the software stack there?
  - Compilers
  - Message passing (MPI)
    - Model performance
  - Batch system
  - Load balanced distribution of interactive login sessions
  - Monitoring
- Is it manageable ?
  - Operating system installation and upgrades
    - Centralised ?
    - Downtimes ?
    - Ability to revert ?
  - Time it takes to shutdown / reboot individual nodes
  - Time it takes to shutdown / reboot the whole cluster
  - Maintenance and support
- Is it robust and reliable ?

Cluster configuration: Options and decision

- Issued Request for Proposals for a small cluster on 30 January 2004
- Offers contained configuration choices; we made choices according to price, and current / future expected performance and scalability:
  - Interconnect: Infiniband. Not: Myrinet, Quadrix, Dolphin SCI
  - Filesystems: Lustre. Not: NFS, CXFS, GPFS, Sistina GFS, Polyserve
  - Batch systems: Sun Grid Engine. Not: SLURM, open PBS, PBS Pro, LSF
  - Monitoring: Ganglia. Not: Vendor specific products
  - Cluster management: Vendor specific - Clusterworx + LinuxBios.

Hardware and Operating System

- Supplied by Linux Networx
- Installed 7 May 2004
- 32 nodes plus 1 master node
  - Includes 6 I/O nodes with Fibre Channel HBAs
- Dual 2.2 GHz AMD Opteron CPUs, 4 GB Memory on each node
- SuSE 9.0 Operating System
- On 17 May, started evaluation with focus on use of cluster for ECMWF HPC

Fast Interconnect: Infiniband

- Using pre-beta release kernel modules
- Stable after initial cabling issues
- Up to 750 Megabytes/sec measured for MPI traffic

Shared scalable file system: Lustre

- Only over Gigabit Ethernet, Infiniband not supported then
- Stable
- One bug found, to be fixed in next release .. ended up being almost one year later
- Throughput depends on number of I/O nodes multiplied by per-node throughput to storage device
- Here: 100 MB/sec = controller throughput
- Can be accessed from outside the cluster
  - By installing lustre clients on Linux workstations
  - NFS exporting - eventually

Compilers: PathScale, Portland, Absoft

- Only PathScale compiled IFS code without problems
- MPI: MVAPICH from Ohio State University
- Resulting IFS model performance:
  - Faster than 1.3 GHz IBM Power 4
  - Slower than 1.9 GHz IBM Power 4+

Batch system: Sun Grid Engine

- Designed for Grid computing
- Very configurable
- Can distribute interactive login sessions taking into account load balancing
- Compartmentalise cluster:
  - batch parallel, batch serial, interactive, I/O nodes

Monitoring: Ganglia

- Designed for Grid computing
- Monitors each node
- Consolidate into groups of nodes -> groups of groups of nodes -> etc.
- Can use web interface to present status and “drill down” to isolate a problem

System management

- Centrally from master node:
  - Create and distribute operating system images
  - Reboot / shutdown
  - Power down / power up
- Reboot times:
  - 2 minutes per node, 8 minutes for the cluster
- Operating system installation downtime:
  - 15 minutes per node, same for cluster
- Timings should be independent of cluster size

- Cluster management software built on notion of "images" = complete operating systems
- Configuration change => Re-installation ?
    - Disruptive for users
    - Time consuming and expensive
    - Frequently needed: e.g.
        - To mount another file system
        - To change root passwords
- Not flexible enough: e.g
    - Different IO nodes serve different file systems - one image each ?

Maintenance and reliability

- Need support from different vendors:
    - Linux Networx for cluster hardware, MPI
    - Linux Networx, ClusterFS for Lustre
    - Linux Networx, SuSE for operating system
    - PathScale, Portland Group etc. for compilers
    - IBM for FAStT disk subsystem
    - The "Open Source Community" for software

-> can, and sometimes did, go from "pillar to post"

General problems — the easy ones

- Need highly available master node:
    - Nodes can run standalone (apart from Infiniband), but cannot re-boot if master is down: they download their kernel from master when rebooting
    - Disaster scenario:
        - General power cut
        - Master node does not reboot (e.g. system disk failed)
        - Whole cluster down

General problems — the difficult ones

- Compatibility of hardware components resulting in performance losses, e.g.
    - Concurrent IP traffic on Infiniband and Gigabit Ethernet
    - Data transfer rates to/from FAStT storage
- Finding out which vendor will take responsibility when things do not work at all / as designed / as desired

Potential problems with a very large cluster

- Evaluation of the small cluster did not reveal any obvious scalability issues.
- However ...
- Cumulative effects of software / hardware issues which do not surface on "small" clusters are possible;
- Other sites have reported size related issues that vendors could only reproduce and resolve on-site;
- May need large internal development / support team?
  e.g. 14 staff at LLNL (kernel, cluster tools, resource management, Lustre, operating system)
- Does the Infiniband design scale?

Preparing the cluster for use as General Purpose Server

- Support issues similar to current systems: Many different 3rd party products used
- No MPI requirements, Infiniband not critical
- Problems to solve:

- Get highly available master node
- Choose shared filesystem: Reliability, performance, site-wide accessibility ?
- User software — integration with Linux workstation environment, provide all that is available on (AIX) servers
- System administration
- Workload management: Interactive and batch, scheduled and very often ad-hoc
- Acceptable to users? — Very different from “traditional” server:
  - Where am I working?
  - Where is my job running?
  - Where is my output?
- -> Create environment where things are “taken care of”

Shared filesystems

FAStT via NFS

- Improved NFS access speeds by experimenting with
  - NFS export / mount parameters: NFS version 3, blocksize, ext2/ext3, udp, no_acl
  - Use Write Cache on FAStT
  - FAStT / LVM specific: Can failover devices between I/O nodes, if necessary, by changing ownership / preferred path and rescanning volume groups on new host
- But .. total throughput of FAStT is still below what it could be.

Lustre

- Scalable, but ..
- No easy way to grow file systems or add I/O nodes
- NFS exports not working yet according to Lustre representative => no access to data from outside the cluster, e.g. AIX servers, HPC
- No user quotas yet
- Difficult to have several filesystems on small set of I/O nodes
- No backup tools

Panasas

- Hardware & software solution
  - Based on shelves with blade servers, uses SATA disk drives, connected by Gigabit Ethernet
- Good performance
- Several filesystems ok
- User quotas promised
- 2 modes of access: NFS and Direct Flow client
- Supposed to scale for both modes of access, but not tested yet
- But:
  - Kernel dependencies for Direct Flow
    - Work within-cluster only
    - Do not co-exist with Lustre
  - Need to use NFS access from all other hosts

Performance test: Write 1 GB file (using dd)

Client	Target filesystem (server) Results are MB/s			
	/scratch (AIX server)	/FAStT	/panfs_nfs (Panasas via NFS)	/panfs (Panasas via direct flow)
AIX server (not /scratch)	10	20	18	N/A
Cluster node	2	37	50	83
Linux workstation	2	8.5	9	N/A
/scratch server (AIX)	20	*	*	*
/FAStT server	+	256	+	+

* Same as AIX server (not /scratch) + Same as cluster node

Performance test: Untar 480 GB/ 40500 files

Client	Target filesystem (server) Results are elapsed time			
	/scratch (AIX server)	/FAStT	/panfs_nfs (Panasas via NFS)	/panfs (Panasas via direct flow)
AIX server (not /scratch)	9m4.87s	9m59.23s	10m6.36s	N/A
Cluster node	6m33.52s	6m27.14s	5m28.48s	5m20.20s
Linux workstation	6m56.13s	6m59.62s	5m03.30s	N/A
/scratch server (AIX)	2m35.90s	*	*	*
/FAStT server	+	0m13.49s	+	+

Performance test: Delete complex directory structure (40500 files)

Client	Target filesystem (server) Results are elapsed time			
	/scratch (AIX server)	/FAStT	/panfs_nfs (Panasas via NFS)	/panfs (Panasas via direct flow)
AIX server (not /scratch)	2m56.74s	2m39.46s	1m20.00s	N/A
Cluster node	5m34.53s	1m36.88s	6m58.13s	0m25.88s
Linux workstation	2m15.42s	1m19.96s	1m19.95s	N/A
/scratch server (AIX)	1m39.11s	2m39.46s	1m20.00s	N/A
/FAStT server	N/A	0m02.27s	N/A	N/A

Workload management: Sun Grid Engine

- All required features, no problems so far, free -> keep to initial choice
- Configuration:
  - 3 types of node:
    - Interactive work
    - Batch work
    - Services (e.g. web server)
  - Reach batch nodes only through SGE / batch queues
  - Encourage interactive access only through interactive queues
  - Allow job submission from all systems, not just from within the cluster
    - Use SGE software on all Linux systems
    - Use wrappers on all others (but SGE versions for those are available)
- Availability:
  - Master and shadow master on 2 cluster nodes
  - Automatic failover between these 2 nodes
  - Easy to configure more master nodes:
    - List of hosts in config used at startup of SGE daemons
  - Define SGE host-groups, assign those to SGE queues:
    - Move work between nodes by changing host_group definition
    - Change is instant, no restart required
    - Useful for node failure and system session (e.g. OS upgrade, reboot, etc)

User software

- Goal: Provide same working environment as on workstations and servers.
- Problem:
  - Cluster nodes are 64-bit
  - Linux workstations are 32-bit
  - Compatibility ?
- Approach taken so far:

  - Build both versions, use the one appropriate for the architecture
  - Almost all software is available now, some still being worked on
  - 32 bit versions in general run on both workstations and cluster nodes (there may be OS-level dependencies though).

Compilers

- Initially, only PathScale compiled IFS without problems
- But:
  - IFS is not run routinely on general purpose servers (HPC systems are used for this).
  - All Linux workstations use Portland Group compilers -> use it on the cluster too, if possible.
  - Many Member States use Portland, not Pathscale.
  - Latest version of Portland also compiles IFS now.

Portland Compiler Evaluation

- Used RAPS8 IFS release for evaluation (IFS cycle 28R3)
- Portland version 5.2-4
  - Problems with unassociated/unallocated array sections passed as arguments on subroutine calls (3 routines had to be modified)
  - No other problems at -O0 (with no optimisation)
  - 4 routines produced incorrect results at -O3 optimisation
- Portland version 6.0
  - One runtime problem identified at -O0 (no optimisation)
    - Relating to pointer/target attribute
    - Workaround found and test case produced and submitted
- Portland compilers usable with no optimisation
- Reliability problems at high optimisation
  - All compilers have problems at high optimisation levels
  - Portland appears to have more than other compilers

Performance comparison
Portland v5.2 v PathScale v1.2

IFS runs on ECMWF linux cluster using 8 CPUs			
	Times in secs		
	pgf90	pgf90	pathf90
	-O0	-O3	-O2
T159 model	1136	682	639
T159 4D-Var	4360	2180	1968

Portland performance is less than PathScale but acceptable

System administration

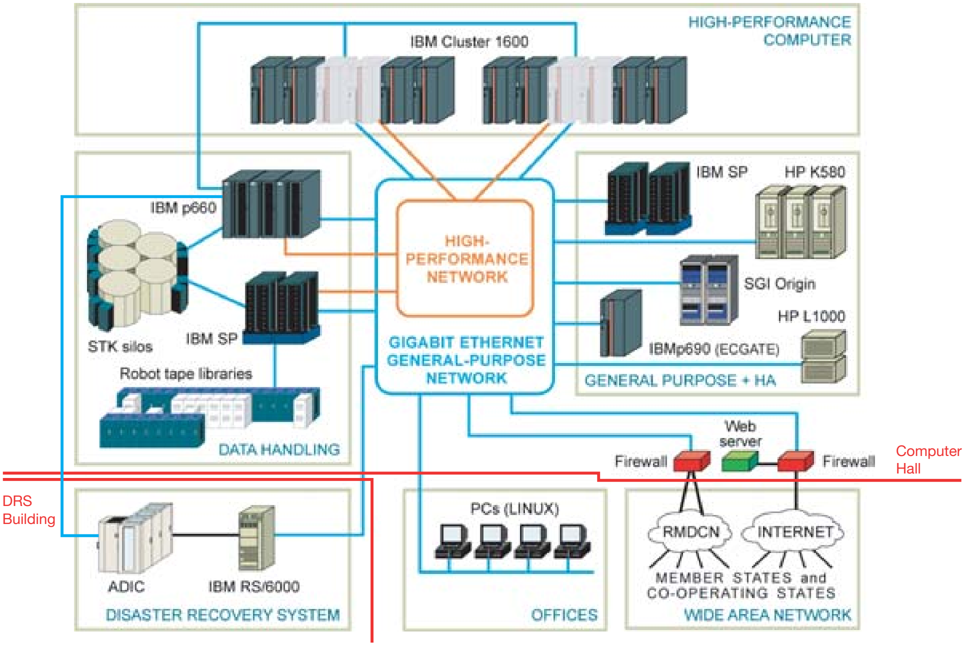
- Extend tools beyond “management by image”:
  - Added own software that “pulls” node specific configuration files at boot time and uses them.
  - Use image changes only for system changes, e.g. additional software installed from distribution.
  - Use same image on all nodes if possible.
  - Activate changes on the running system if possible, avoid reboots.

Current cluster use

- For some Operations Department tasks, in particular the production of charts for the web.
- Research Department have used it for one-off large tasks
- Not being used for day-to-day research work:
  - replacement for the verify package very close to completion, but not ready yet
  - performance figures show that using the AIX server for serving data to the cluster is not a good idea
  - need to finish evaluation of Panasas, decide whether to use this or FAStT, then move data, together with user work
- Substantial speed-up for tasks moved off from the AIX systems - typically 2 to 3 times faster (single cpu)

ECMWF Disaster Recovery Plans — *F. Dequenne*

Computer hall setup 2004



The former DRS

- The DRS building contained only:
  - Second copy of some ECFS and MARS data, partially stored in a robot.
  - Systems backup tapes
  - Tiny TSM server with
    - Backups of the critical DHS metadata
    - Backups of some servers' data (e.g. NFS servers, General Purpose servers..)

If the computer hall was lost...

- Super-Computers:
  - Require installation of new super-computers (months).
  - In the short term: find a site able to run our models for a while.
- Other servers:
  - Require the installation of new hardware (weeks), plus bare-metal restore from DRS backups.
- DHS:
  - The critical data would be saved, but no hardware to access it would have been available.
  - Require installation of new platforms (weeks), plus bare-metal restore of systems and metadata (HPSS, MARS, ECFS)
  - Never fully tested.

There was scope for improvement

- A disaster in the computer hall might have stopped ECMWF activities for weeks.
- In an ideal world:
  - Create an alternative site in another part of Europe.
  - Distribute or duplicate our equipment to this new site.

  - Duplicate all data to this site.
  - Install high speed links between the 2 sites.
  - But may be difficult to finance...
- How can we protect ourselves better, while keeping the costs under control?

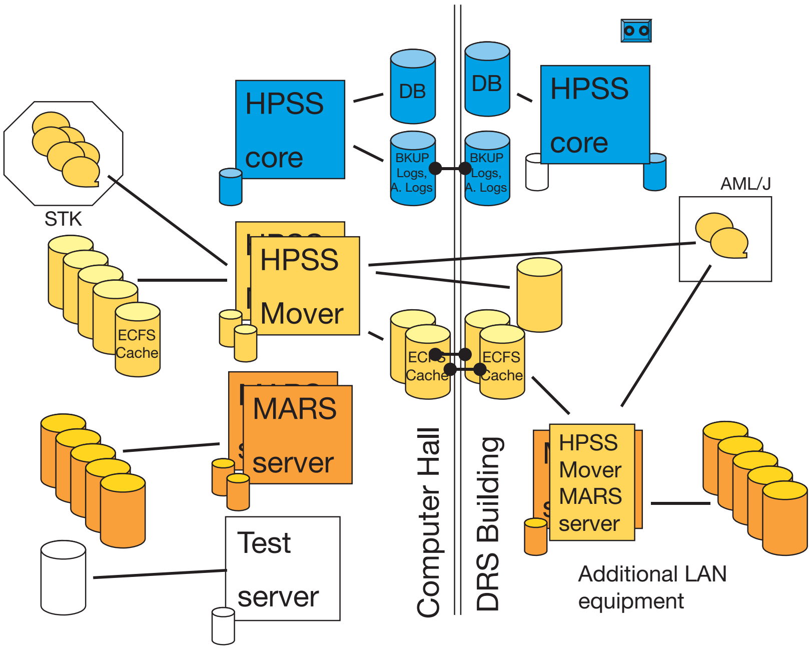
First step

- Weather Community is ready to help.
  - Following an NCEP disaster, NCEP operational workload was distributed to several sites.
  - When ECMWF's Cray C90 burned down, an alternative site was identified in a few hours (UK Met).
  - Finding alternative super-computer sites is possible.
- Make use of the second computer hall being planned.
  - Distribute equipment between the 2 halls.
  - Increase the chance that part of our equipment would survive a disaster.
- First priority:
  - How to provide access to the required data?

What we wanted to achieve (DHS)

- Provide access quickly to the DHS data stored in the DRS building.
  - Critical data could be exported to external sites.
  - Data could be provided to unaffected equipment onsite.
  - Transfer data to other sites:
    - By tape.
    - Possibly in the future by connection of the DRS equipment to the WAN.
- Provide a minimal DHS service to support unaffected equipment.
- Test regularly that a service can be restored.
- Costs have to be kept low.

New layout (DHS)

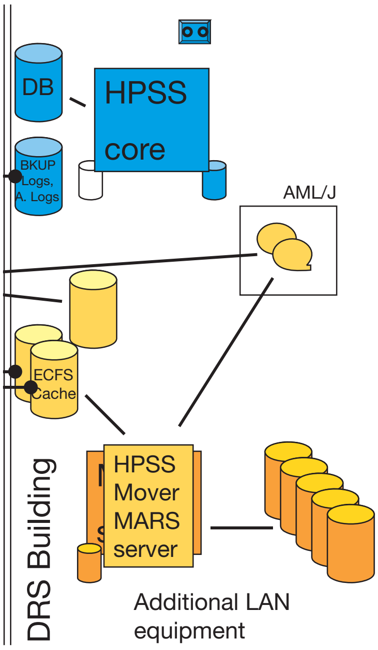


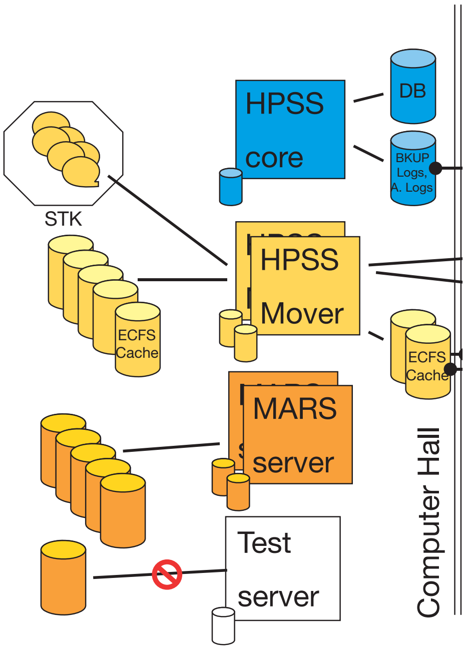
Time to recover: 4 to 5 hours.

Data lost:
- Old data which is not backed up.

In particular:
- MARS RD
- ECFS data without backup;
- Recent data not yet copied to DRS tapes.

The service is expected to be very limited:
- 12 tape drives only
- small disk cache
- limited CPU resources





The only affected service is one MARS server.
It will be restored on one of the surviving MARS server platforms.

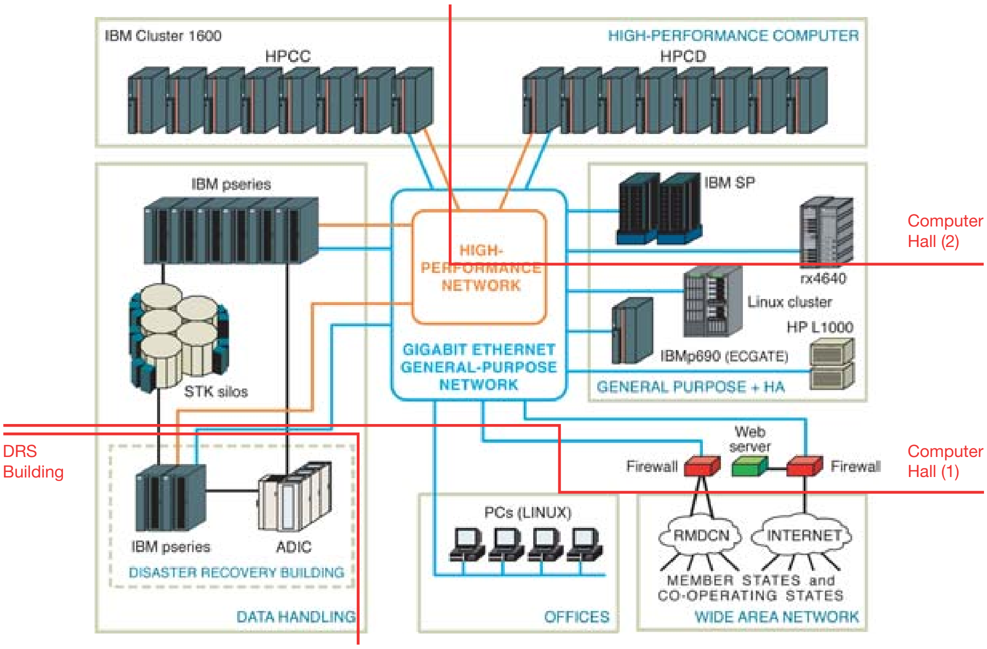
Data lost: anything from that server which was not yet copied to tape; ECFS data which was on un-mirrored disks in the DRS building and not copied to tape.

Service will be affected to some extent.

Current Status (DHS)

- First large scale test was performed in April.
- We still need to:
  - Resolve some issues discovered during previous tests;
  - Test the restoration of one MARS server in the computer hall;
  - Evaluate the management of some end cases;
  - Introduce a regular testing schedule (twice a year?).
- We are reasonably confident that we would be able to provide a service after a computer hall loss.

Computer hall setup



Protection of non-DHS servers: Short term

- Install supercomputer clusters in different computer halls.
- Other servers
  - Work has started on confirming whether some critical workload can be moved between various servers. (e.g. nfs service)
  - These servers could then be distributed between the 2 computer halls.
- Resilient LAN connections between DRS building and both computer halls.
- Split of telecoms area.
- These proposals are under investigation, no decisions have been taken yet.

In the future:

- Static subsets of popular data could be distributed to other sites
  - Already done for ERA-40 data.
- ECMWF may investigate the ability to distribute a minimal subset of data geographically.
  - This may require additional bandwidth.
- Consider an alternative WAN connection to the DRS building.
- Distribute DHS equipment across computer halls.
- Consider extending or replacing the Disaster Recovery Building.
- An Integrated Disaster Recovery Action plan will be designed.

M. Pithon asked how far away from the main building the Disaster Recovery building was. F. Dequenne replied that it was approximately 50 metres away and was supposed to be built in such a way as to withstand any major incident in the main building.

C. Hammerschmid asked for more information on duplicated networking equipment. F. Dequenne replied that 2 small Gigabit routers were the only networking equipment that had to be acquired to run the disaster recovery system.

User Registration, Update on the interface — *Paul Dando*

Concepts

- New system: EMS = Entity Management System
  - database used to store and define user access rights
- Entities:
  - users, applications, web domains
- Policies:
  - rules that define access rights
- EMS database contains two core data sets
  - user and organisation data
  - Policies (maintained by ECMWF)
- Registrator:
  - the person performing the registration

Underlying principles

- Based on a concept of access rights
- Rules defining access rights are called "Policies"
- Registrator decides which policies should be applied to a user
- Policies are based on:
  1. User's employer (National Met Service, University, ECMWF, WMO, etc)
  2. Projects the user works on (e.g., Special Projects)
- Access rights can be:
  - Default - assigned to all holders of the policy
  - Additional requirements — assigned on a case-by-case basis

Advantages of EMS

- Easy to use, web-based interface for user registration
- Provides a flexible, consistent & co-ordinated approach
- Fast turnaround:
  - Can register users and supply them with a spare SecurID card
  - User should be able to start working within ~30 minutes
- More guidance:
  - Registration pages created dynamically
  - Input on first page defines options available on following pages
- Easier to modify user info and access rights
  - e.g., can grant or deny access to current forecast data, hpcd, etc
- On-line query of user info and access rights
  - Up-to-date information obtainable directly from the EMS database

Range of possible actions

- System can be used to register:
  - Member State or Special Project users with host login access to ECMWF computing systems (e.g., access MARS, ecgate, hpcd)
  - Users with web-only access
- Modify or query personal details or access rights for existing users

- Comp Reps CANNOT use system (yet !) to:
  - deregister users
  - register or delete Special Projects
  - register new Section Identifiers
  - change user quotas

Please contact User Support (advisory@ecmwf.int) for these cases

Logging in to the EMS and security

- First log on to the ECMWF web site at: **http://www.ecmwf.int/login/**
- For security reasons:
  - You MUST login using your SecurID passcode
  - login expires after 1 hour of inactivity
  - a logout button is provided on each screen so that the registrator can log off the system at any time
- Access is limited strictly to those persons authorised by ECMWF
- All access to the system is logged in the EMS logs

Main registration menu

- Accessed at: **http://www.ecmwf.int/services/ems/d/registration/**

Three options are available:

- Entity management
  - to register new users
  - to query or modify info or access rights for existing users
- Organisation management
  - to add new or modify existing employer/organisation information
- Registration Guide
  - to access an up-to-date version of the documentation

Web access classes — authorised domain

User class	Auth Method	Browse MARS data	Retrieve Data		Your Room	Real-time charts	Restricted Computing Docs
			Archived	Real-time			
Unregistered	Domain	✓	✗	✗	✗	✓	✓
Self registered	Domain + web password	✓	✓	✗	✓	✓	✓
Registered by Comp Rep	Domain + web password	✓	✓	🧑‍💻	✓	✓	✓
	Roaming password	✓	✓	🧑‍💻	✓	✓	✓
	SecurID	✓	✓	🧑‍💻	✓	✓	✓

Web access classes — other domains

User class	Auth Method	Browse MARS data	Retrieve Data		Your Room	Real-time charts	Restricted Computing Docs
			Archived	Real-time			
unregistered	Not applicable						
Self registered	Not applicable						
Registered by Comp Rep	web password	Insufficient privilege					
	Roaming password	✓	✓	🧑‍💻	✓	🧑‍💻	✓
	SecurID	✓	✓	🧑‍💻	✓	🧑‍💻	✓

Paperwork

- User registration forms can still be used in parallel
  - If Comp Rep uses EMS to register users, there's no need to use a form
  - Forms should be sent to User Support, if you want ECMWF to register the users.
- Current registration forms will be changed to reflect new "policy based" system.
- Users will still need to contact authorising organisation
  - Access authorised by Comp Reps (as at present).
- Users still need to sign the SecurID declaration.
- New User Packs
  - Once the system is active, all information will be sent by e-mail or made available electronically.

Possible future developments

- Web-based registration for users
  - Users complete on-line registration forms.
  - Computing Representatives authorise registration and assign access rights via EMS interface.
  - No more paper forms!
- On-line acceptance of "ECMWF Terms and Conditions"

Availability

- Core system operational since December 2003
  - Used by Calldesk and User Support for all user registrations since then.
- Already being tried out by two Member State Comp Reps
  - Thank you to Hans and Roddy!
- Available for use in the next few weeks.

BUT...

- Use is NOT compulsory:
  - Comp Reps can still send the registration forms to ECMWF

H. Bjornsson asked what facilities roaming passwords would provide access to. U. Modigliani replied that they would provide access to most web services, including data retrievals via WebMARS, but would not allow general access to ECMWF computer systems for job submission. P. Dando added that roaming passwords would require regular renewal, although the expiry period had not yet been decided.

M. Pithon asked whether registrators would be able to register new projects. P. Dando replied that this function would remain with ECMWF, as projects are regarded as policies. I. Weger suggested that it might be possible to introduce web based forms for new project registration.

Results of the survey of external users — *Carsten Maaß*

Background

The survey of all registered external computer users had the following aims:

- Determine the level of user satisfaction with the computing services provided
- Identify issues of current concern
- Gather quantitative and qualitative data
- Improve the service provided

Response

An invitation to take part in the online survey was sent to 1267 registered users

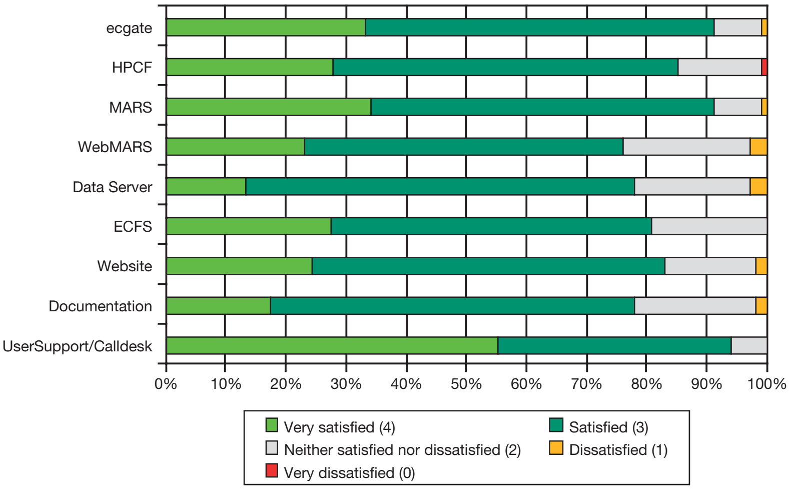
468 (37%) provided very useful and detailed answers

719 (57%) did not respond

Full information from the responses received has been published at: http://www.ecmwf.int/services/computing/survey/

The following gives highlights.

Overall satisfaction (active users only)



Ecgate

Comments:

- Performance has improved

Problems mentioned:

- X-connection (time-out, lost connections)
- Environment (shell)
- Bandwidth between user’s machine and ecgate
- Disk space
- Slow (probably refers to MARS)
- SecurID cards

HPCF

Reasons for not using HPCF:

- No need
- Easier access to supercomputer at own organisation

- Lack of training
- Porting
- Not allowed to access

HPCF

Problems mentioned:

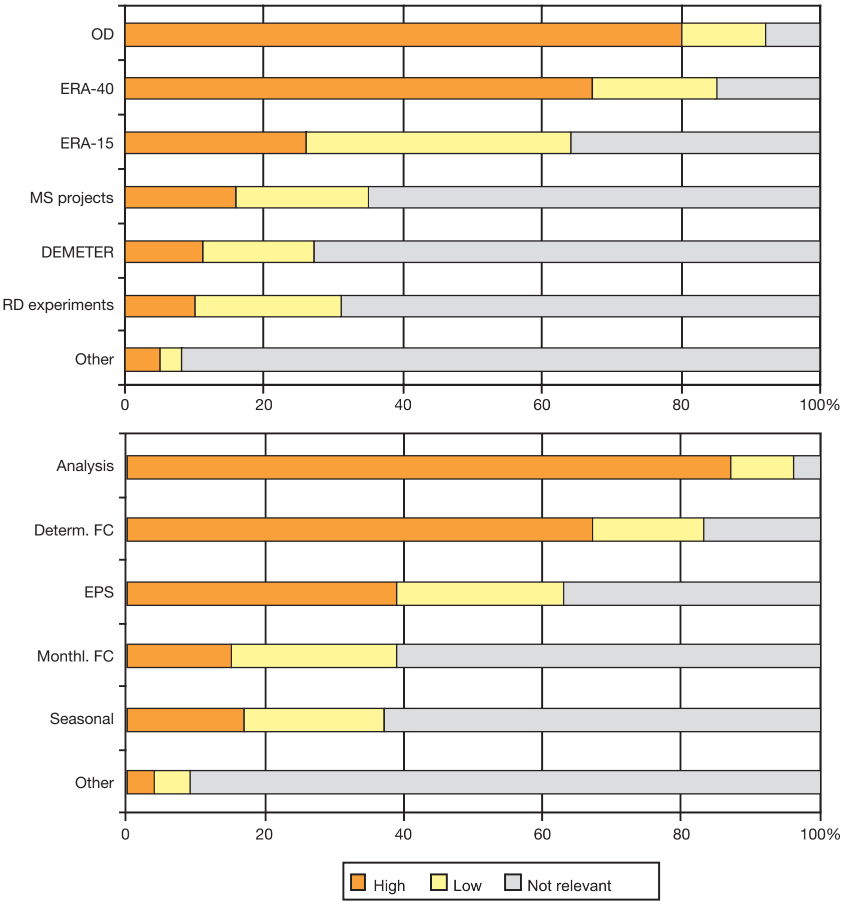
- Disk space
- Data transfer to local system
- Scheduling of very long jobs
- Restriction to ksh
- Users from Co-operating States would like to have access

MARS

Comments:

- Too slow
- Easier interface / query language
- Limited post-processing: interpolation, vertical profiles, GRIB header, formats (NetCDF, HDF, ASCII)
- Poor documentation
- Error messages not clear enough
- Observations and satellite data difficult to access

MARS data usage



WebMARS

61% have used WebMARS at least occasionally, of which the majority are satisfied (53%) or very satisfied (23%)

Users would like to have:

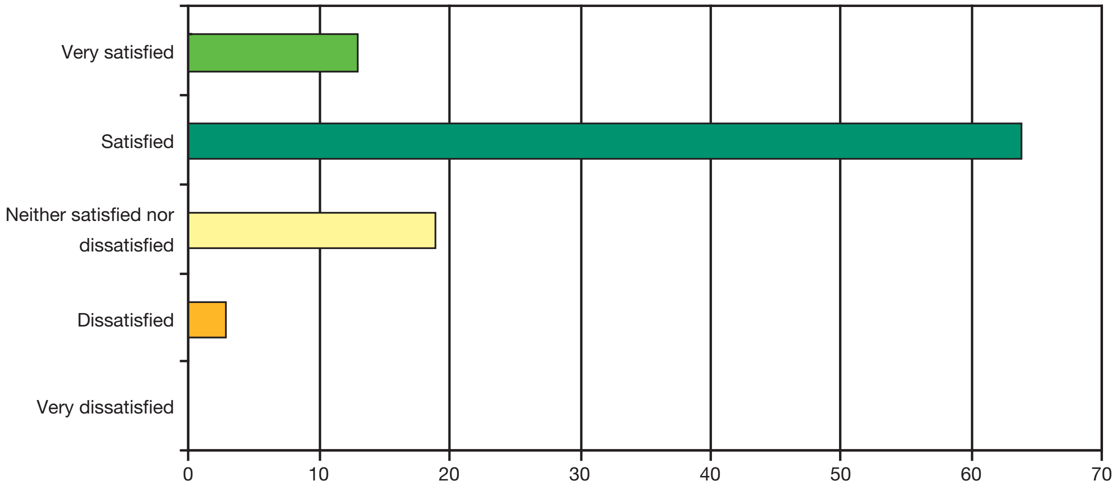
- Better (meta data) documentation
- More flexible graphical tools
- Timely information.

Reasons why users haven't used WebMARS yet:

- No need
- Did not know about it (26 users)
- Prefer traditional request.

Data Server

23% of users have used the data server; of those:



ECFS

51% of users use ECFS at least occasionally. In this group:

- majority are satisfied (53%) or very satisfied (27%) with the service
- 74% find ECFS easy or very easy to use.

Comments

- Slow
- Move command and usage of wildcards missing
- Audit file was useful.

Website

- 99% of users use the website.
- Majority find it, overall, useful (55%) or very useful (41%).
- Satisfaction with various characteristics is below 90%.

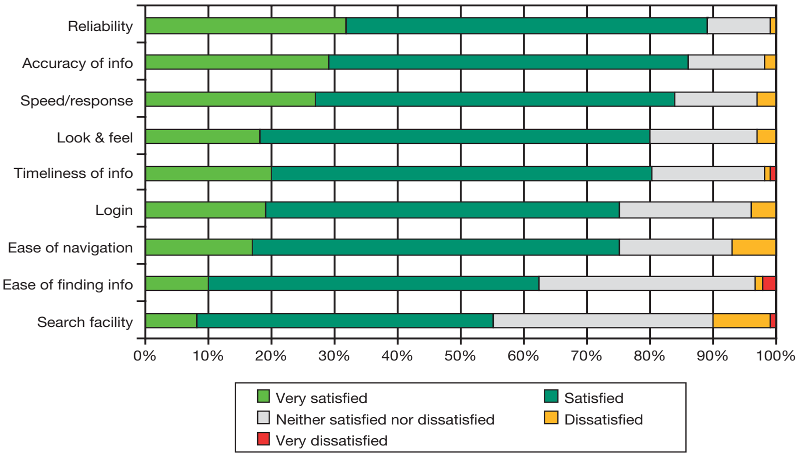
Frequent comment:

- Information is difficult to find.

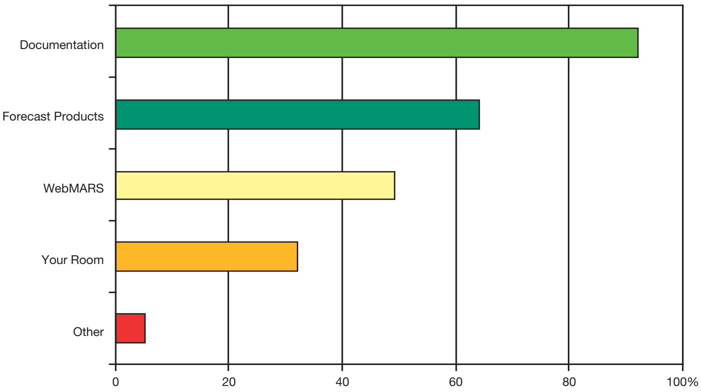
Users would visit the Website more often if it provided:

- Easy access to more (real time, short range) forecast products
- "What's new".

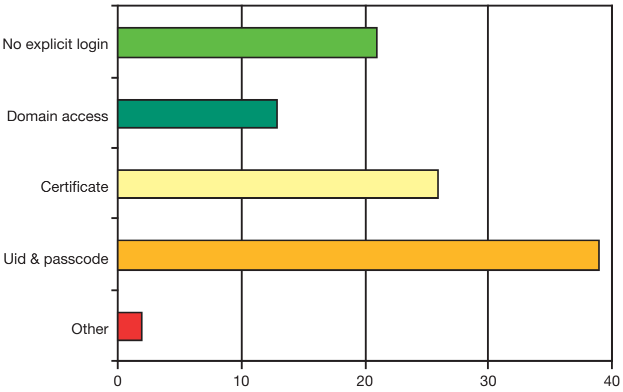
Website — Satisfaction with some characteristics



Which areas are accessed?



Login



75% of users are satisfied with the login.

Users use more than one method.

Comments

- Lack of documentation explaining the login
- Login status not clear
- Problems with login.

Documentation

77% of users are satisfied or very satisfied with both on-line and paper based documentation.

Users would mainly like to see the following documents improved:

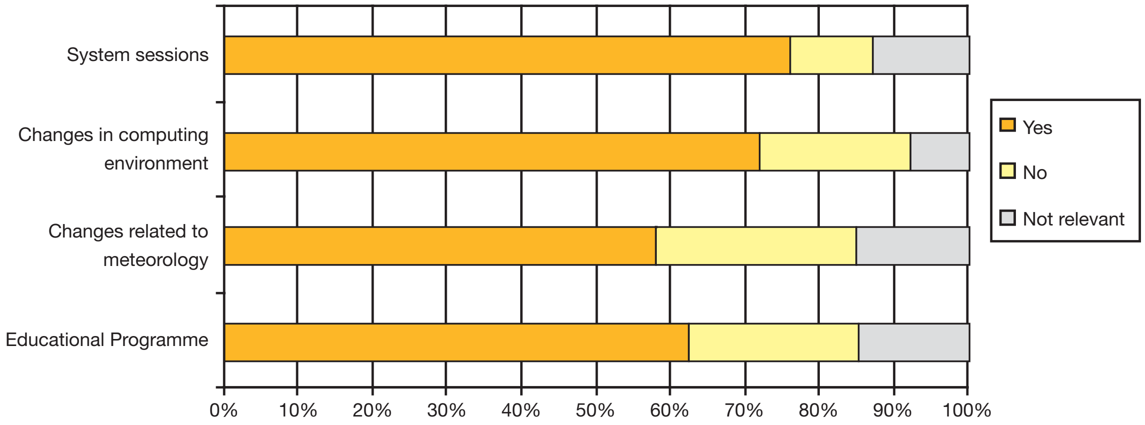
- MARS
- Emoslib and GRIB decoding
- (Prep)IFS
- Model skill.

User Services

- 74% of users contact User Support/Calldesk at least occasionally.
- 98% found the services provided by User Support/Calldesk helpful (33%) or very helpful (65%).
- In case of problems, users contact:
  1. User Support
  2. Colleague
  3. Computing Representative
  4. Call desk
  5. ECMWF expert.
- 84% prefer email to telephone.
- Advice in their own language is important to 30% of users.

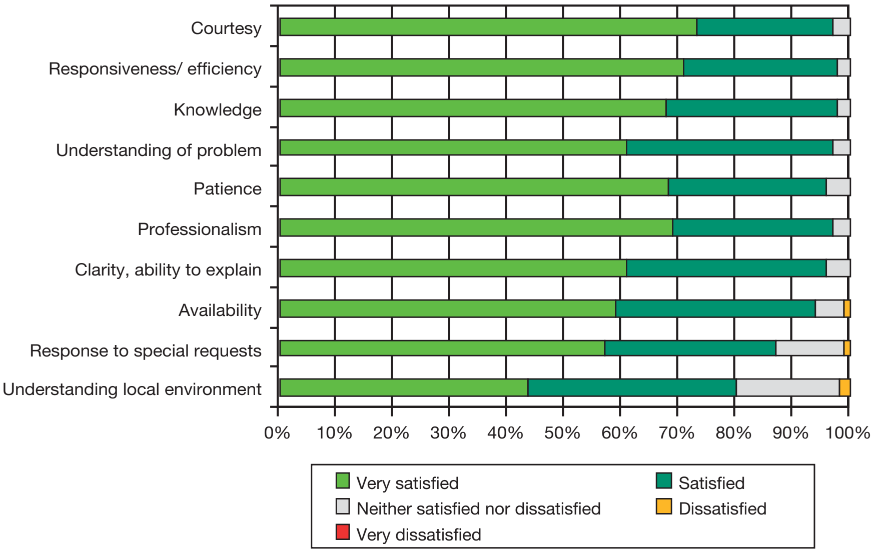
User Services — Flow of information

Do you think you are adequately informed?



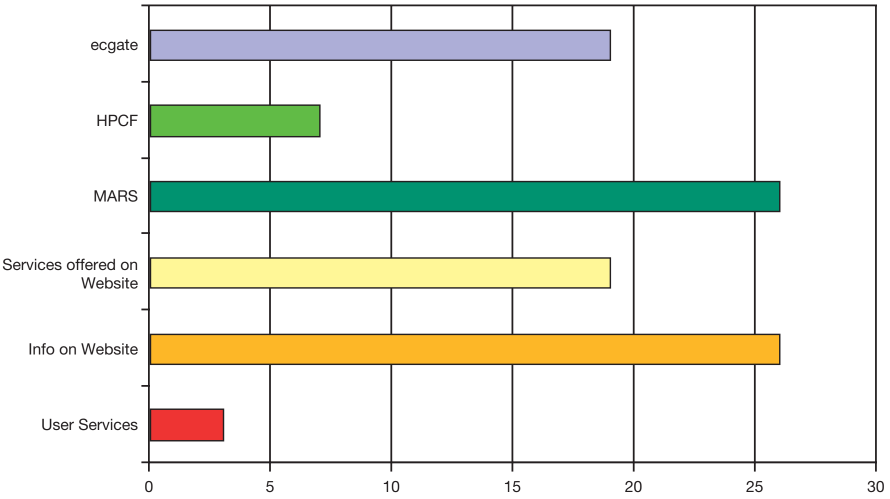
- Users outside Meteorological Services feel less informed
- Frequent comment: mailing lists

User Services — Satisfaction



Areas for improvement

Users suggested the following areas should be improved:



General user suggestions

How can ECMWF improve?

- Abolish SecurID card
- Make software (Metview, GRIB/BUFR) available under GNU licence
- Offer mailing lists/FAQs
- Training in MS open to users outside met. services
- Change data policy.

What would improve users' productivity?

- More disk space
- Tools to convert GRIB to other formats
- Faster MARS
- More bandwidth between MS and ECMWF.

M. Pithon commented that MétéoFrance users had not received the survey at all.

C. Maass thought that the most likely reason was that the mails had been filtered out by their system as suspected Spam. U. Modigliani added that the mails to Denmark had been bounced back, so could be resent; this was the only problem ECMWF had been aware of. The lower than expected response from France had been noticed but as there was also the possibility of anonymous reply, this was difficult to follow up. I Weger encouraged M. Pithon to invite French users to send any particular comments they might wish to make after the meeting. They could still be added to the final summary report.

ANY OTHER BUSINESS

Mailing Lists

R. Rudsar noted that there were still only a few States using SMS and wondered whether there was any interest in setting up a Mailing List for the exchange of information on and comparing experience in using SMS. The representatives from France, Norway, Romania, Germany, Spain and CTBTO expressed their interest.

U. Modigliani noted that this interest would be taken into account when the mailing lists were set up and added that as well as a general list for announcements etc, specialist lists for such topics as Magics and Metview were planned.

NEXT MEETING

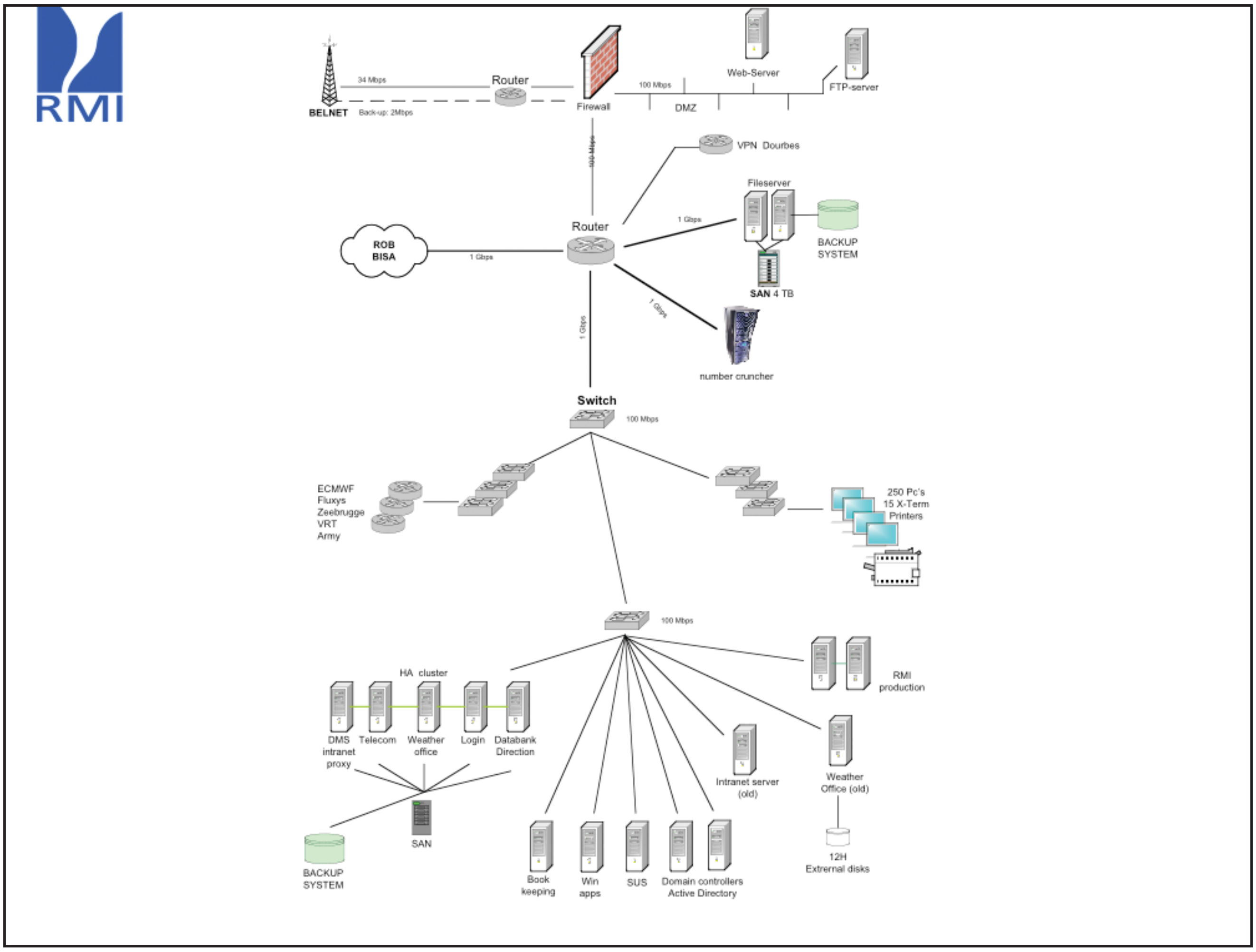
There was strong support for having the next meeting in spring 2006.

PART II

Member States' and Cooperating States' Presentations

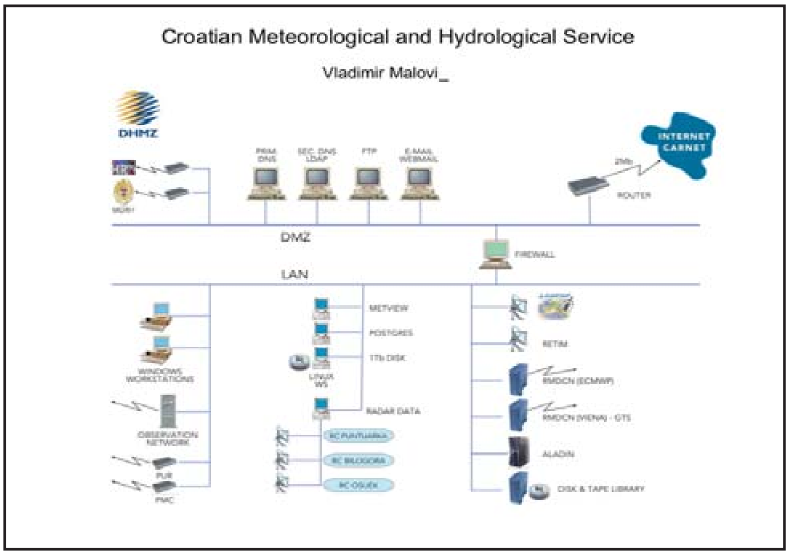
BELGIUM BELGIUM

Liliane Frappez – Royal Meteorological Institute, Brussels



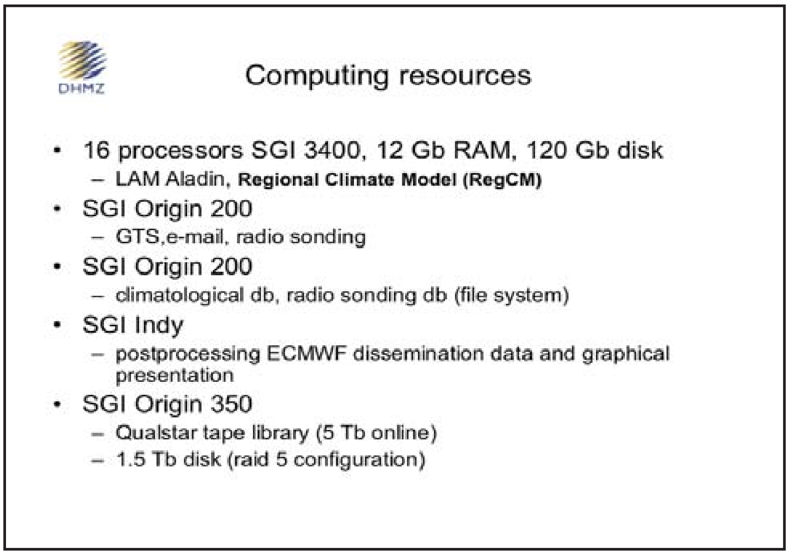
CROATIA CROATIA

Vladomir Malović – Croatian Meteorological and Hydrological Service



Croatian Meteorological and Hydrological Service

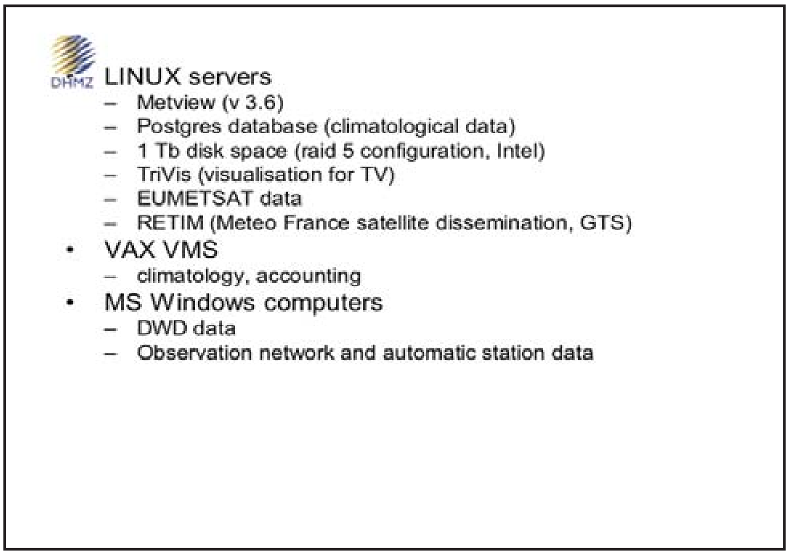
Vladimir Malovi_



DHMZ

Computing resources

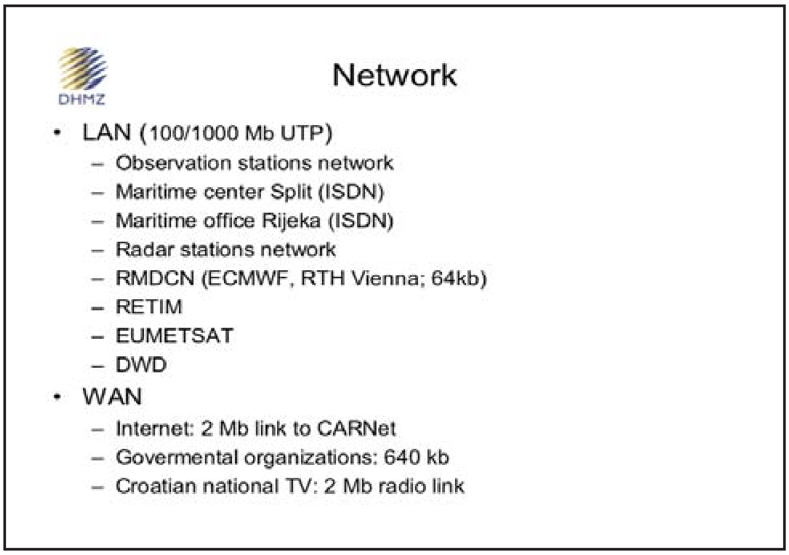
- 16 processors SGI 3400, 12 Gb RAM, 120 Gb disk
  - LAM Aladin, **Regional Climate Model (RegCM)**
- SGI Origin 200
  - GTS,e-mail, radio sonding
- SGI Origin 200
  - climatological db, radio sonding db (file system)
- SGI Indy
  - postprocessing ECMWF dissemination data and graphical presentation
- SGI Origin 350
  - Qualstar tape library (5 Tb online)
  - 1.5 Tb disk (raid 5 configuration)



DHMZ LINUX servers

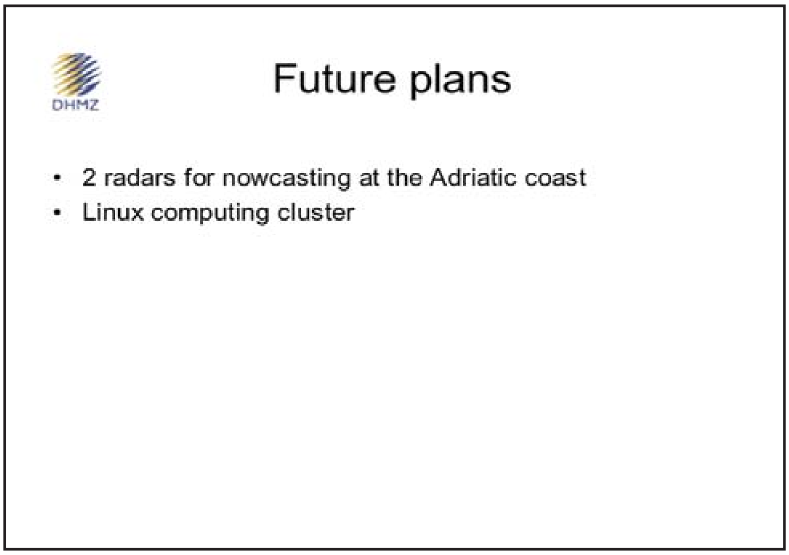
- Metview (v 3.6)
- Postgres database (climatological data)
- 1 Tb disk space (raid 5 configuration, Intel)
- TriVis (visualisation for TV)
- EUMETSAT data
- RETIM (Meteo France satellite dissemination, GTS)

- VAX VMS
  - climatology, accounting
- MS Windows computers
  - DWD data
  - Observation network and automatic station data



Network

- LAN (100/1000 Mb UTP)
  - Observation stations network
  - Maritime center Split (ISDN)
  - Maritime office Rijeka (ISDN)
  - Radar stations network
  - RMDCN (ECMWF, RTH Vienna; 64kb)
  - RETIM
  - EUMETSAT
  - DWD
- WAN
  - Internet: 2 Mb link to CARNet
  - Govermental organizations: 640 kb
  - Croatian national TV: 2 Mb radio link

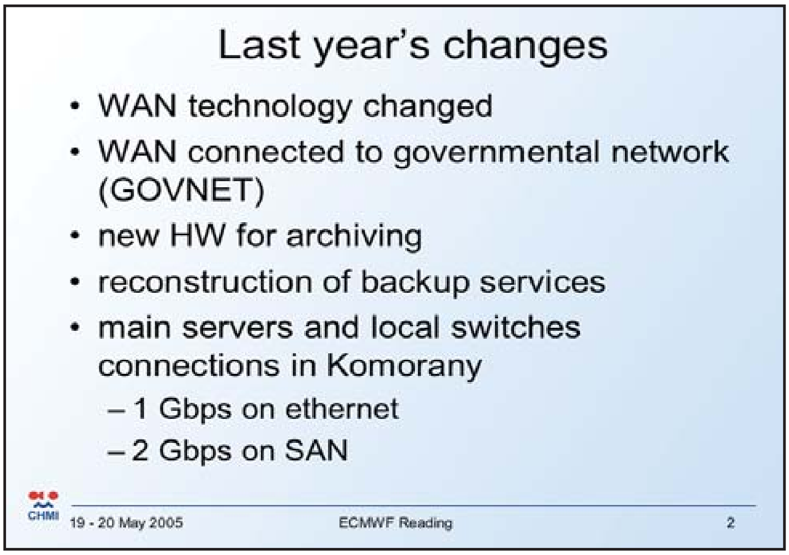


Future plans

- 2 radars for nowcasting at the Adriatic coast
- Linux computing cluster

CZECH REPUBLIC

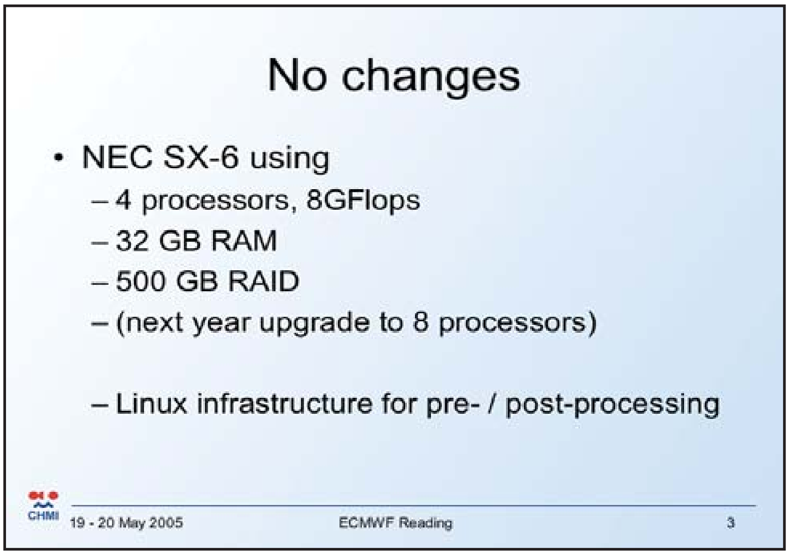
Karel Ostatnicky, Karel Pesata – Czech Hydrometeorological Institute



Last year’s changes

- WAN technology changed
- WAN connected to governmental network (GOVNET)
- new HW for archiving
- reconstruction of backup services
- main servers and local switches connections in Komorany
  - 1 Gbps on ethernet
  - 2 Gbps on SAN

CHMI 19 - 20 May 2005 ECMWF Reading 2

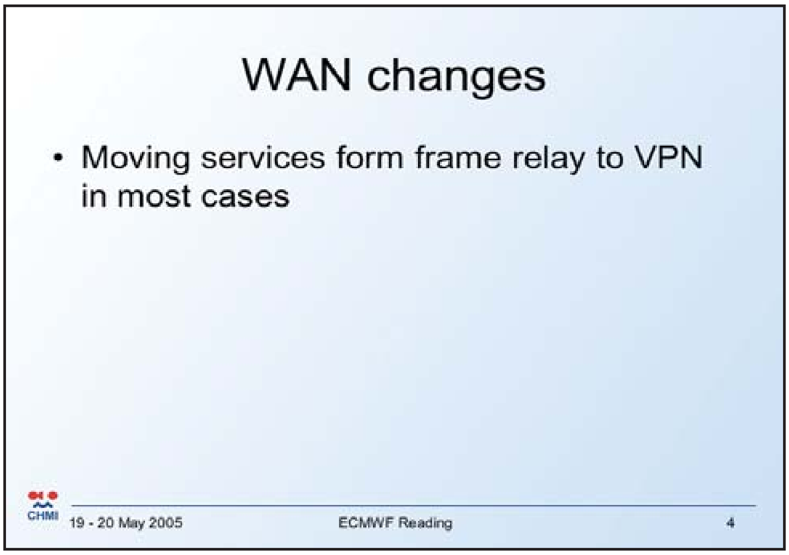


No changes

- NEC SX-6 using
  - 4 processors, 8GFlops
  - 32 GB RAM
  - 500 GB RAID
  - (next year upgrade to 8 processors)

  - Linux infrastructure for pre- / post-processing

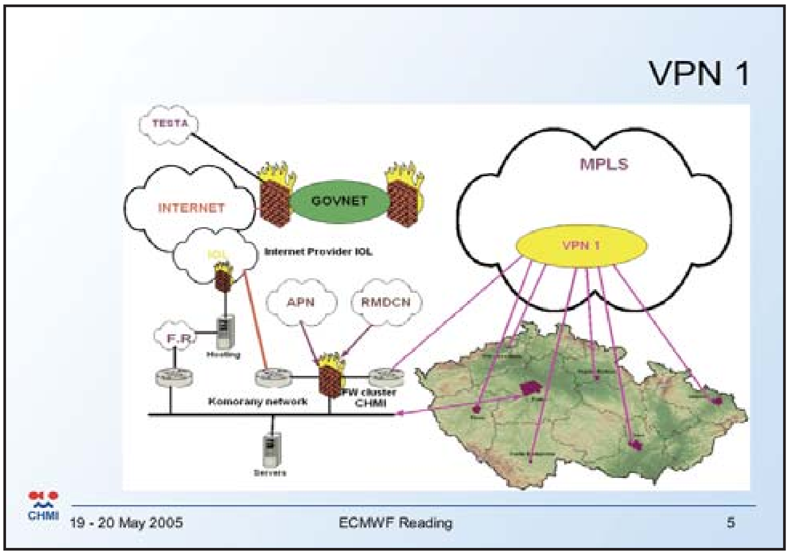
CHMI 19 - 20 May 2005 ECMWF Reading 3



WAN changes

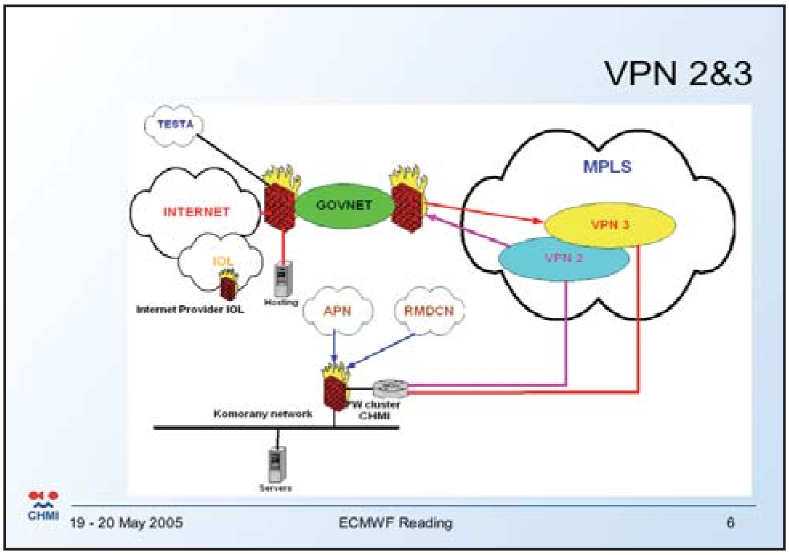
- Moving services form frame relay to VPN in most cases

CHMI 19 - 20 May 2005 ECMWF Reading 4



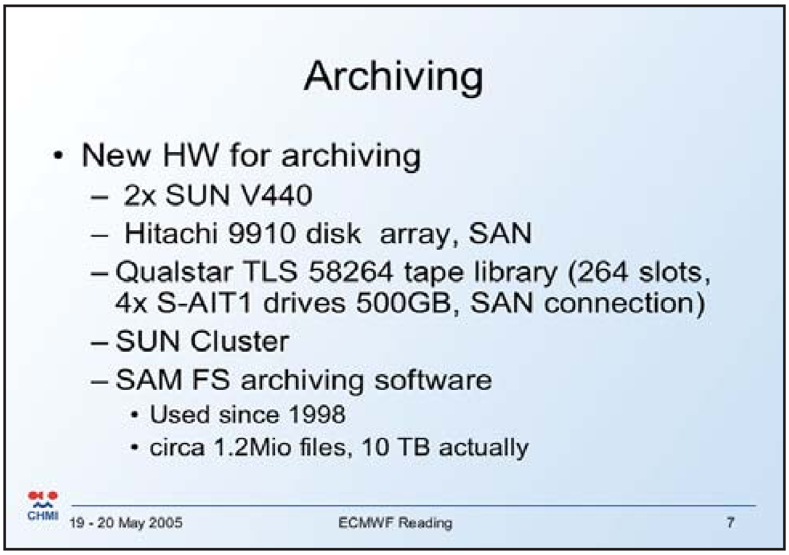
VPN 1

CHMI 19 - 20 May 2005 ECMWF Reading 5



VPN 2&3

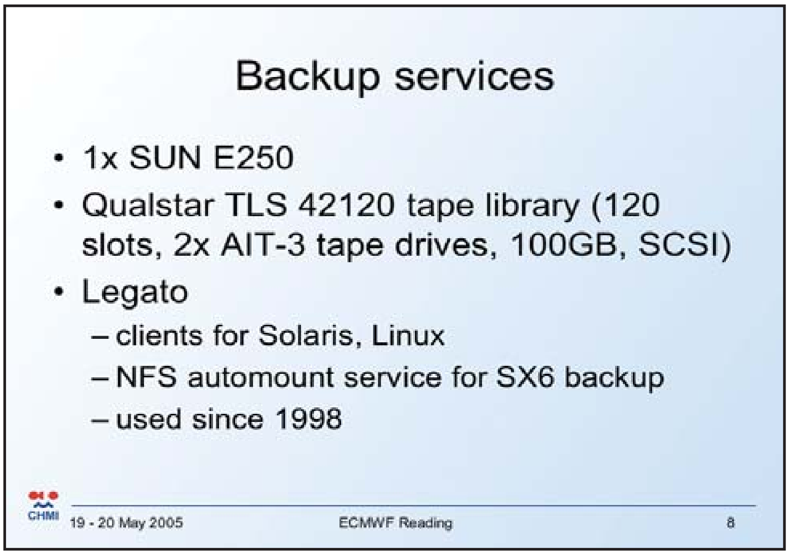
CHMI 19 - 20 May 2005 ECMWF Reading 6



Archiving

- New HW for archiving
  - 2x SUN V440
  - Hitachi 9910 disk array, SAN
  - Qualstar TLS 58264 tape library (264 slots, 4x S-AIT1 drives 500GB, SAN connection)
  - SUN Cluster
  - SAM FS archiving software
    - Used since 1998
    - circa 1.2Mio files, 10 TB actually

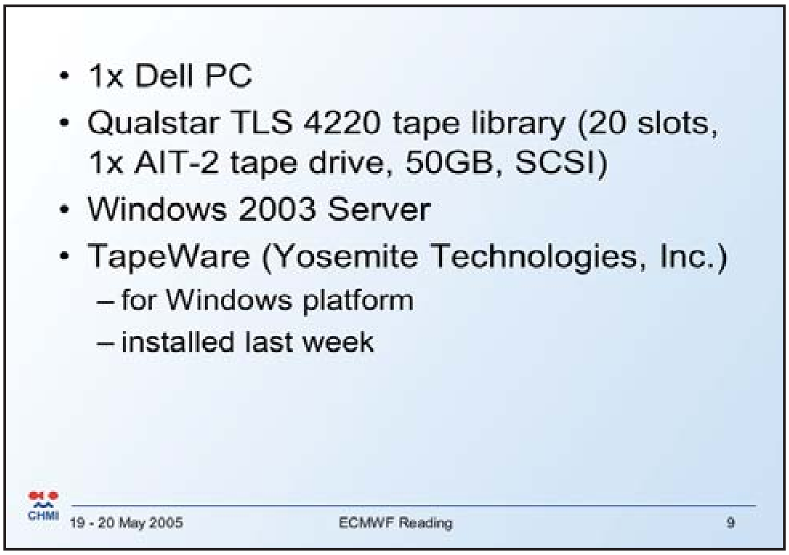
CHMI 19 - 20 May 2005 ECMWF Reading 7



Backup services

- 1x SUN E250
- Qualstar TLS 42120 tape library (120 slots, 2x AIT-3 tape drives, 100GB, SCSI)
- Legato
  - clients for Solaris, Linux
  - NFS automount service for SX6 backup
  - used since 1998

CHMI 19 - 20 May 2005 ECMWF Reading 8

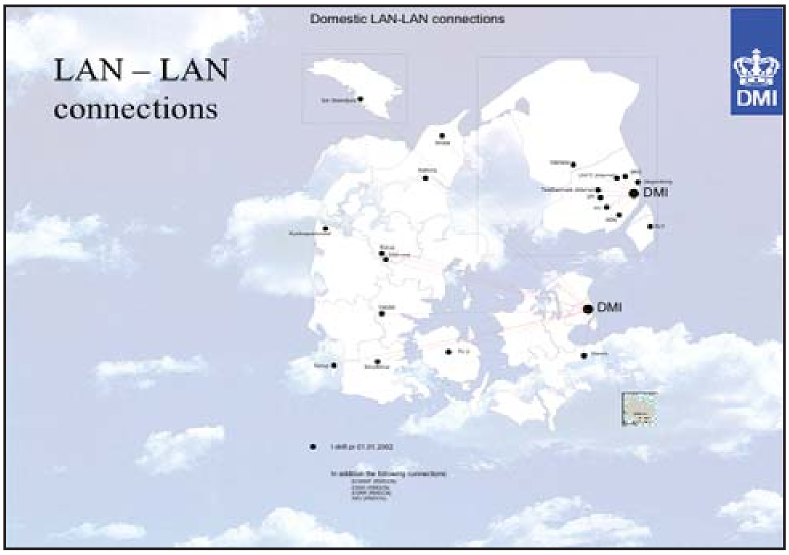


- 1x Dell PC
- Qualstar TLS 4220 tape library (20 slots, 1x AIT-2 tape drive, 50GB, SCSI)
- Windows 2003 Server
- TapeWare (Yosemite Technologies, Inc.)
  - for Windows platform
  - installed last week

CHMI 19 - 20 May 2005 ECMWF Reading 9

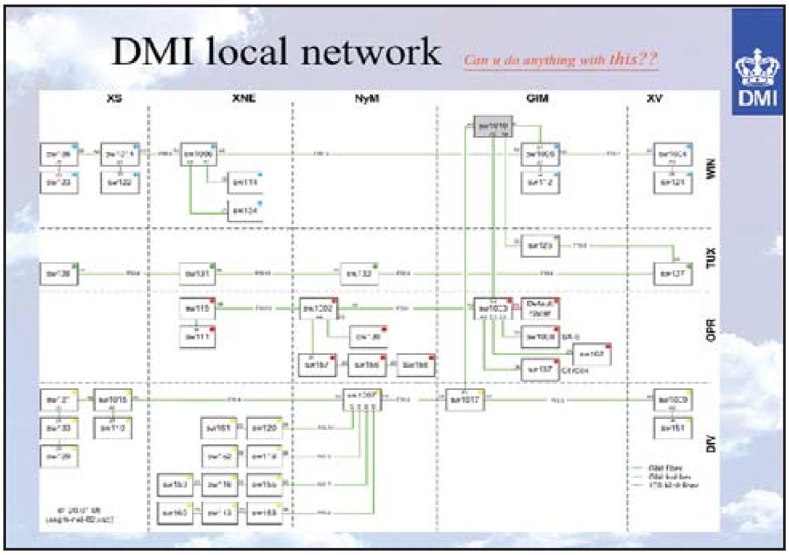
DENMARK DENMARK

Niels Olsen – Danish Meteorological Institute



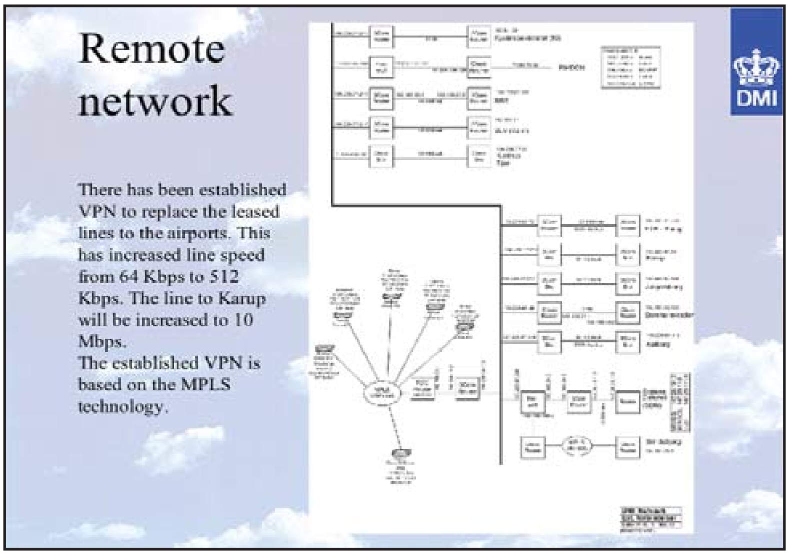
LAN – LAN connections

Domestic LAN-LAN connections



DMI local network

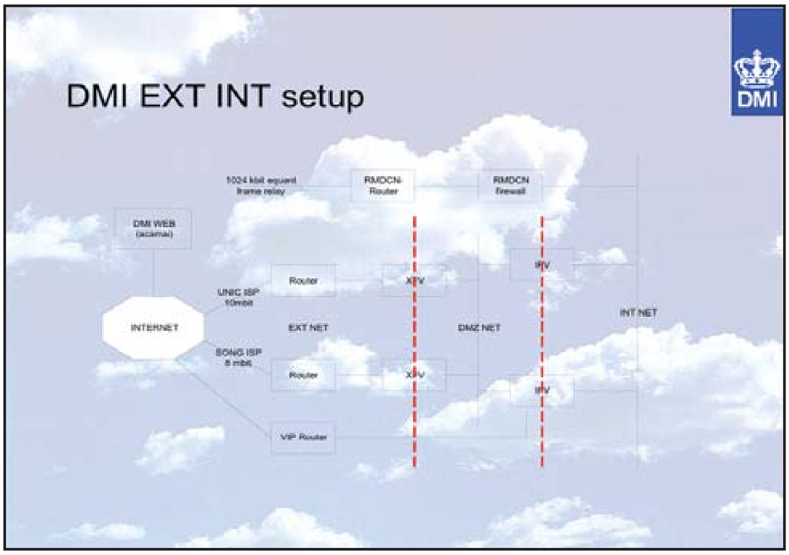
Can u do anything with this??



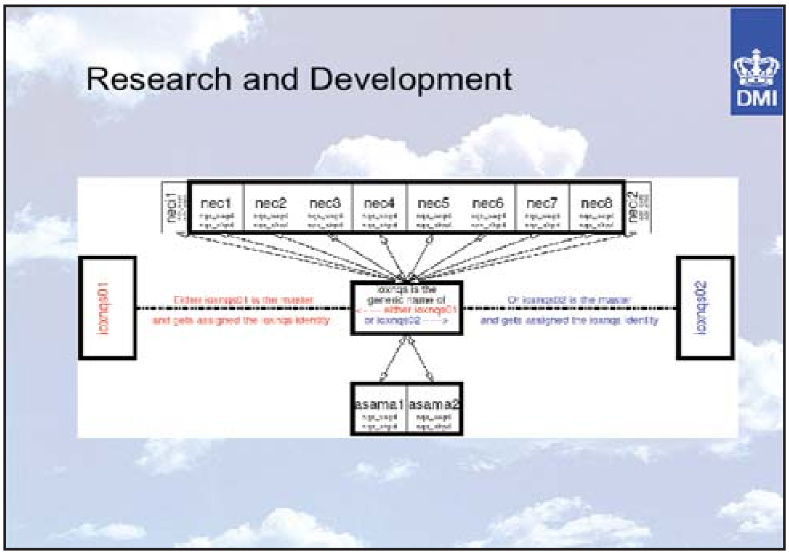
Remote network

There has been established VPN to replace the leased lines to the airports. This has increased line speed from 64 Kbps to 512 Kbps. The line to Karup will be increased to 10 Mbps.
The established VPN is based on the MPLS technology.

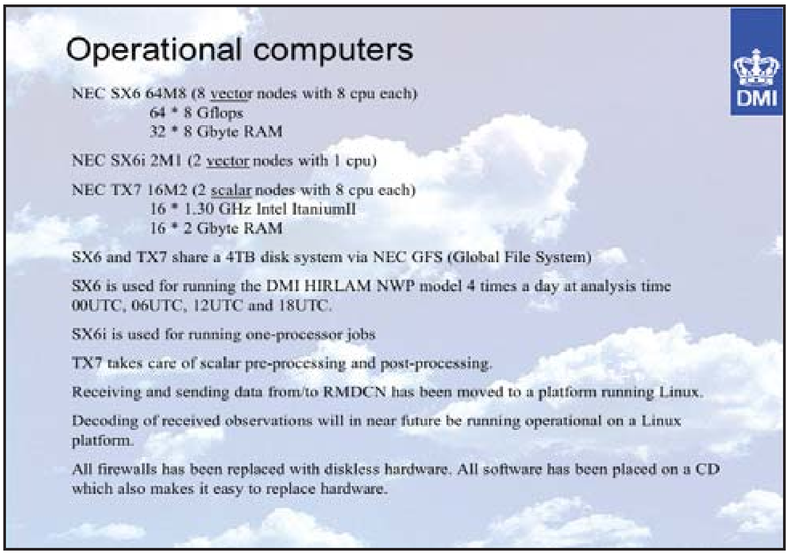
DENMARK DENMARK



DMI EXT INT setup



Research and Development



Operational computers

NEC SX6 64M8 (8 vector nodes with 8 cpu each)
 64 * 8 Gflops
 32 * 8 Gbyte RAM

NEC SX6i 2M1 (2 vector nodes with 1 cpu)

NEC TX7 16M2 (2 scalar nodes with 8 cpu each)
 16 * 1.30 GHz Intel ItaniumII
 16 * 2 Gbyte RAM

SX6 and TX7 share a 4TB disk system via NEC GFS (Global File System)

SX6 is used for running the DMI HIRLAM NWP model 4 times a day at analysis time 00UTC, 06UTC, 12UTC and 18UTC.

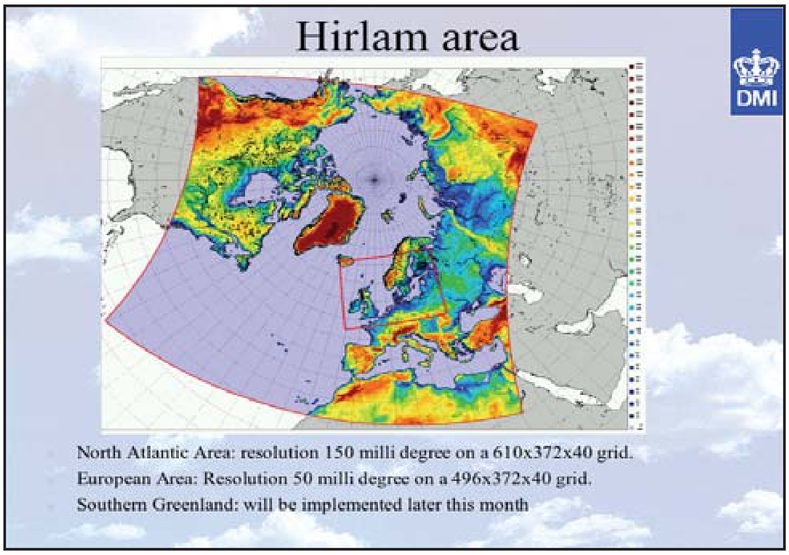
SX6i is used for running one-processor jobs

TX7 takes care of scalar pre-processing and post-processing.

Receiving and sending data from/to RMDCN has been moved to a platform running Linux.

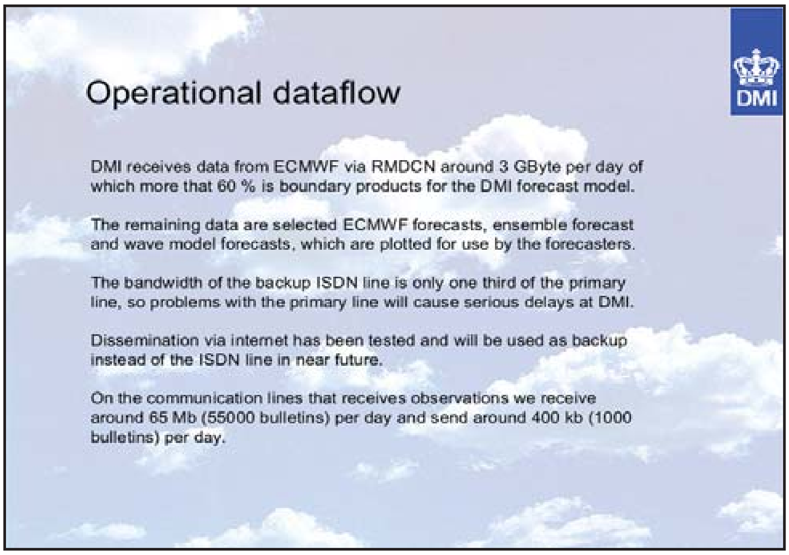
Decoding of received observations will in near future be running operational on a Linux platform.

All firewalls has been replaced with diskless hardware. All software has been placed on a CD which also makes it easy to replace hardware.



Hirlam area

- North Atlantic Area: resolution 150 milli degree on a 610x372x40 grid.
- European Area: Resolution 50 milli degree on a 496x372x40 grid.
- Southern Greenland: will be implemented later this month



Operational dataflow

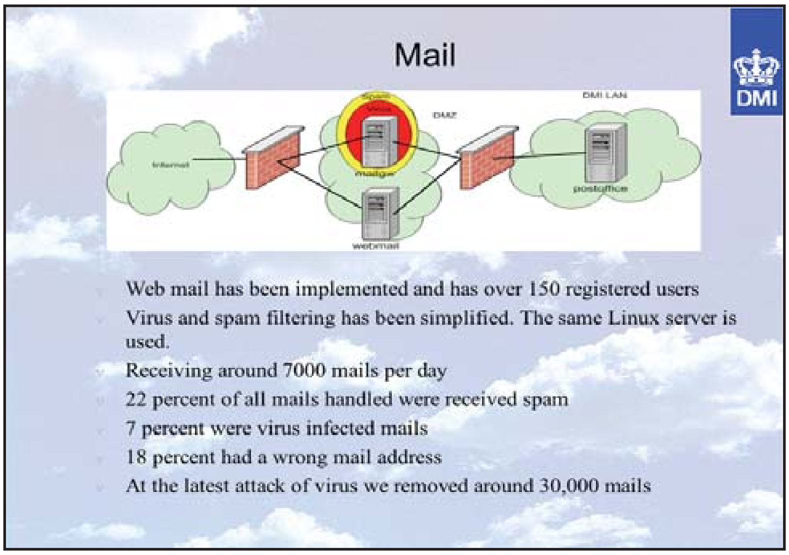
DMI receives data from ECMWF via RMDCN around 3 GByte per day of which more that 60 % is boundary products for the DMI forecast model.

The remaining data are selected ECMWF forecasts, ensemble forecast and wave model forecasts, which are plotted for use by the forecasters.

The bandwidth of the backup ISDN line is only one third of the primary line, so problems with the primary line will cause serious delays at DMI.

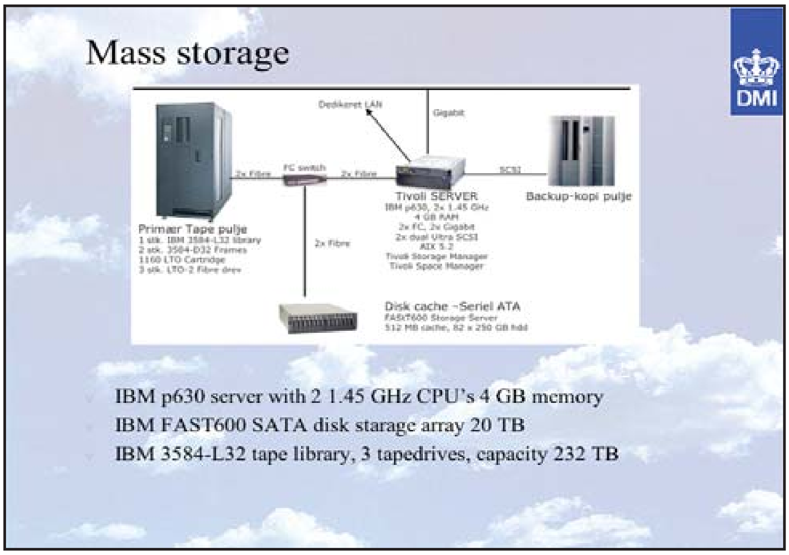
Dissemination via internet has been tested and will be used as backup instead of the ISDN line in near future.

On the communication lines that receives observations we receive around 65 Mb (55000 bulletins) per day and send around 400 kb (1000 bulletins) per day.



Mail

- Web mail has been implemented and has over 150 registered users
- Virus and spam filtering has been simplified. The same Linux server is used.
- Receiving around 7000 mails per day
- 22 percent of all mails handled were received spam
- 7 percent were virus infected mails
- 18 percent had a wrong mail address
- At the latest attack of virus we removed around 30,000 mails



Mass storage

Dedikeret LAN

Gigabit

2x Fibre

FC switch

2x Fibre

SCSI

Tivoli SERVER
IBM p630, 2x 1.45 GHz
4 GB RAM
2x FC, 2x Gigabit
2x dual Ultra SCSI
AIX 5.2
Tivoli Storage Manager
Tivoli Space Manager

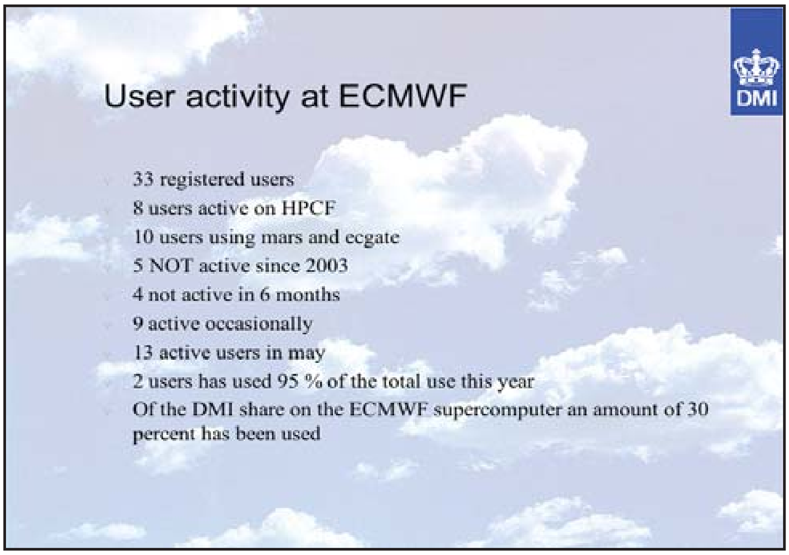
Backup-kopi pulje

Primær Tape pulje
1 stk. IBM 3584-L32 library
2 stk. 3584-D32 Frames
1160 LTO Cartridge
3 stk. LTO-2 Fibre drev

2x Fibre

Disk cache –Seriel ATA
FAStT600 Storage Server
512 MB cache, 82 x 250 GB hdd

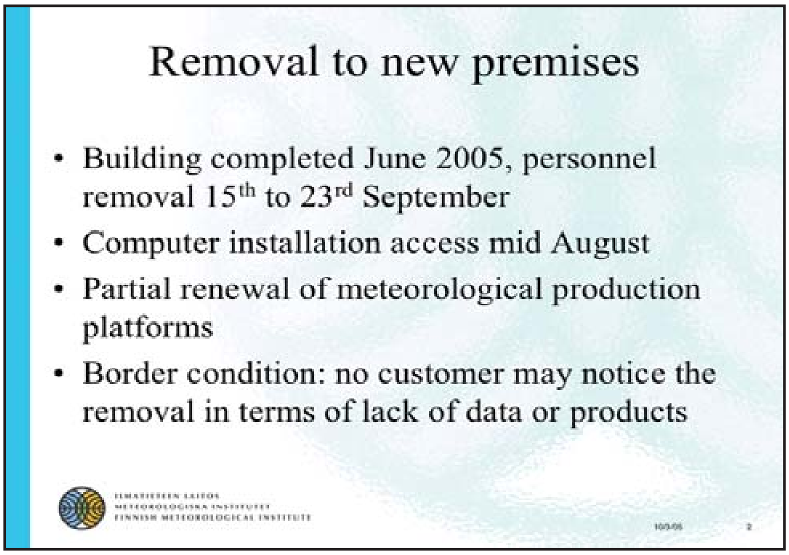
- IBM p630 server with 2 1.45 GHz CPU's 4 GB memory
- IBM FAST600 SATA disk starage array 20 TB
- IBM 3584-L32 tape library, 3 tapedrives, capacity 232 TB



User activity at ECMWF

- 33 registered users
- 8 users active on HPCF
- 10 users using mars and ecgate
- 5 NOT active since 2003
- 4 not active in 6 months
- 9 active occasionally
- 13 active users in may
- 2 users has used 95 % of the total use this year
- Of the DMI share on the ECMWF supercomputer an amount of 30 percent has been used

Kari Niemelä – Finnish Meteorological Institute

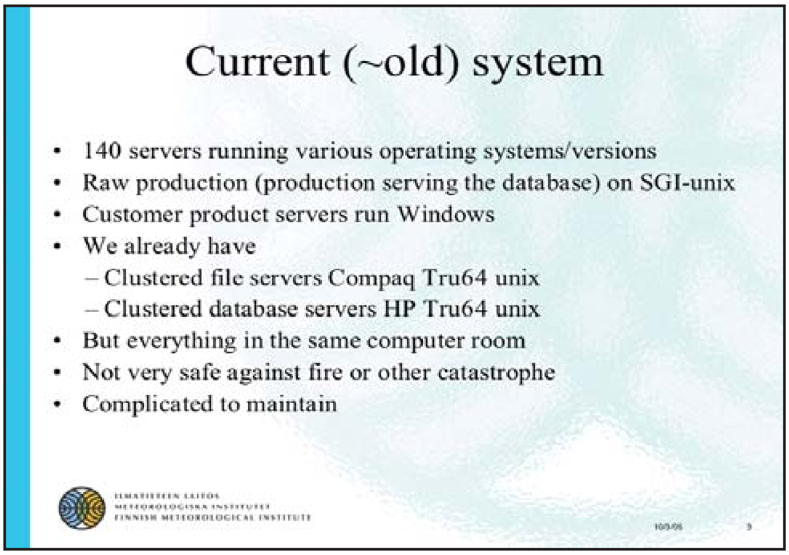


Removal to new premises

- Building completed June 2005, personnel removal 15th to 23rd September
- Computer installation access mid August
- Partial renewal of meteorological production platforms
- Border condition: no customer may notice the removal in terms of lack of data or products

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

10/3/05 2

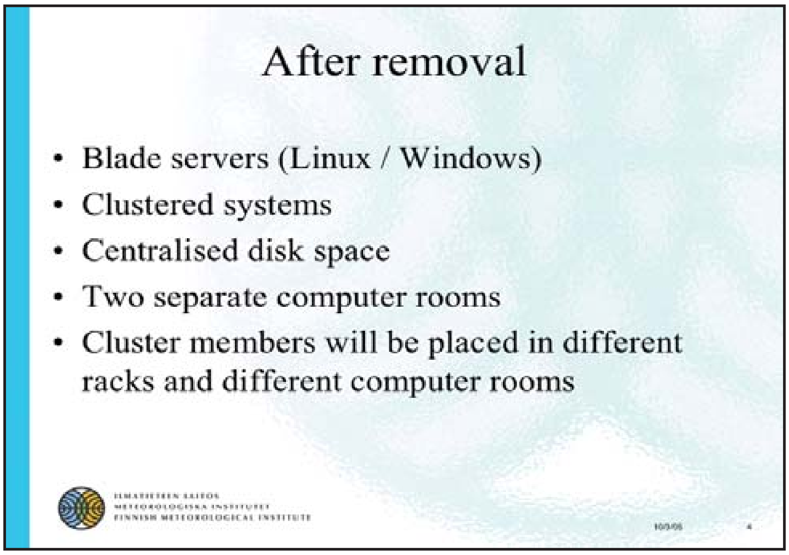


Current (~old) system

- 140 servers running various operating systems/versions
- Raw production (production serving the database) on SGI-unix
- Customer product servers run Windows
- We already have
  - Clustered file servers Compaq Tru64 unix
  - Clustered database servers HP Tru64 unix
- But everything in the same computer room
- Not very safe against fire or other catastrophe
- Complicated to maintain

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

10/3/05 3

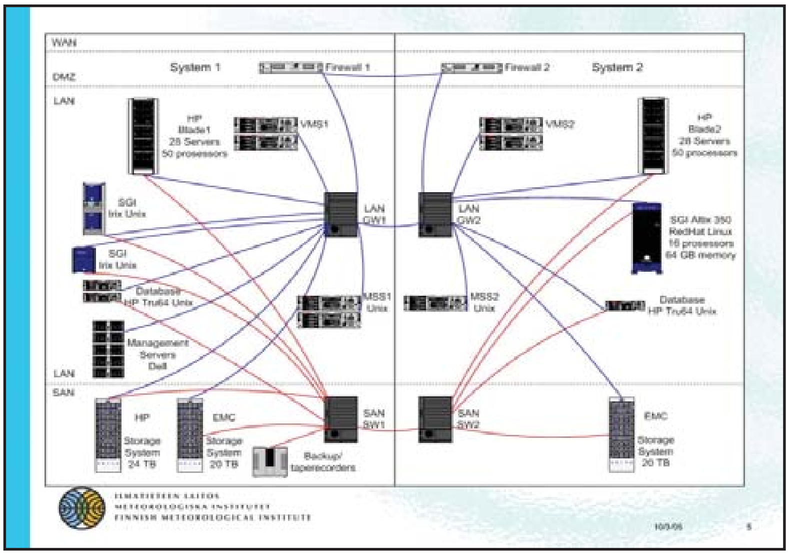


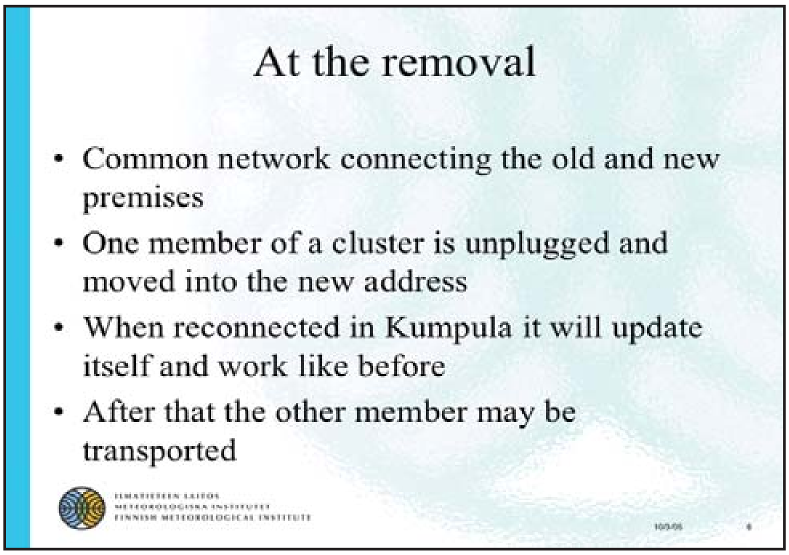
After removal

- Blade servers (Linux / Windows)
- Clustered systems
- Centralised disk space
- Two separate computer rooms
- Cluster members will be placed in different racks and different computer rooms

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

10/3/05 4



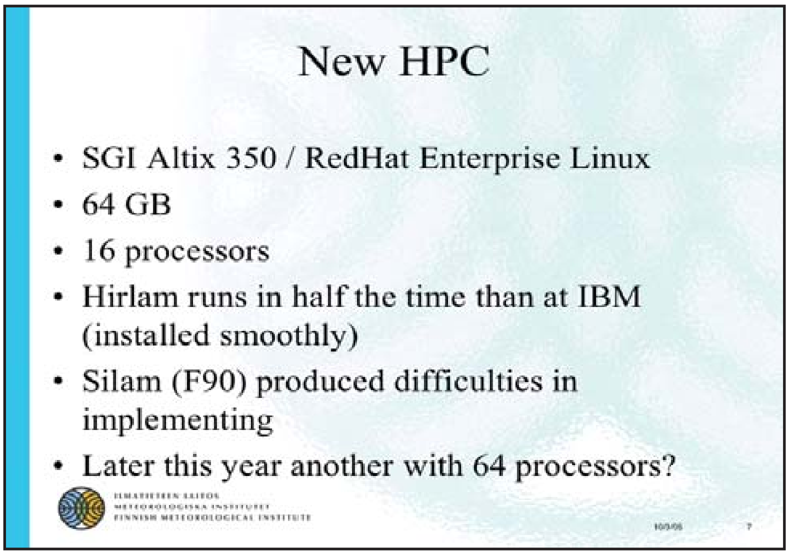


At the removal

- Common network connecting the old and new premises
- One member of a cluster is unplugged and moved into the new address
- When reconnected in Kumpula it will update itself and work like before
- After that the other member may be transported

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

10/3-05 6



New HPC

- SGI Altix 350 / RedHat Enterprise Linux
- 64 GB
- 16 processors
- Hirlam runs in half the time than at IBM (installed smoothly)
- Silam (F90) produced difficulties in implementing
- Later this year another with 64 processors?

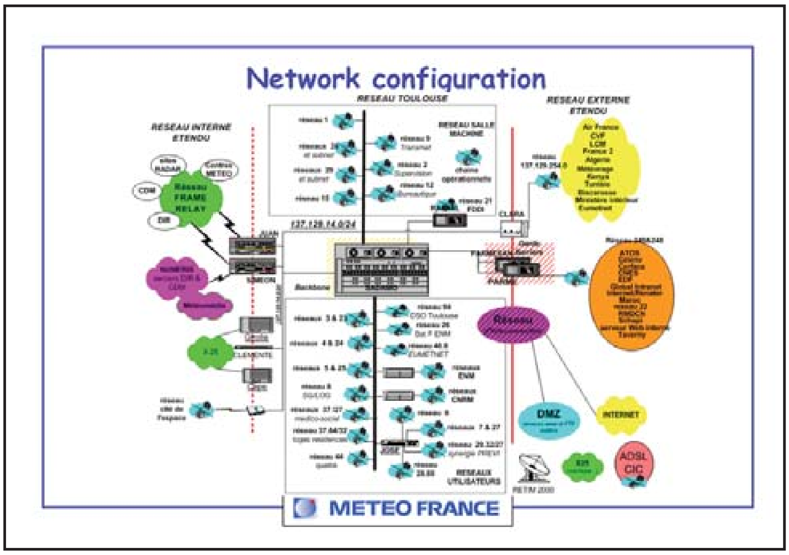
ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

10/3-05 7

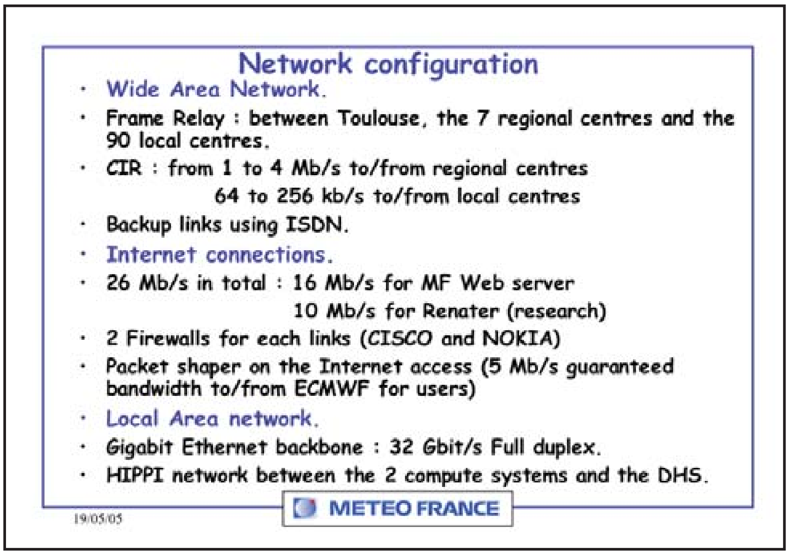
I. Weger asked the configuration of their Disaster Recovery System. K. Niemela replied that all vital systems and data are replicated exactly.

FRANCE

Marion Pithon – Météo-France



Network configuration

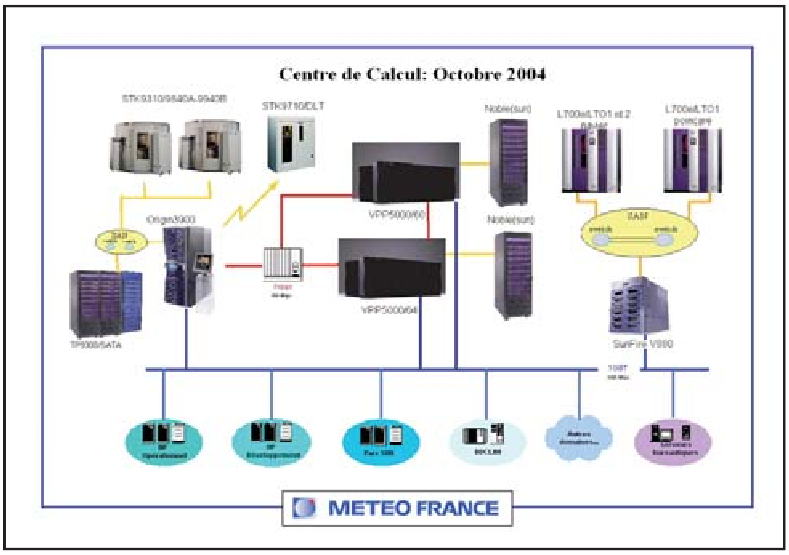


Network configuration

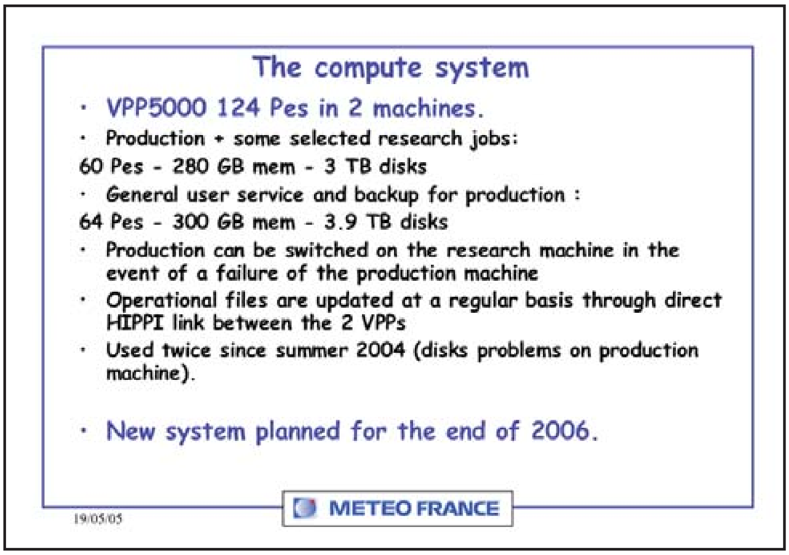
- Wide Area Network.
- Frame Relay : between Toulouse, the 7 regional centres and the 90 local centres.
- CIR : from 1 to 4 Mb/s to/from regional centres
  64 to 256 kb/s to/from local centres
- Backup links using ISDN.
- Internet connections.
- 26 Mb/s in total : 16 Mb/s for MF Web server
  10 Mb/s for Renater (research)
- 2 Firewalls for each links (CISCO and NOKIA)
- Packet shaper on the Internet access (5 Mb/s guaranteed bandwidth to/from ECMWF for users)
- Local Area network.
- Gigabit Ethernet backbone : 32 Gbit/s Full duplex.
- HIPPI network between the 2 compute systems and the DHS.

19/05/05

METEO FRANCE



Centre de Calcul: Octobre 2004



The compute system

- VPP5000 124 Pes in 2 machines.
- Production + some selected research jobs:

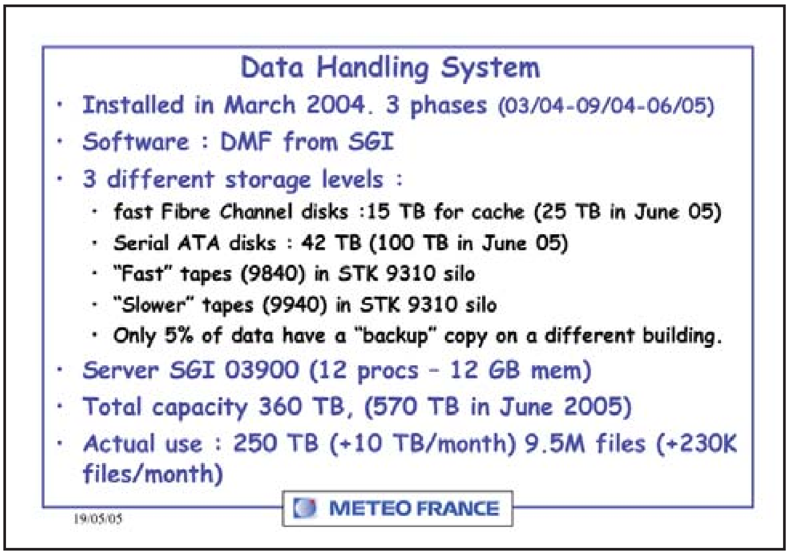
60 Pes - 280 GB mem - 3 TB disks

- General user service and backup for production :

64 Pes - 300 GB mem - 3.9 TB disks

- Production can be switched on the research machine in the event of a failure of the production machine
- Operational files are updated at a regular basis through direct HIPPI link between the 2 VPPs
- Used twice since summer 2004 (disks problems on production machine).
- New system planned for the end of 2006.

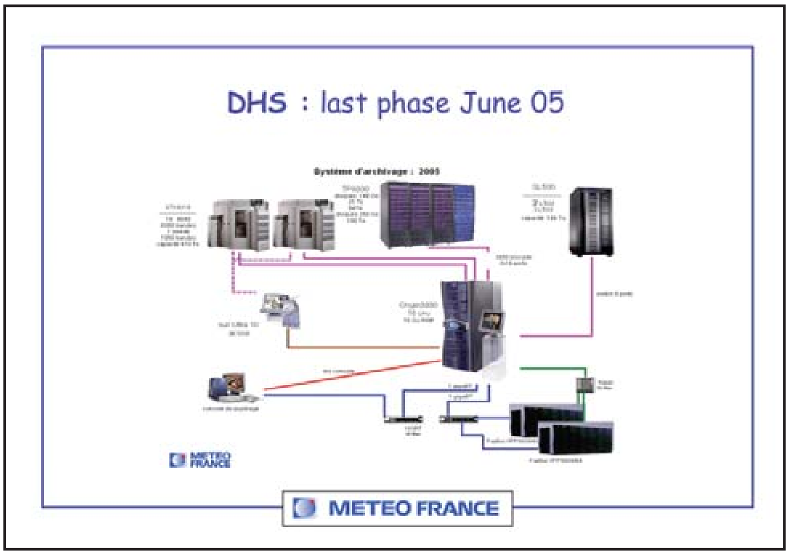
19/05/05 METEO FRANCE



Data Handling System

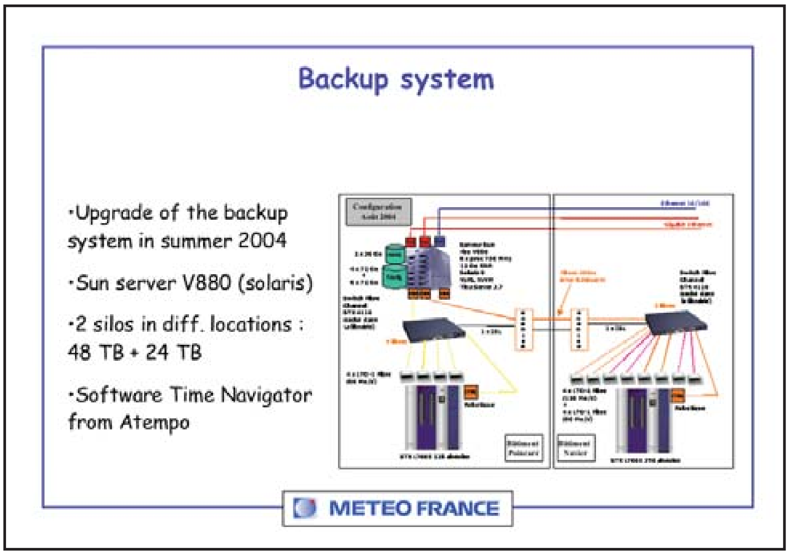
- Installed in March 2004. 3 phases (03/04-09/04-06/05)
- Software : DMF from SGI
- 3 different storage levels :
  - fast Fibre Channel disks :15 TB for cache (25 TB in June 05)
  - Serial ATA disks : 42 TB (100 TB in June 05)
  - "Fast" tapes (9840) in STK 9310 silo
  - "Slower" tapes (9940) in STK 9310 silo
  - Only 5% of data have a "backup" copy on a different building.
- Server SGI 03900 (12 procs – 12 GB mem)
- Total capacity 360 TB, (570 TB in June 2005)
- Actual use : 250 TB (+10 TB/month) 9.5M files (+230K files/month)

19/05/05 METEO FRANCE



DHS : last phase June 05

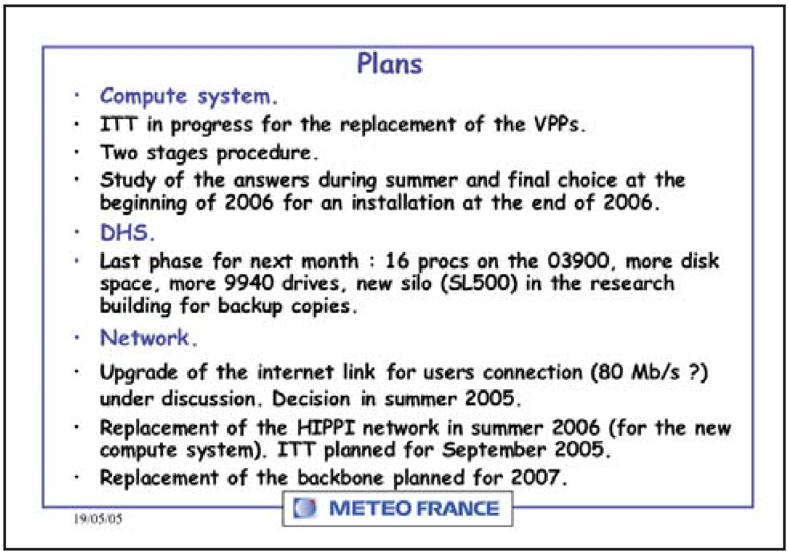
METEO FRANCE



Backup system

- Upgrade of the backup system in summer 2004
- Sun server V880 (solaris)
- 2 silos in diff. locations : 48 TB + 24 TB
- Software Time Navigator from Atempo

METEO FRANCE

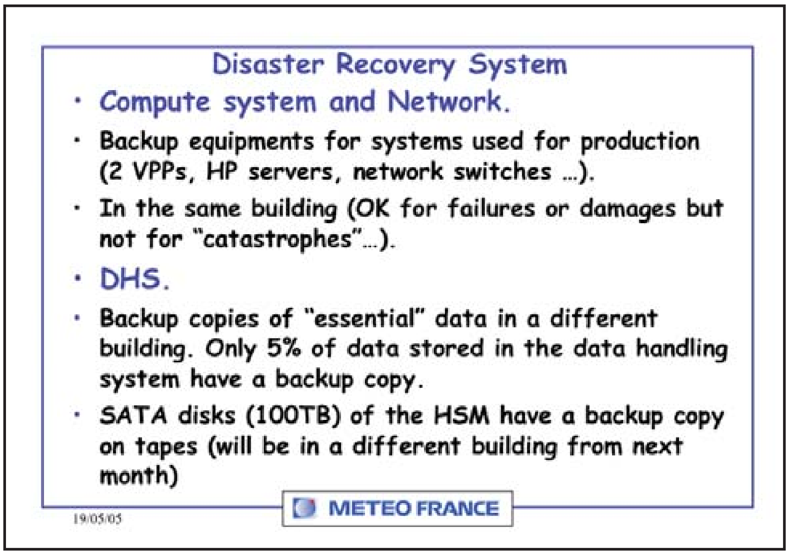


Plans

- Compute system.
- ITT in progress for the replacement of the VPPs.
- Two stages procedure.
- Study of the answers during summer and final choice at the beginning of 2006 for an installation at the end of 2006.
- DHS.
- Last phase for next month : 16 procs on the O3900, more disk space, more 9940 drives, new silo (SL500) in the research building for backup copies.
- Network.
- Upgrade of the internet link for users connection (80 Mb/s ?) under discussion. Decision in summer 2005.
- Replacement of the HIPPI network in summer 2006 (for the new compute system). ITT planned for September 2005.
- Replacement of the backbone planned for 2007.

19/05/05

METEO FRANCE

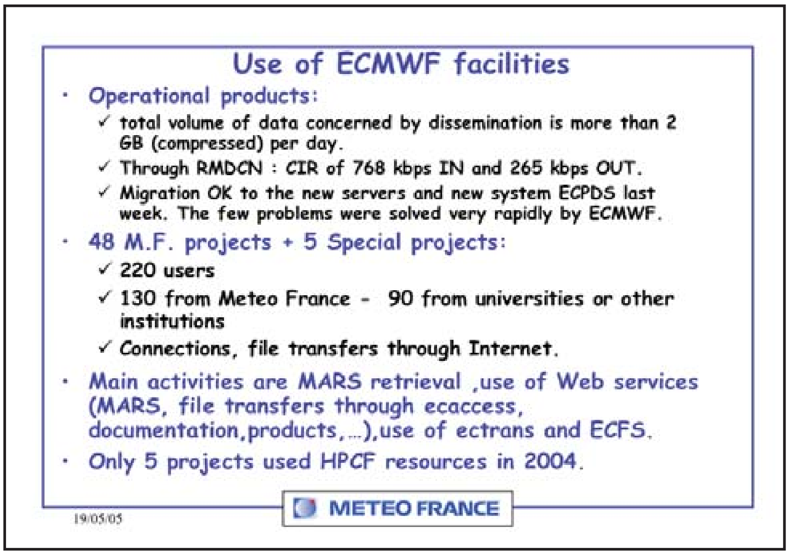


Disaster Recovery System

- Compute system and Network.
- Backup equipments for systems used for production (2 VPPs, HP servers, network switches …).
- In the same building (OK for failures or damages but not for "catastrophes"…).
- DHS.
- Backup copies of "essential" data in a different building. Only 5% of data stored in the data handling system have a backup copy.
- SATA disks (100TB) of the HSM have a backup copy on tapes (will be in a different building from next month)

19/05/05

METEO FRANCE

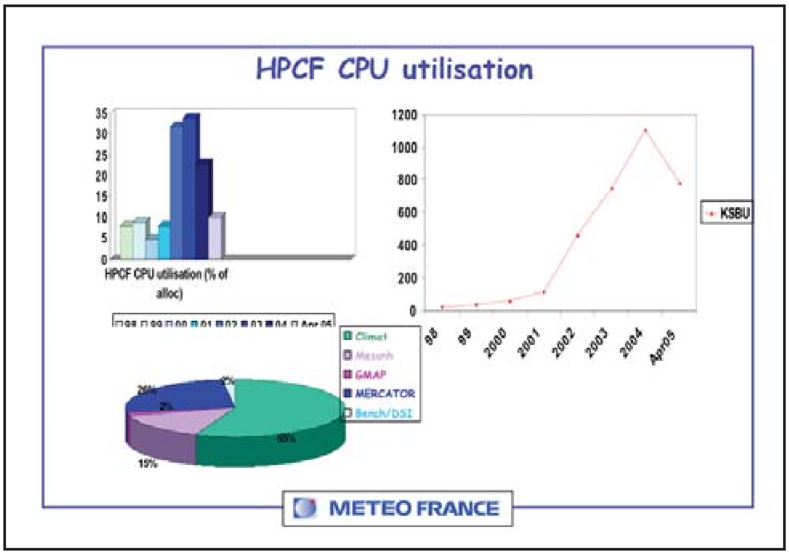


Use of ECMWF facilities

- Operational products:
  - ✓ total volume of data concerned by dissemination is more than 2 GB (compressed) per day.
  - ✓ Through RMDCN : CIR of 768 kbps IN and 265 kbps OUT.
  - ✓ Migration OK to the new servers and new system ECPDS last week. The few problems were solved very rapidly by ECMWF.
- 48 M.F. projects + 5 Special projects:
  - ✓ 220 users
  - ✓ 130 from Meteo France - 90 from universities or other institutions
  - ✓ Connections, file transfers through Internet.
- Main activities are MARS retrieval ,use of Web services (MARS, file transfers through ecaccess, documentation,products,…),use of ectrans and ECFS.
- Only 5 projects used HPCF resources in 2004.

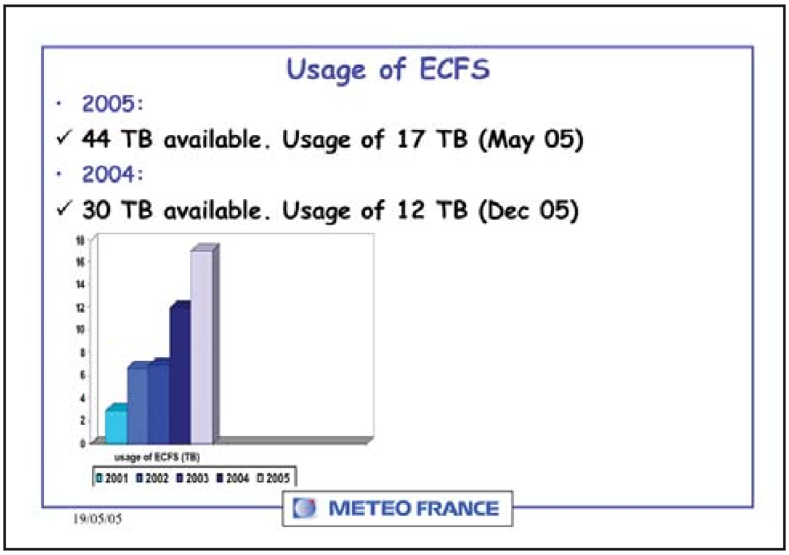
19/05/05

METEO FRANCE



HPCF CPU utilisation

METEO FRANCE



Usage of ECFS

- 2005:

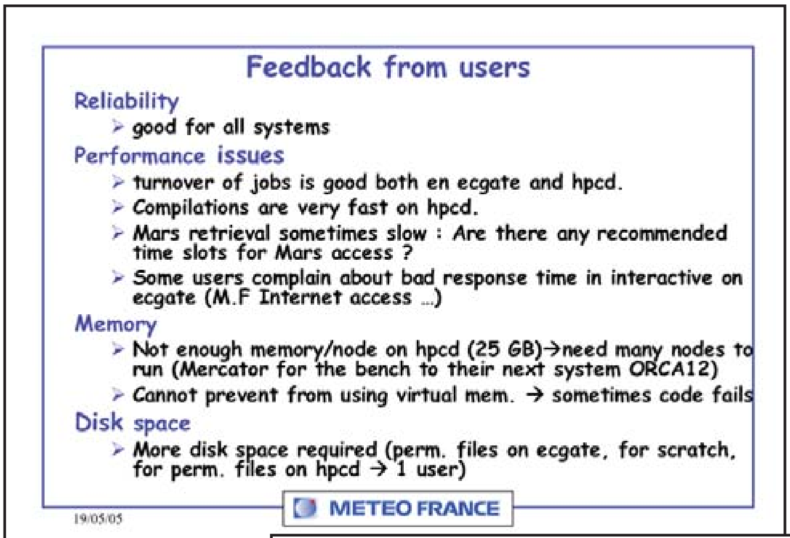
✓ 44 TB available. Usage of 17 TB (May 05)

- 2004:

✓ 30 TB available. Usage of 12 TB (Dec 05)

19/05/05

METEO FRANCE



Feedback from users

Reliability

- good for all systems

Performance issues

- turnover of jobs is good both en ecgate and hpcd.
- Compilations are very fast on hpcd.
- Mars retrieval sometimes slow : Are there any recommended time slots for Mars access ?
- Some users complain about bad response time in interactive on ecgate (M.F Internet access ...)

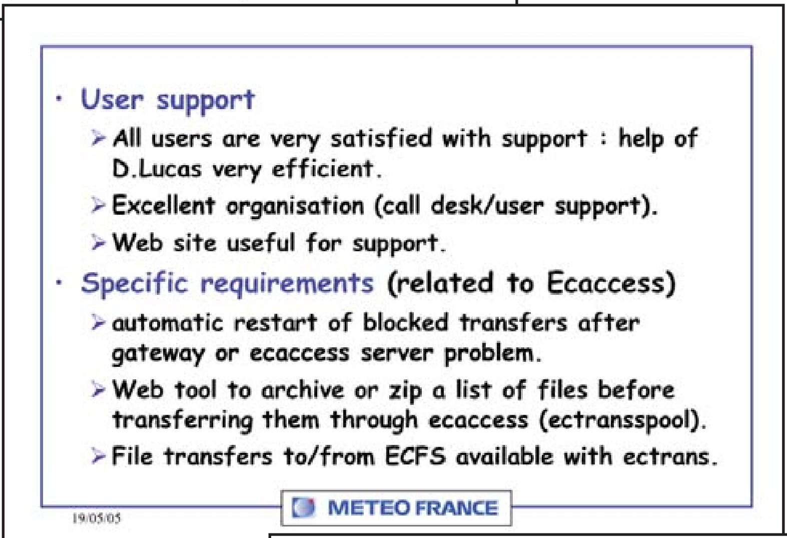
Memory

- Not enough memory/node on hpcd (25 GB)→need many nodes to run (Mercator for the bench to their next system ORCA12)
- Cannot prevent from using virtual mem. → sometimes code fails

Disk space

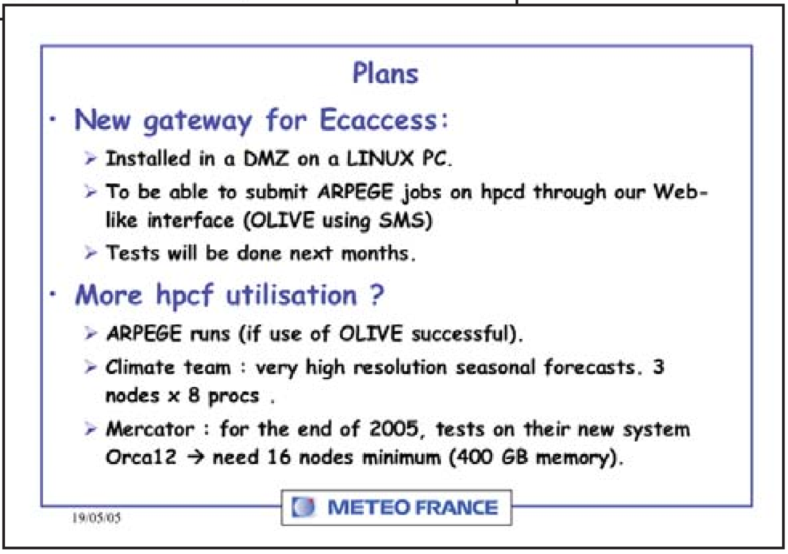
- More disk space required (perm. files on ecgate, for scratch, for perm. files on hpcd → 1 user)

19/05/05 METEO FRANCE



- User support
  - All users are very satisfied with support : help of D.Lucas very efficient.
  - Excellent organisation (call desk/user support).
  - Web site useful for support.
- Specific requirements (related to Ecaccess)
  - automatic restart of blocked transfers after gateway or ecaccess server problem.
  - Web tool to archive or zip a list of files before transferring them through ecaccess (ectransspool).
  - File transfers to/from ECFS available with ectrans.

19/05/05 METEO FRANCE



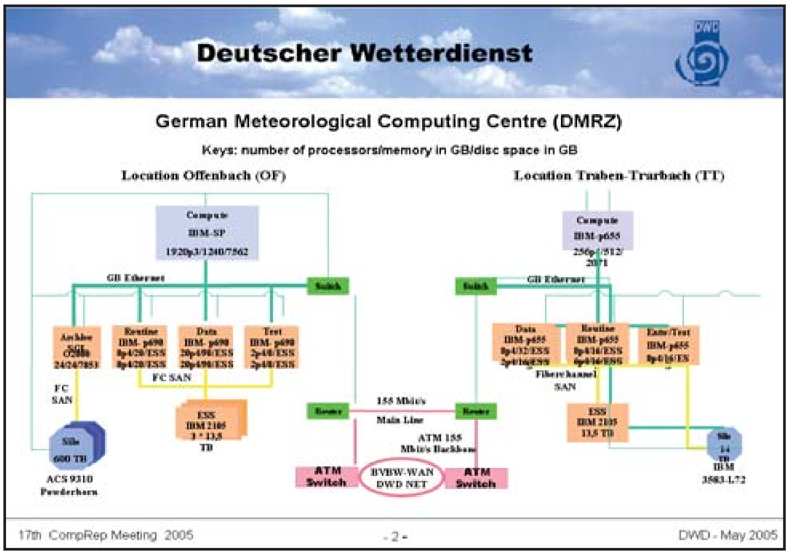
Plans

- New gateway for Ecaccess:
  - Installed in a DMZ on a LINUX PC.
  - To be able to submit ARPEGE jobs on hpcd through our Web-like interface (OLIVE using SMS)
  - Tests will be done next months.
- More hpcf utilisation ?
  - ARPEGE runs (if use of OLIVE successful).
  - Climate team : very high resolution seasonal forecasts. 3 nodes x 8 procs .
  - Mercator : for the end of 2005, tests on their new system Orca12 → need 16 nodes minimum (400 GB memory).

19/05/05 METEO FRANCE

With reference to the automatic restart of failed ECaccess transfers, L. Gougeon commented that this facility had already been implemented on ECPDS, so it must be possible to implement it also on ECaccess. File transfers to/from ECFS were possible with ecgate1 but a problem seems to have developed with the transfer to ecgate. Both problems will be solved.

Elizabeth Krenzien – Deutscher Wetterdienst

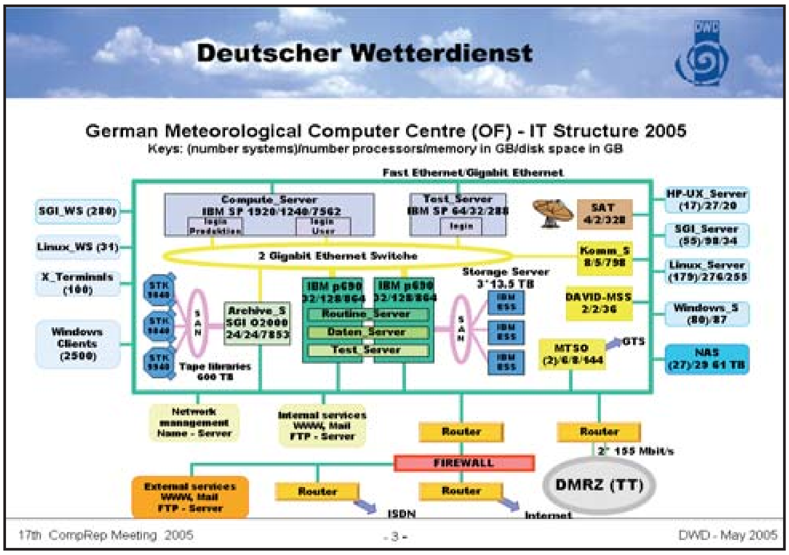


Deutscher Wetterdienst

German Meteorological Computing Centre (DMRZ)

Keys: number of processors/memory in GB/disc space in GB

17th CompRep Meeting 2005 - 2 - DWD - May 2005

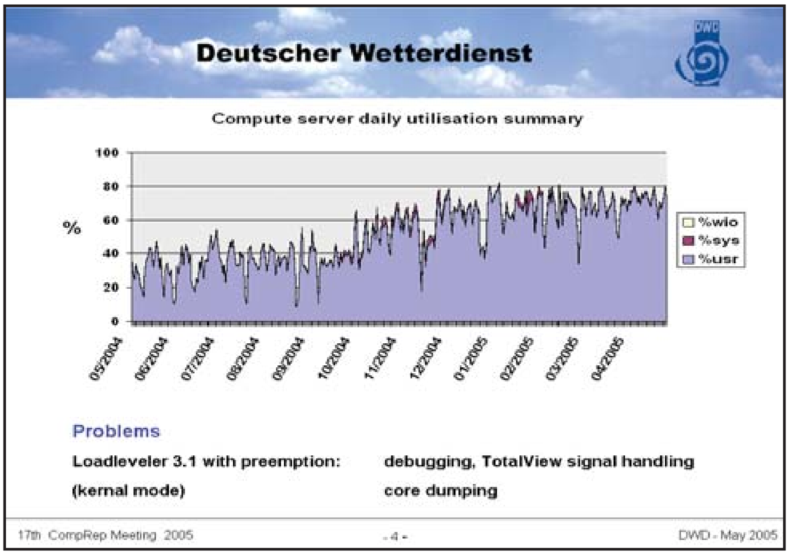


Deutscher Wetterdienst

German Meteorological Computer Centre (OF) - IT Structure 2005

Keys: (number systems)/number processors/memory in GB/disk space in GB

17th CompRep Meeting 2005 - 3 - DWD - May 2005



Deutscher Wetterdienst

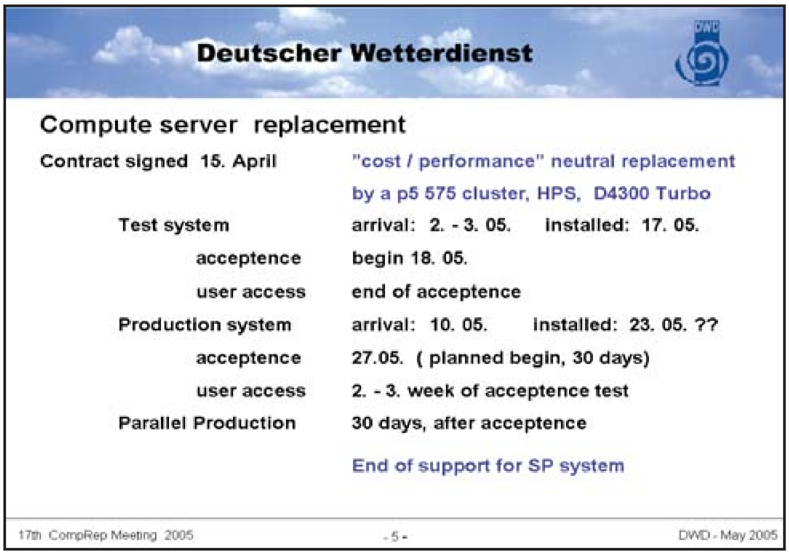
Compute server daily utilisation summary

Problems

Loadleveler 3.1 with preemption: debugging, TotalView signal handling

(kernal mode) core dumping

17th CompRep Meeting 2005 - 4 - DWD - May 2005



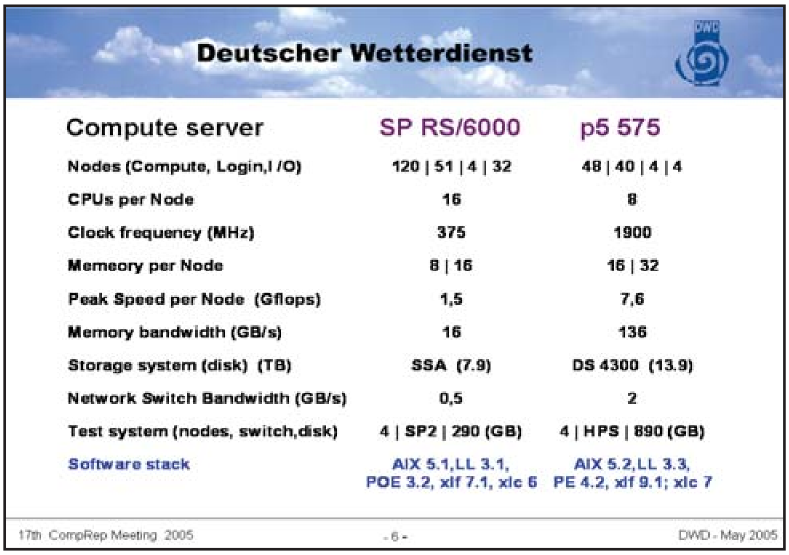
Deutscher Wetterdienst

DWD

Compute server replacement

Contract signed 15. April		"cost / performance" neutral replacement by a p5 575 cluster, HPS, D4300 Turbo
Test system		arrival: 2. - 3. 05. installed: 17. 05.
	acceptence	begin 18. 05.
	user access	end of acceptence
Production system		arrival: 10. 05. installed: 23. 05. ??
	acceptence	27.05. (planned begin, 30 days)
	user access	2. - 3. week of acceptence test
Parallel Production		30 days, after acceptence
		End of support for SP system

17th CompRep Meeting 2005 - 5 - DWD - May 2005

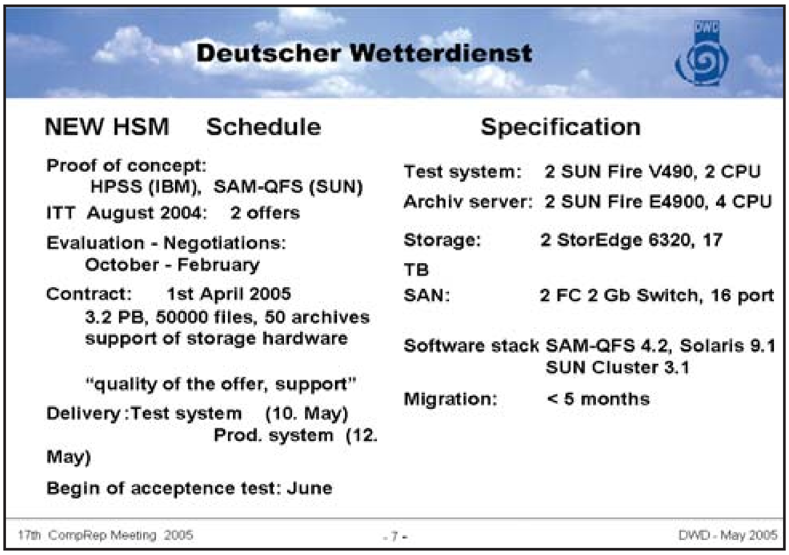


Deutscher Wetterdienst

DWD

Compute server	SP RS/6000	p5 575
Nodes (Compute, Login,I /O)	120 | 51 | 4 | 32	48 | 40 | 4 | 4
CPUs per Node	16	8
Clock frequency (MHz)	375	1900
Memeory per Node	8 | 16	16 | 32
Peak Speed per Node (Gflops)	1,5	7,6
Memory bandwidth (GB/s)	16	136
Storage system (disk) (TB)	SSA (7.9)	DS 4300 (13.9)
Network Switch Bandwidth (GB/s)	0,5	2
Test system (nodes, switch,disk)	4 | SP2 | 290 (GB)	4 | HPS | 890 (GB)
Software stack	AIX 5.1,LL 3.1, POE 3.2, xlf 7.1, xlc 6	AIX 5.2,LL 3.3, PE 4.2, xlf 9.1; xlc 7

17th CompRep Meeting 2005 - 6 - DWD - May 2005



Deutscher Wetterdienst

DWD

NEW HSM Schedule

Proof of concept:
HPSS (IBM), SAM-QFS (SUN)

ITT August 2004: 2 offers

Evaluation - Negotiations:
October - February

Contract: 1st April 2005
3.2 PB, 50000 files, 50 archives
support of storage hardware

"quality of the offer, support"

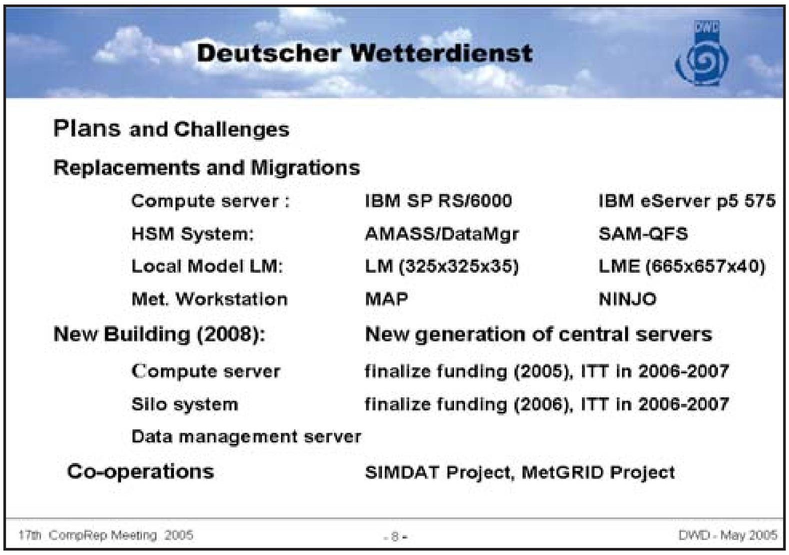
Delivery :Test system (10. May)
Prod. system (12. May)

Begin of acceptence test: June

Specification

Test system:	2 SUN Fire V490, 2 CPU
Archiv server:	2 SUN Fire E4900, 4 CPU
Storage:	2 StorEdge 6320, 17 TB
SAN:	2 FC 2 Gb Switch, 16 port
Software stack	SAM-QFS 4.2, Solaris 9.1 SUN Cluster 3.1
Migration:	< 5 months

17th CompRep Meeting 2005 - 7 - DWD - May 2005



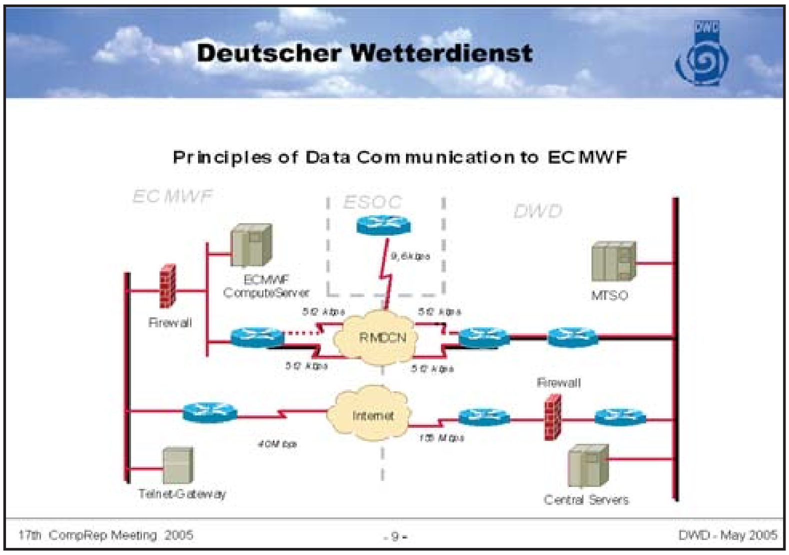
Deutscher Wetterdienst

Plans and Challenges

Replacements and Migrations

Compute server :	IBM SP RS/6000	IBM eServer p5 575
HSM System:	AMASS/DataMgr	SAM-QFS
Local Model LM:	LM (325x325x35)	LME (665x657x40)
Met. Workstation	MAP	NINJO
New Building (2008):	**New generation of central servers**	
Compute server	finalize funding (2005), ITT in 2006-2007	
Silo system	finalize funding (2006), ITT in 2006-2007	
Data management server		
Co-operations	SIMDAT Project, MetGRID Project	

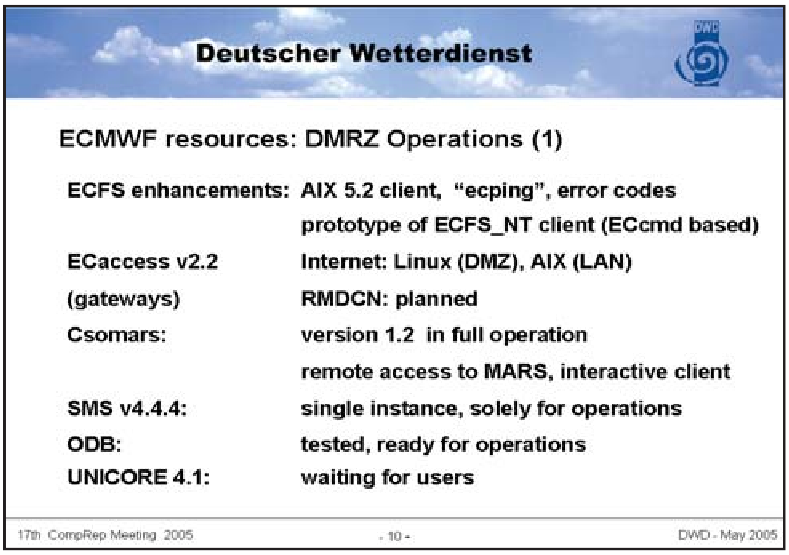
17th CompRep Meeting 2005 - 8 - DWD - May 2005



Deutscher Wetterdienst

Principles of Data Communication to ECMWF

17th CompRep Meeting 2005 - 9 - DWD - May 2005

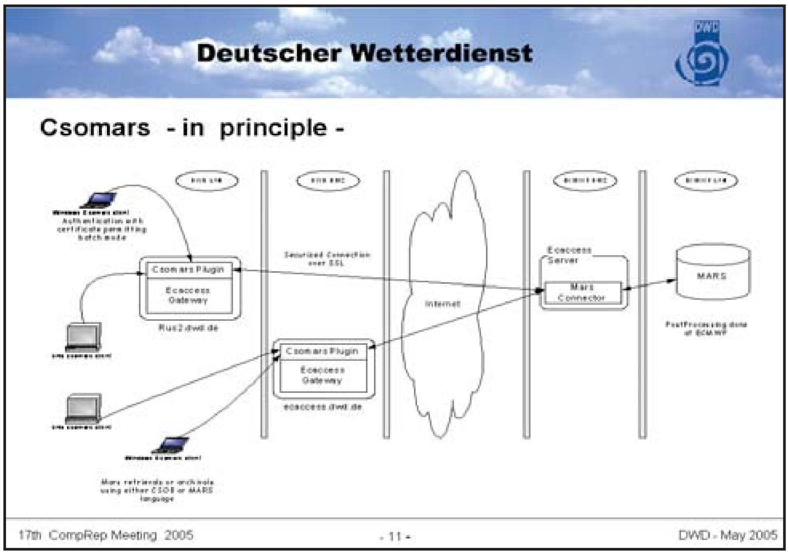


Deutscher Wetterdienst

ECMWF resources: DMRZ Operations (1)

ECFS enhancements:	AIX 5.2 client, "ecping", error codes prototype of ECFS_NT client (ECcmd based)
ECaccess v2.2	Internet: Linux (DMZ), AIX (LAN)
(gateways)	RMDCN: planned
Csomars:	version 1.2 in full operation remote access to MARS, interactive client
SMS v4.4.4:	single instance, solely for operations
ODB:	tested, ready for operations
UNICORE 4.1:	waiting for users

17th CompRep Meeting 2005 - 10 - DWD - May 2005

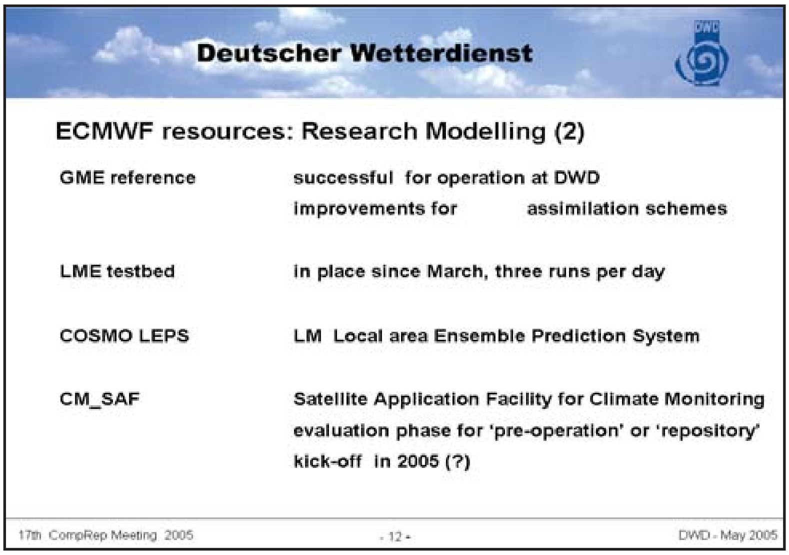


Deutscher Wetterdienst

Csomars - in principle -

Authentication with certificate permitting batch mode
Csomars Plugin
Ecaccess Gateway
Rus2.dwd.de
Securized Connection over SSL
Csomars Plugin
Ecaccess Gateway
ecaccess.dwd.de
Internet
Ecaccess Server
Mars Connector
MARS
PostProcessing done at ECMWF
Mars retrievals or archivals using either CSOB or MARS language

17th CompRep Meeting 2005 - 11 - DWD - May 2005

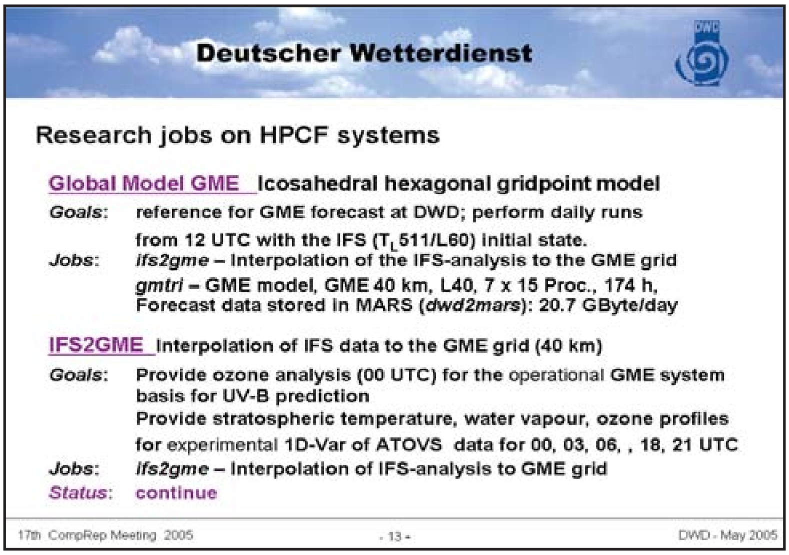


Deutscher Wetterdienst

ECMWF resources: Research Modelling (2)

GME reference	successful for operation at DWD improvements for assimilation schemes
LME testbed	in place since March, three runs per day
COSMO LEPS	LM Local area Ensemble Prediction System
CM_SAF	Satellite Application Facility for Climate Monitoring evaluation phase for 'pre-operation' or 'repository' kick-off in 2005 (?)

17th CompRep Meeting 2005 - 12 - DWD - May 2005



Deutscher Wetterdienst

Research jobs on HPCF systems

Global Model GME Icosahedral hexagonal gridpoint model

Goals: reference for GME forecast at DWD; perform daily runs from 12 UTC with the IFS (T_L511/L60) initial state.

Jobs: *ifs2gme* – Interpolation of the IFS-analysis to the GME grid
gmtri – GME model, GME 40 km, L40, 7 x 15 Proc., 174 h, Forecast data stored in MARS (*dwd2mars*): 20.7 GByte/day

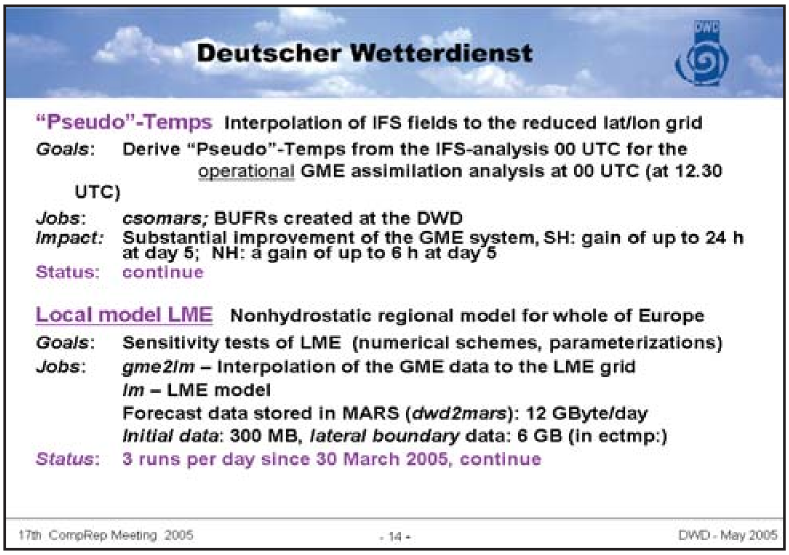
IFS2GME Interpolation of IFS data to the GME grid (40 km)

Goals: Provide ozone analysis (00 UTC) for the operational GME system basis for UV-B prediction
Provide stratospheric temperature, water vapour, ozone profiles for experimental 1D-Var of ATOVS data for 00, 03, 06, , 18, 21 UTC

Jobs: *ifs2gme* – Interpolation of IFS-analysis to GME grid

Status: continue

17th CompRep Meeting 2005 - 13 - DWD - May 2005



Deutscher Wetterdienst

"Pseudo"-Temps Interpolation of IFS fields to the reduced lat/lon grid

Goals: Derive "Pseudo"-Temps from the IFS-analysis 00 UTC for the operational GME assimilation analysis at 00 UTC (at 12.30 UTC)

Jobs: *csomars;* BUFRs created at the DWD

Impact: Substantial improvement of the GME system, SH: gain of up to 24 h at day 5; NH: a gain of up to 6 h at day 5

Status: continue

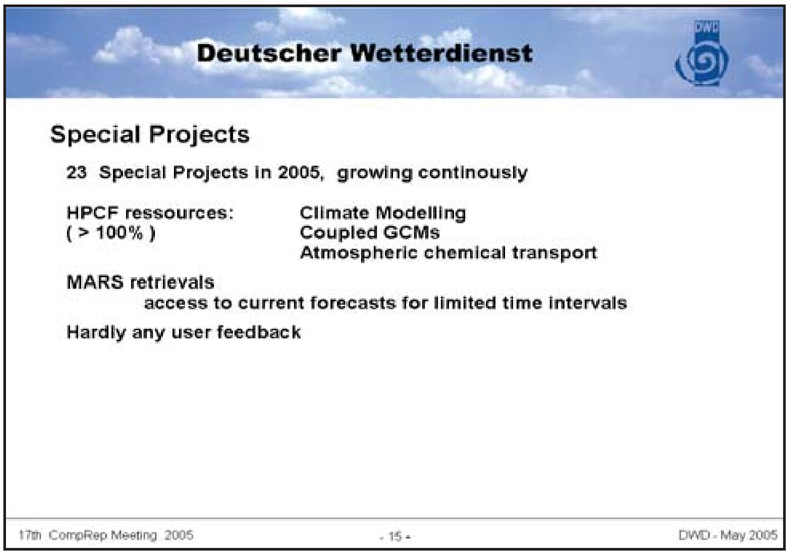
Local model LME Nonhydrostatic regional model for whole of Europe

Goals: Sensitivity tests of LME (numerical schemes, parameterizations)

Jobs: *gme2lm* – Interpolation of the GME data to the LME grid
lm – LME model
Forecast data stored in MARS (*dwd2mars*): 12 GByte/day
Initial data: 300 MB, *lateral boundary* data: 6 GB (in ectmp:)

Status: 3 runs per day since 30 March 2005, continue

17th CompRep Meeting 2005 - 14 - DWD - May 2005



Deutscher Wetterdienst

Special Projects

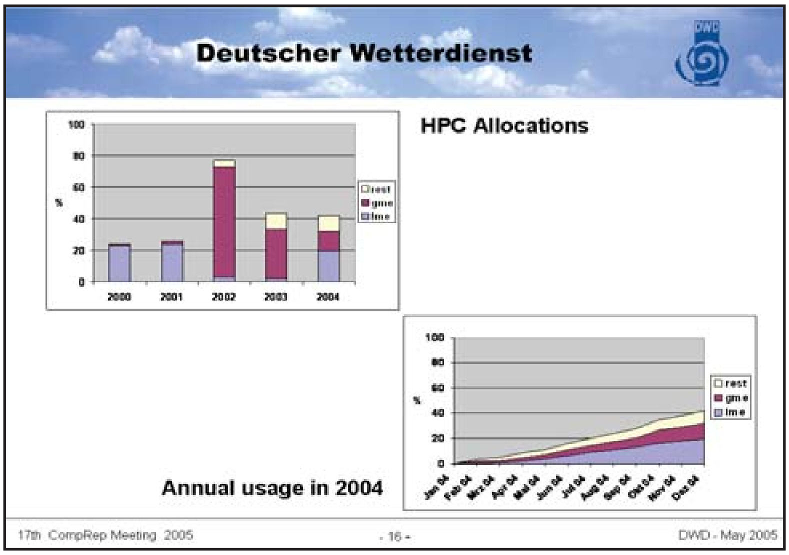
23 Special Projects in 2005, growing continously

HPCF ressources: (> 100%)
Climate Modelling
Coupled GCMs
Atmospheric chemical transport

MARS retrievals
access to current forecasts for limited time intervals

Hardly any user feedback

17th CompRep Meeting 2005 - 15 - DWD - May 2005

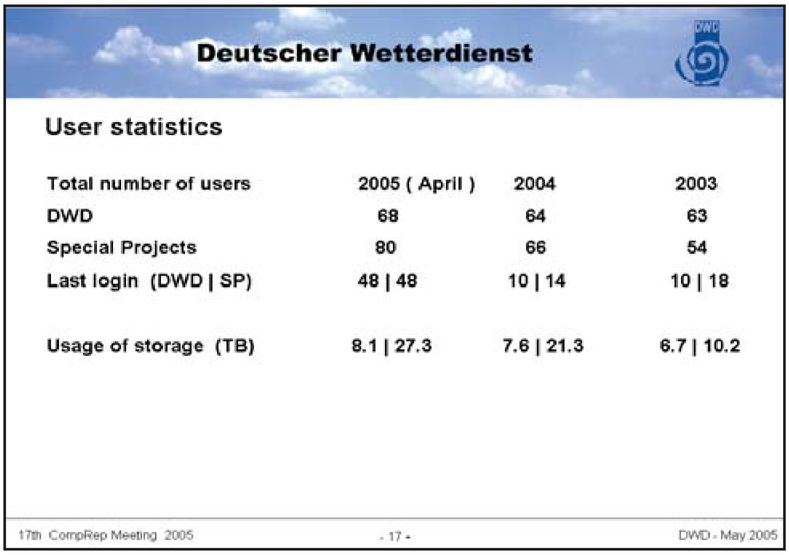


Deutscher Wetterdienst

HPC Allocations

Annual usage in 2004

17th CompRep Meeting 2005 - 16 - DWD - May 2005



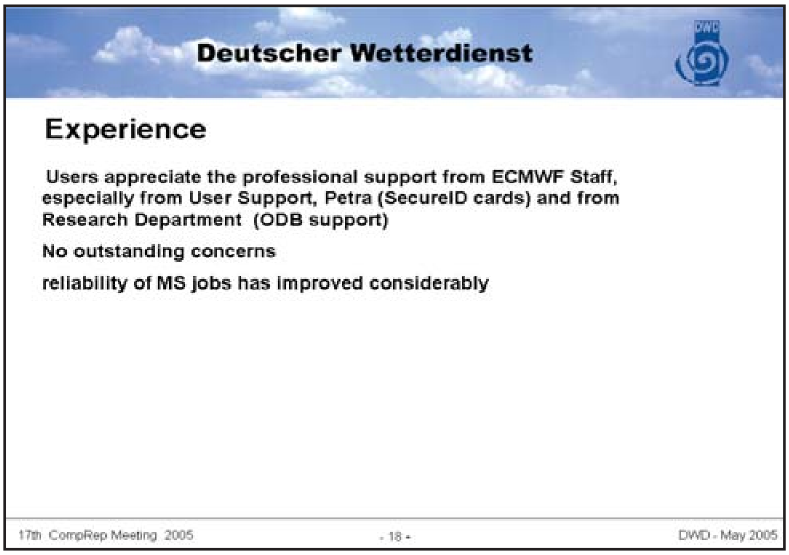
Deutscher Wetterdienst

DWD

User statistics

Total number of users	2005 (April)	2004	2003
DWD	68	64	63
Special Projects	80	66	54
Last login (DWD | SP)	48 | 48	10 | 14	10 | 18
Usage of storage (TB)	8.1 | 27.3	7.6 | 21.3	6.7 | 10.2

17th CompRep Meeting 2005 - 17 - DWD - May 2005



Deutscher Wetterdienst

DWD

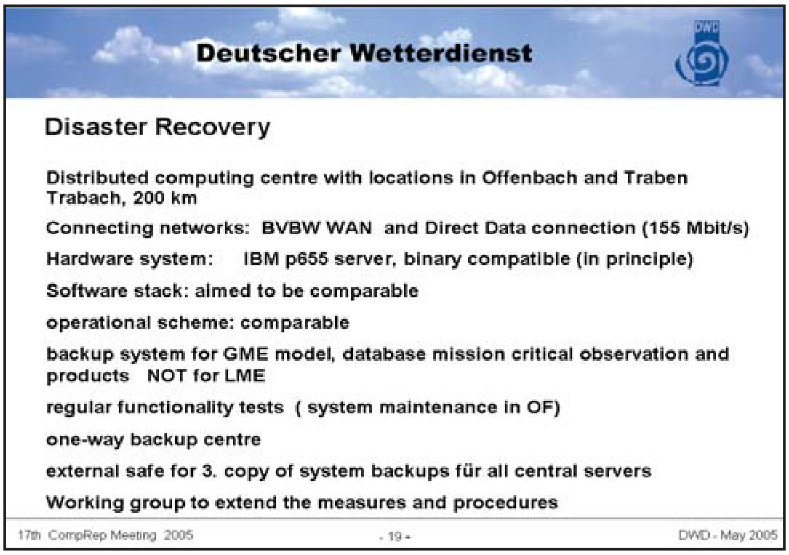
Experience

Users appreciate the professional support from ECMWF Staff, especially from User Support, Petra (SecureID cards) and from Research Department (ODB support)

No outstanding concerns

reliability of MS jobs has improved considerably

17th CompRep Meeting 2005 - 18 - DWD - May 2005



Deutscher Wetterdienst

DWD

Disaster Recovery

Distributed computing centre with locations in Offenbach and Traben Trabach, 200 km

Connecting networks: BVBW WAN and Direct Data connection (155 Mbit/s)

Hardware system: IBM p655 server, binary compatible (in principle)

Software stack: aimed to be comparable

operational scheme: comparable

backup system for GME model, database mission critical observation and products NOT for LME

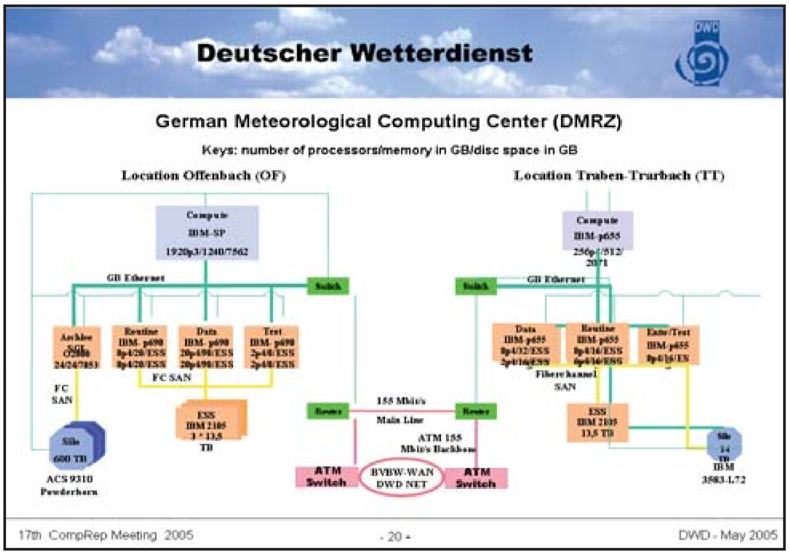
regular functionality tests (system maintenance in OF)

one-way backup centre

external safe for 3. copy of system backups für all central servers

Working group to extend the measures and procedures

17th CompRep Meeting 2005 - 19 - DWD - May 2005

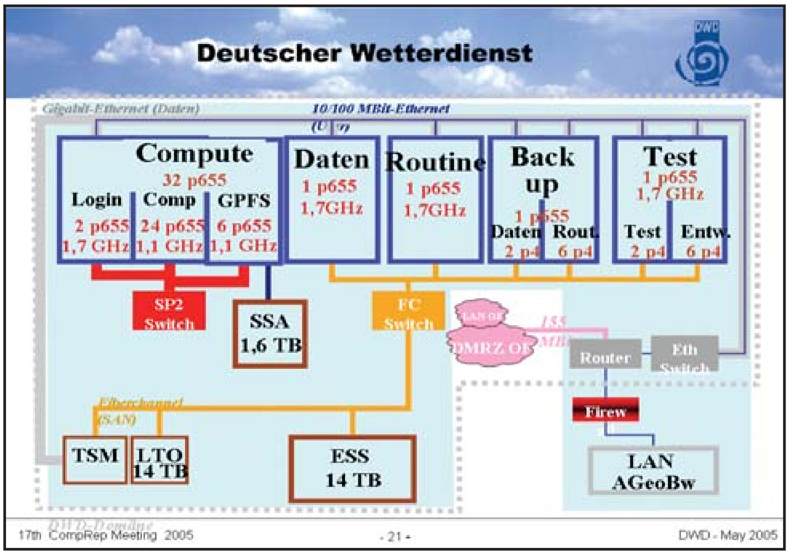


Deutscher Wetterdienst

German Meteorological Computing Center (DMRZ)

Keys: number of processors/memory in GB/disc space in GB

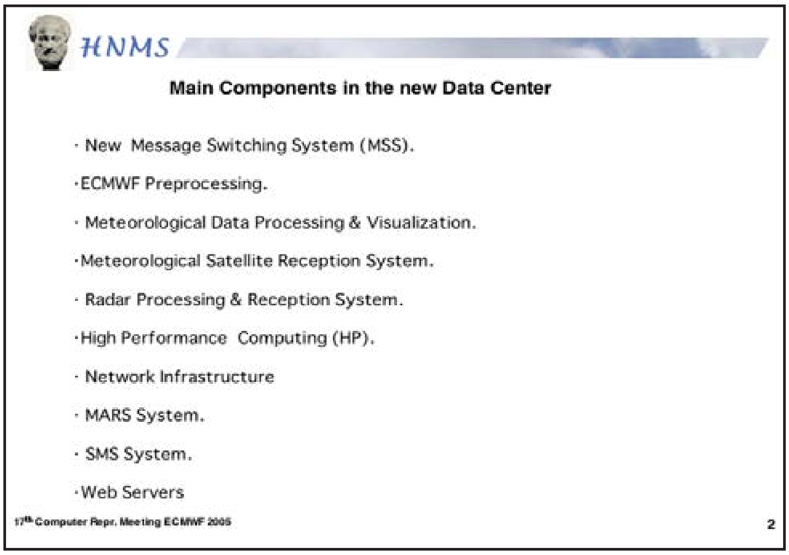
17th CompRep Meeting 2005 - 20 - DWD - May 2005



Deutscher Wetterdienst

17th CompRep Meeting 2005 - 21 - DWD - May 2005

Ioannis Alexiou – Hellenic Meteorological Service

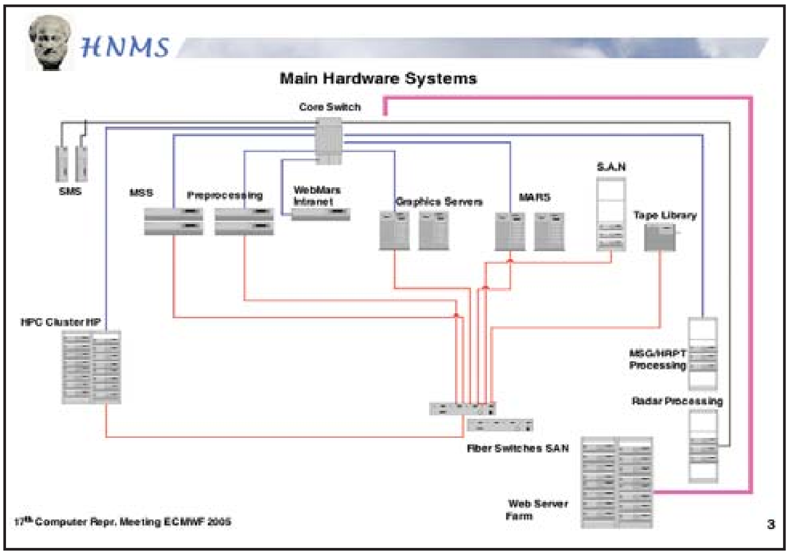


HNMS

Main Components in the new Data Center

- New Message Switching System (MSS).
- ECMWF Preprocessing.
- Meteorological Data Processing & Visualization.
- Meteorological Satellite Reception System.
- Radar Processing & Reception System.
- High Performance Computing (HP).
- Network Infrastructure
- MARS System.
- SMS System.
- Web Servers

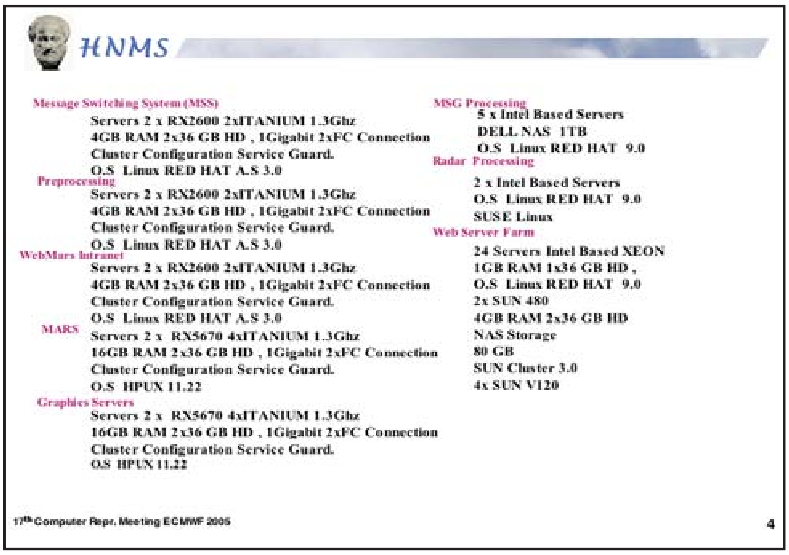
17th Computer Repr. Meeting ECMWF 2005 2



HNMS

Main Hardware Systems

17th Computer Repr. Meeting ECMWF 2005 3



HNMS

Message Switching System (MSS)
Servers 2 x RX2600 2xITANIUM 1.3Ghz
4GB RAM 2x36 GB HD , 1Gigabit 2xFC Connection
Cluster Configuration Service Guard.
O.S Linux RED HAT A.S 3.0

Preprocessing
Servers 2 x RX2600 2xITANIUM 1.3Ghz
4GB RAM 2x36 GB HD , 1Gigabit 2xFC Connection
Cluster Configuration Service Guard.
O.S Linux RED HAT A.S 3.0

WebMars Intranet
Servers 2 x RX2600 2xITANIUM 1.3Ghz
4GB RAM 2x36 GB HD , 1Gigabit 2xFC Connection
Cluster Configuration Service Guard.
O.S Linux RED HAT A.S 3.0

MARS
Servers 2 x RX5670 4xITANIUM 1.3Ghz
16GB RAM 2x36 GB HD , 1Gigabit 2xFC Connection
Cluster Configuration Service Guard.
O.S HPUX 11.22

Graphics Servers
Servers 2 x RX5670 4xITANIUM 1.3Ghz
16GB RAM 2x36 GB HD , 1Gigabit 2xFC Connection
Cluster Configuration Service Guard.
O.S HPUX 11.22

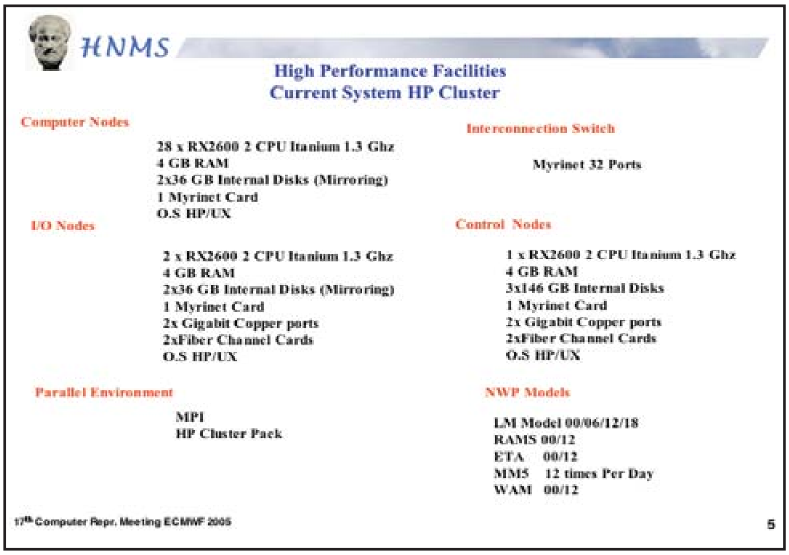
MSG Processing
5 x Intel Based Servers
DELL NAS 1TB
O.S Linux RED HAT 9.0

Radar Processing
2 x Intel Based Servers
O.S Linux RED HAT 9.0
SUSE Linux

Web Server Farm
24 Servers Intel Based XEON
1GB RAM 1x36 GB HD ,
O.S Linux RED HAT 9.0
2x SUN 480
4GB RAM 2x36 GB HD
NAS Storage
80 GB
SUN Cluster 3.0
4x SUN V120

17th Computer Repr. Meeting ECMWF 2005 4

GREECE



HNMS

High Performance Facilities
Current System HP Cluster

Computer Nodes

28 x RX2600 2 CPU Itanium 1.3 Ghz
4 GB RAM
2x36 GB Internal Disks (Mirroring)
1 Myrinet Card
O.S HP/UX

Interconnection Switch

Myrinet 32 Ports

I/O Nodes

2 x RX2600 2 CPU Itanium 1.3 Ghz
4 GB RAM
2x36 GB Internal Disks (Mirroring)
1 Myrinet Card
2x Gigabit Copper ports
2xFiber Channel Cards
O.S HP/UX

Control Nodes

1 x RX2600 2 CPU Itanium 1.3 Ghz
4 GB RAM
3x146 GB Internal Disks
1 Myrinet Card
2x Gigabit Copper ports
2xFiber Channel Cards
O.S HP/UX

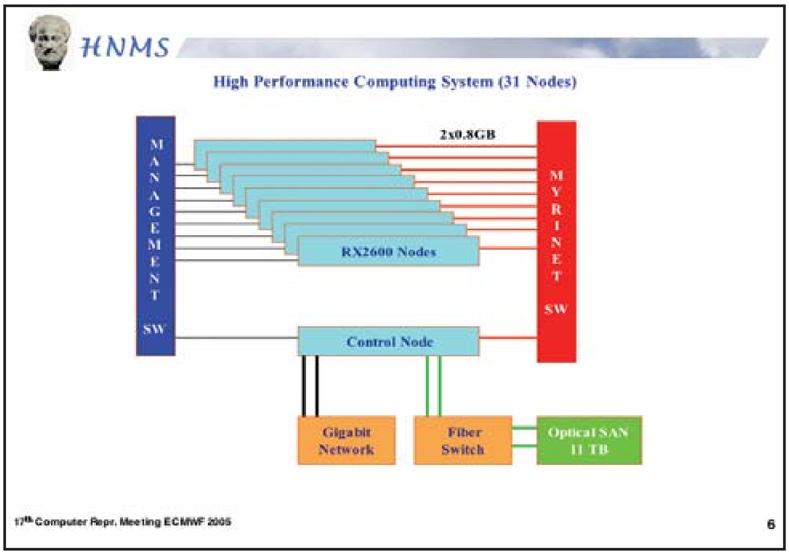
Parallel Environment

MPI
HP Cluster Pack

NWP Models

LM Model 00/06/12/18
RAMS 00/12
ETA 00/12
MM5 12 times Per Day
WAM 00/12

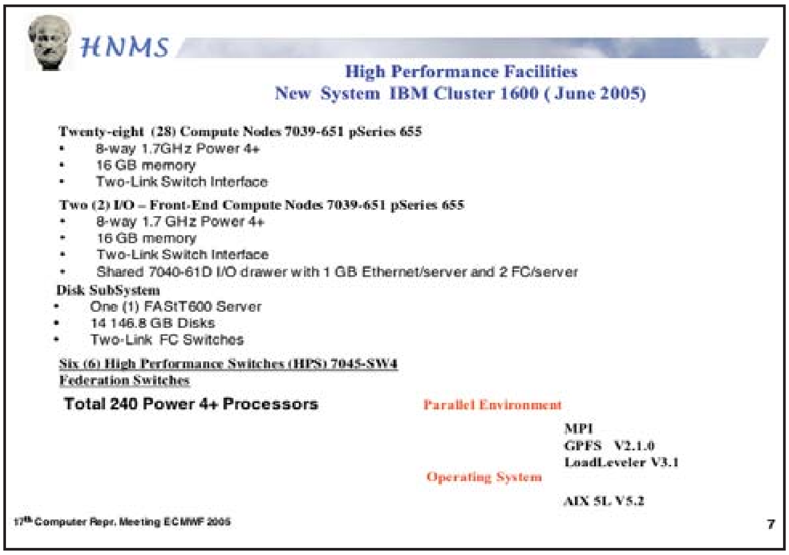
17th Computer Repr. Meeting ECMWF 2005 5



HNMS

High Performance Computing System (31 Nodes)

17th Computer Repr. Meeting ECMWF 2005 6



HNMS

High Performance Facilities
New System IBM Cluster 1600 (June 2005)

Twenty-eight (28) Compute Nodes 7039-651 pSeries 655
- 8-way 1.7GHz Power 4+
- 16 GB memory
- Two-Link Switch Interface

Two (2) I/O – Front-End Compute Nodes 7039-651 pSeries 655
- 8-way 1.7 GHz Power 4+
- 16 GB memory
- Two-Link Switch Interface
- Shared 7040-61D I/O drawer with 1 GB Ethernet/server and 2 FC/server

Disk SubSystem
- One (1) FAStT600 Server
- 14 146.8 GB Disks
- Two-Link FC Switches

Six (6) High Performance Switches (HPS) 7045-SW4
Federation Switches

Total 240 Power 4+ Processors

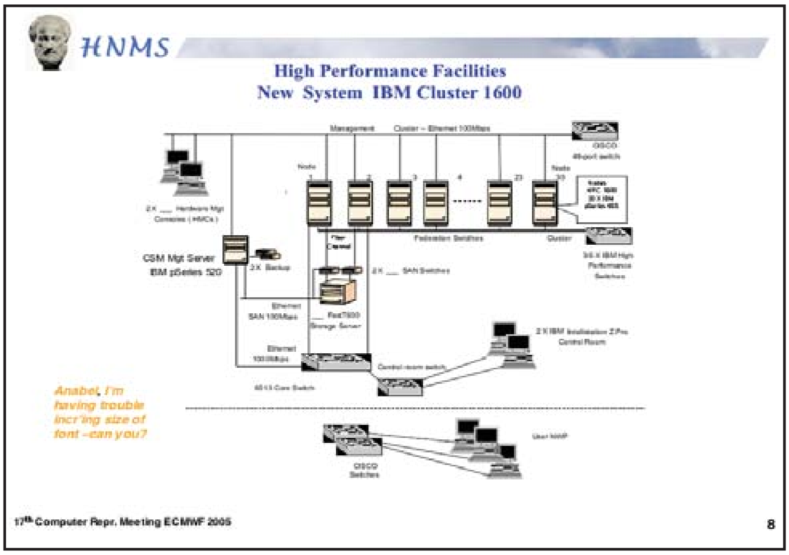
Parallel Environment

MPI
GPFS V2.1.0
LoadLeveler V3.1

Operating System

AIX 5L V5.2

17th Computer Repr. Meeting ECMWF 2005 7

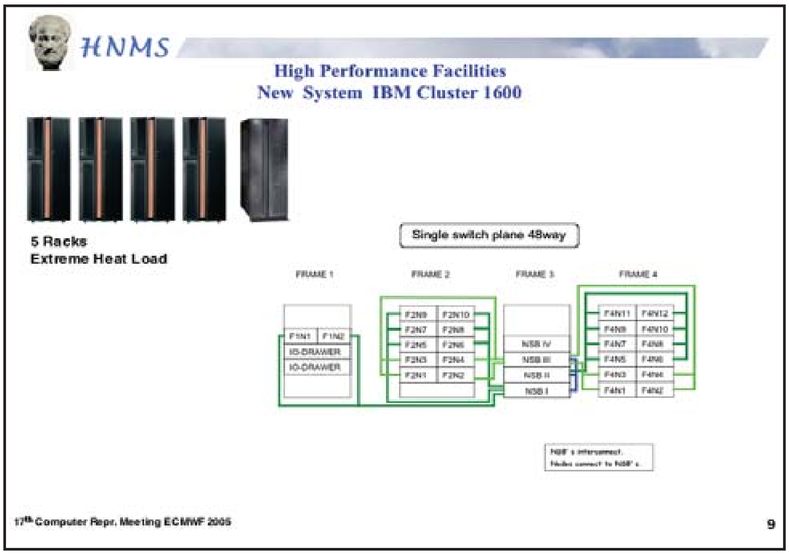


High Performance Facilities
New System IBM Cluster 1600

Anabel, I'm having trouble incr'ing size of font - can you?

17th Computer Repr. Meeting ECMWF 2005

8

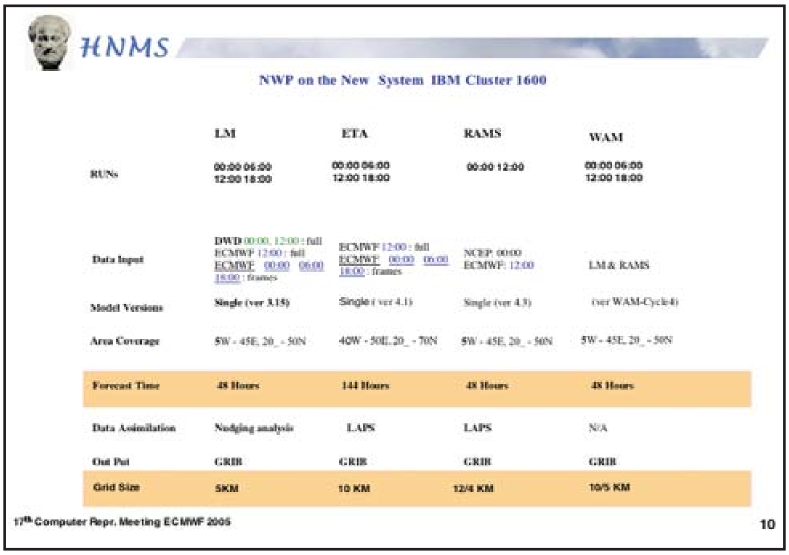


High Performance Facilities
New System IBM Cluster 1600

5 Racks
Extreme Heat Load

17th Computer Repr. Meeting ECMWF 2005

9

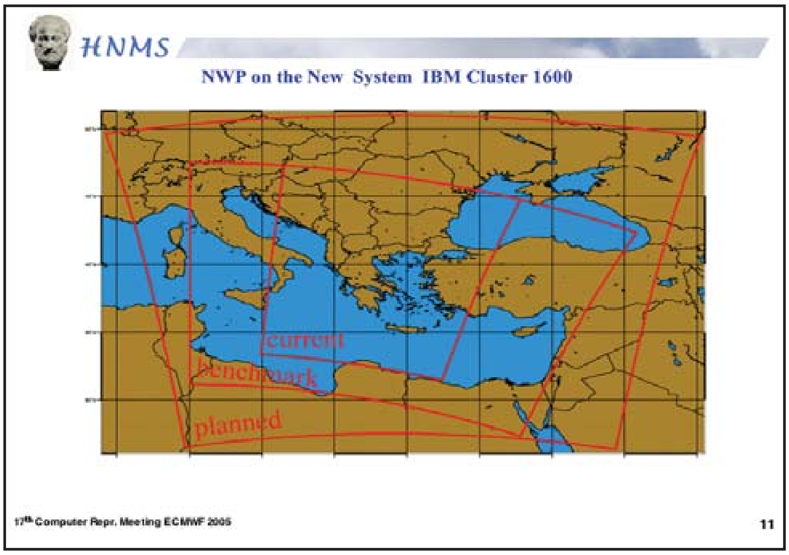


NWP on the New System IBM Cluster 1600

	LM	ETA	RAMS	WAM
RUNs	00:00 06:00 12:00 18:00	00:00 06:00 12:00 18:00	00:00 12:00	00:00 06:00 12:00 18:00
Data Input	DWD 00:00, 12:00 : full ECMWF 12:00 : full ECMWF 00:00 06:00 18:00 : frames	ECMWF 12:00 : full ECMWF 00:00 06:00 18:00 : frames	NCEP: 00:00 ECMWF: 12:00	LM & RAMS
Model Versions	Single (ver 3.15)	Single (ver 4.1)	Single (ver 4.3)	(ver WAM-Cycle4)
Area Coverage	5W - 45E, 20_ - 50N	40W - 50E, 20_ - 70N	5W - 45E, 20_ - 50N	5W - 45E, 20_ - 50N
Forecast Time	48 Hours	144 Hours	48 Hours	48 Hours
Data Asimilation	Nudging analysis	LAPS	LAPS	N/A
Out Put	GRIB	GRIB	GRIB	GRIB
Grid Size	5KM	10 KM	12/4 KM	10/5 KM

17th Computer Repr. Meeting ECMWF 2005

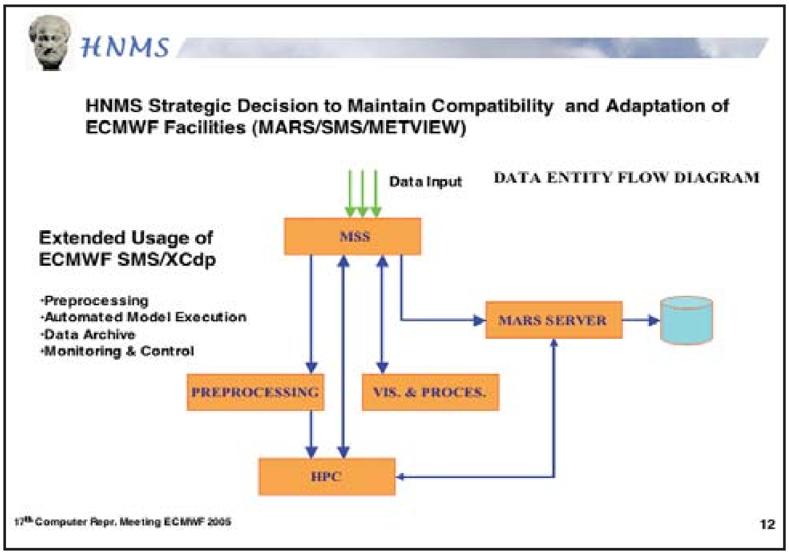
10



HNMS

NWP on the New System IBM Cluster 1600

17th Computer Repr. Meeting ECMWF 2005 11



HNMS

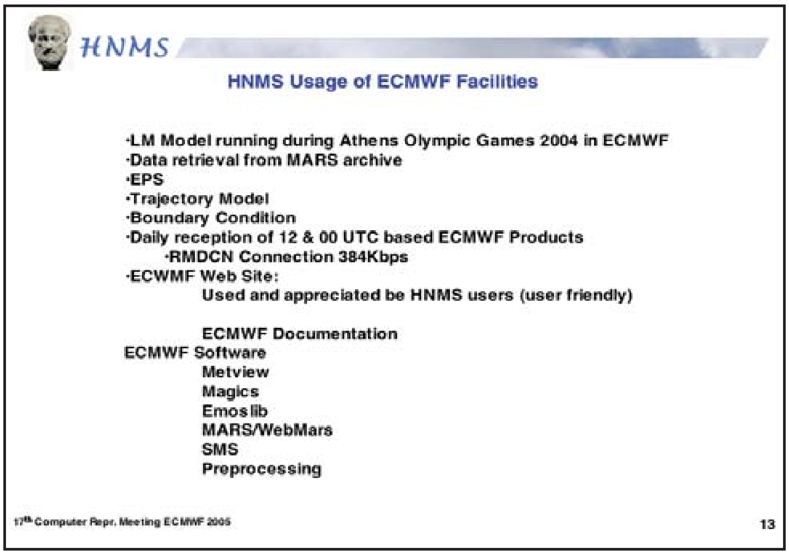
HNMS Strategic Decision to Maintain Compatibility and Adaptation of ECMWF Facilities (MARS/SMS/METVIEW)

DATA ENTITY FLOW DIAGRAM

Extended Usage of ECMWF SMS/XCdp

- Preprocessing
- Automated Model Execution
- Data Archive
- Monitoring & Control

17th Computer Repr. Meeting ECMWF 2005 12



HNMS

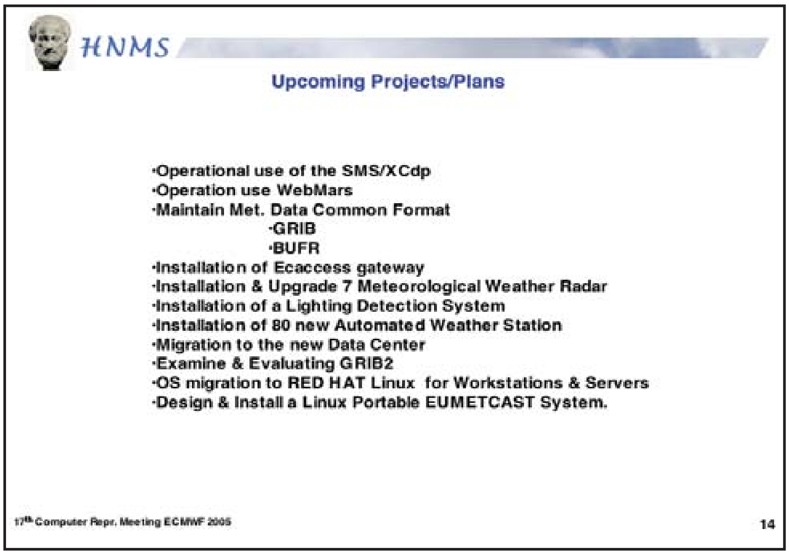
HNMS Usage of ECMWF Facilities

- LM Model running during Athens Olympic Games 2004 in ECMWF
- Data retrieval from MARS archive
- EPS
- Trajectory Model
- Boundary Condition
- Daily reception of 12 & 00 UTC based ECMWF Products
  - RMDCN Connection 384Kbps
- ECWMF Web Site:
  Used and appreciated be HNMS users (user friendly)

  ECMWF Documentation

ECMWF Software
  Metview
  Magics
  Emos lib
  MARS/WebMars
  SMS
  Preprocessing

17th Computer Repr. Meeting ECMWF 2005 13



HNMS

Upcoming Projects/Plans

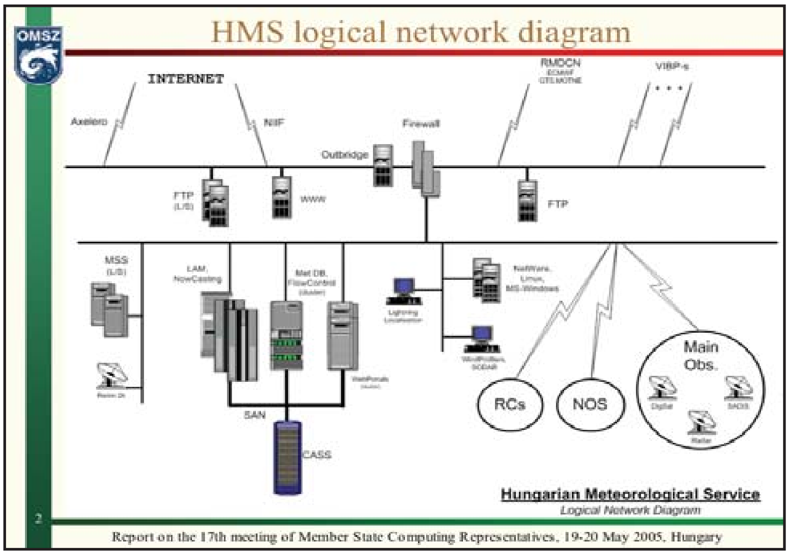
- Operational use of the SMS/XCdp
- Operation use WebMars
- Maintain Met. Data Common Format
  - GRIB
  - BUFR
- Installation of Ecaccess gateway
- Installation & Upgrade 7 Meteorological Weather Radar
- Installation of a Lighting Detection System
- Installation of 80 new Automated Weather Station
- Migration to the new Data Center
- Examine & Evaluating GRIB2
- OS migration to RED HAT Linux for Workstations & Servers
- Design & Install a Linux Portable EUMETCAST System.

17th Computer Repr. Meeting ECMWF 2005

14

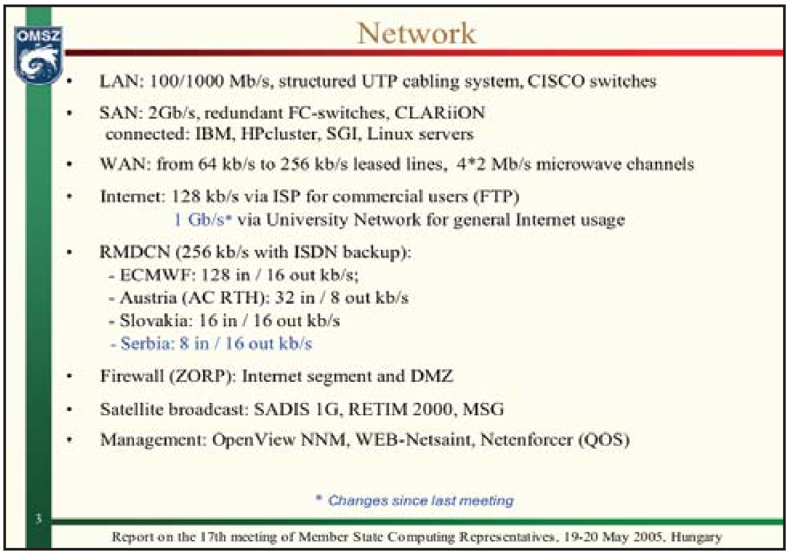
HUNGARY

László Tölgyesi – Hungarian Meteorological Service



HMS logical network diagram

Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



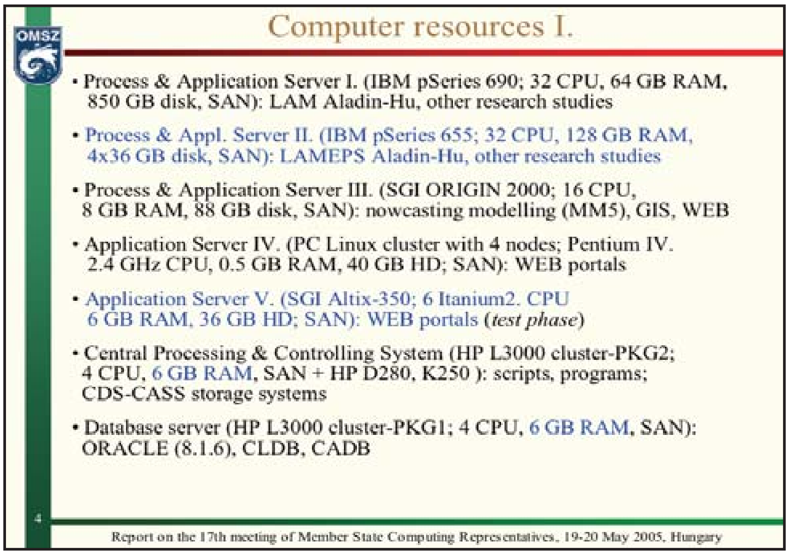
Network

- LAN: 100/1000 Mb/s, structured UTP cabling system, CISCO switches
- SAN: 2Gb/s, redundant FC-switches, CLARiiON connected: IBM, HPcluster, SGI, Linux servers
- WAN: from 64 kb/s to 256 kb/s leased lines, 4*2 Mb/s microwave channels
- Internet: 128 kb/s via ISP for commercial users (FTP)
  1 Gb/s* via University Network for general Internet usage
- RMDCN (256 kb/s with ISDN backup):
  - ECMWF: 128 in / 16 out kb/s;
  - Austria (AC RTH): 32 in / 8 out kb/s
  - Slovakia: 16 in / 16 out kb/s
  - Serbia: 8 in / 16 out kb/s
- Firewall (ZORP): Internet segment and DMZ
- Satellite broadcast: SADIS 1G, RETIM 2000, MSG
- Management: OpenView NNM, WEB-Netsaint, Netenforcer (QOS)

** Changes since last meeting*

3

Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary

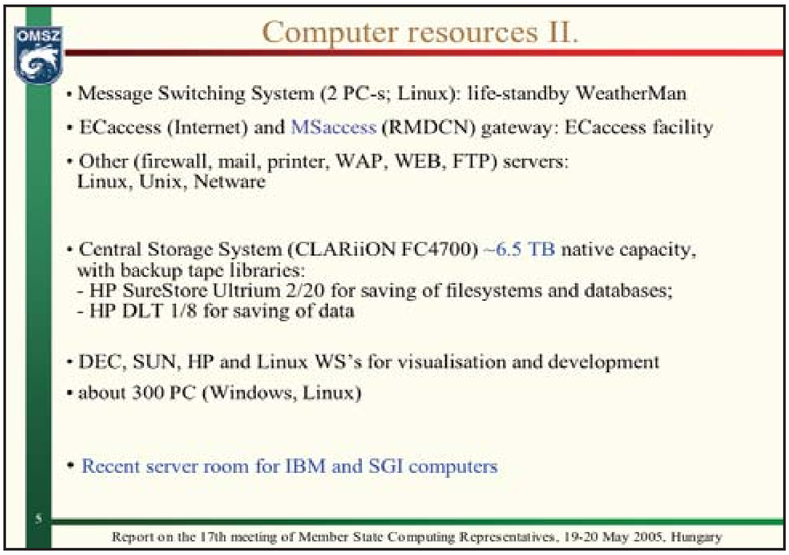


Computer resources I.

- Process & Application Server I. (IBM pSeries 690; 32 CPU, 64 GB RAM, 850 GB disk, SAN): LAM Aladin-Hu, other research studies
- Process & Appl. Server II. (IBM pSeries 655; 32 CPU, 128 GB RAM, 4x36 GB disk, SAN): LAMEPS Aladin-Hu, other research studies
- Process & Application Server III. (SGI ORIGIN 2000; 16 CPU, 8 GB RAM, 88 GB disk, SAN): nowcasting modelling (MM5), GIS, WEB
- Application Server IV. (PC Linux cluster with 4 nodes; Pentium IV. 2.4 GHz CPU, 0.5 GB RAM, 40 GB HD; SAN): WEB portals
- Application Server V. (SGI Altix-350; 6 Itanium2. CPU 6 GB RAM, 36 GB HD; SAN): WEB portals (*test phase*)
- Central Processing & Controlling System (HP L3000 cluster-PKG2; 4 CPU, 6 GB RAM, SAN + HP D280, K250): scripts, programs; CDS-CASS storage systems
- Database server (HP L3000 cluster-PKG1; 4 CPU, 6 GB RAM, SAN): ORACLE (8.1.6), CLDB, CADB

4

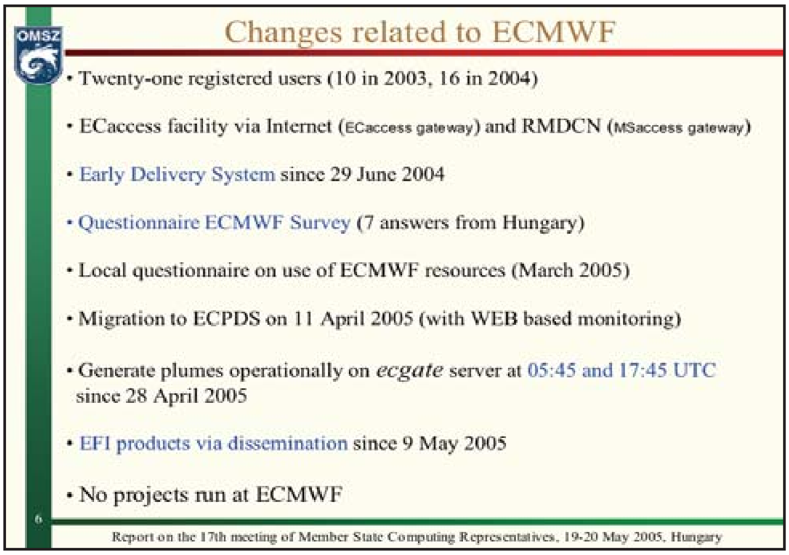
Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



Computer resources II.

- Message Switching System (2 PC-s; Linux): life-standby WeatherMan
- ECaccess (Internet) and MSaccess (RMDCN) gateway: ECaccess facility
- Other (firewall, mail, printer, WAP, WEB, FTP) servers:
  Linux, Unix, Netware
- Central Storage System (CLARiiON FC4700) ~6.5 TB native capacity,
  with backup tape libraries:
  - HP SureStore Ultrium 2/20 for saving of filesystems and databases;
  - HP DLT 1/8 for saving of data
- DEC, SUN, HP and Linux WS's for visualisation and development
- about 300 PC (Windows, Linux)
- Recent server room for IBM and SGI computers

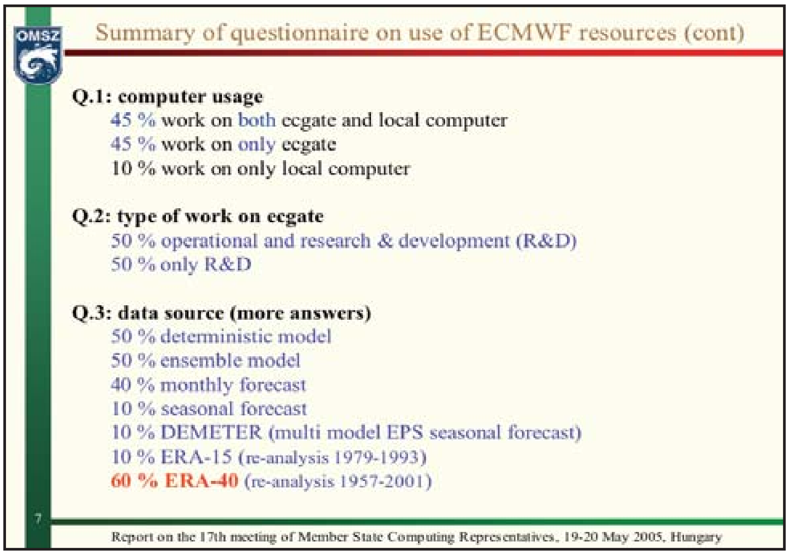
5 Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



Changes related to ECMWF

- Twenty-one registered users (10 in 2003, 16 in 2004)
- ECaccess facility via Internet (ECaccess gateway) and RMDCN (MSaccess gateway)
- Early Delivery System since 29 June 2004
- Questionnaire ECMWF Survey (7 answers from Hungary)
- Local questionnaire on use of ECMWF resources (March 2005)
- Migration to ECPDS on 11 April 2005 (with WEB based monitoring)
- Generate plumes operationally on *ecgate* server at 05:45 and 17:45 UTC since 28 April 2005
- EFI products via dissemination since 9 May 2005
- No projects run at ECMWF

6 Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



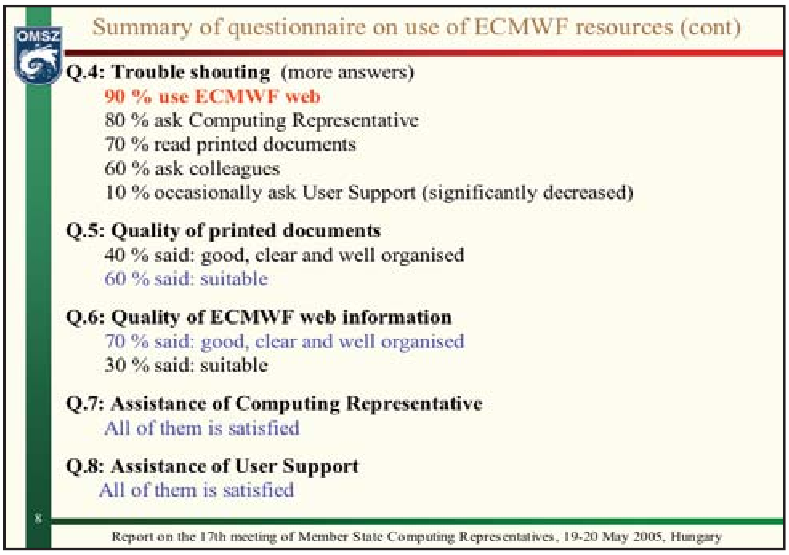
Summary of questionnaire on use of ECMWF resources (cont)

Q.1: computer usage
45 % work on both ecgate and local computer
45 % work on only ecgate
10 % work on only local computer

Q.2: type of work on ecgate
50 % operational and research & development (R&D)
50 % only R&D

Q.3: data source (more answers)
50 % deterministic model
50 % ensemble model
40 % monthly forecast
10 % seasonal forecast
10 % DEMETER (multi model EPS seasonal forecast)
10 % ERA-15 (re-analysis 1979-1993)
60 % ERA-40 (re-analysis 1957-2001)

7 Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



OMSZ

Summary of questionnaire on use of ECMWF resources (cont)

Q.4: Trouble shouting (more answers)

- **90 % use ECMWF web**
- 80 % ask Computing Representative
- 70 % read printed documents
- 60 % ask colleagues
- 10 % occasionally ask User Support (significantly decreased)

Q.5: Quality of printed documents

- 40 % said: good, clear and well organised
- 60 % said: suitable

Q.6: Quality of ECMWF web information

- 70 % said: good, clear and well organised
- 30 % said: suitable

Q.7: Assistance of Computing Representative

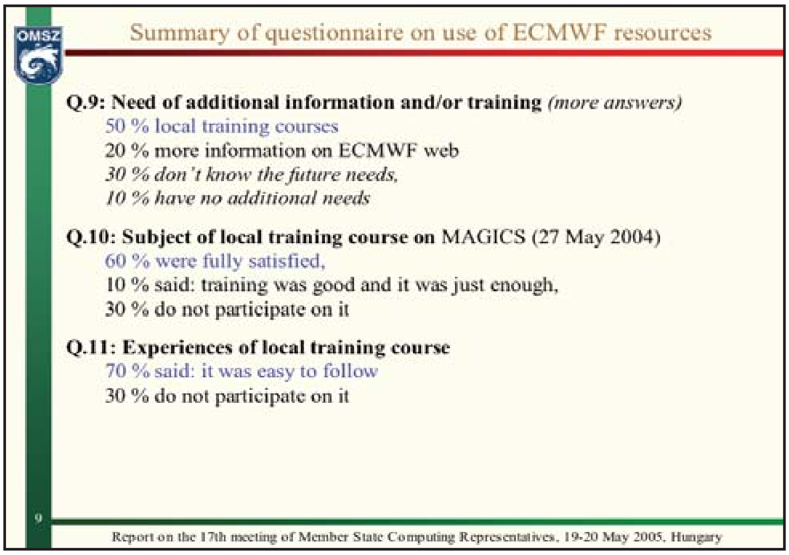
All of them is satisfied

Q.8: Assistance of User Support

All of them is satisfied

8

Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



OMSZ

Summary of questionnaire on use of ECMWF resources

Q.9: Need of additional information and/or training *(more answers)*

- 50 % local training courses
- 20 % more information on ECMWF web
- *30 % don't know the future needs,*
- *10 % have no additional needs*

Q.10: Subject of local training course on MAGICS (27 May 2004)

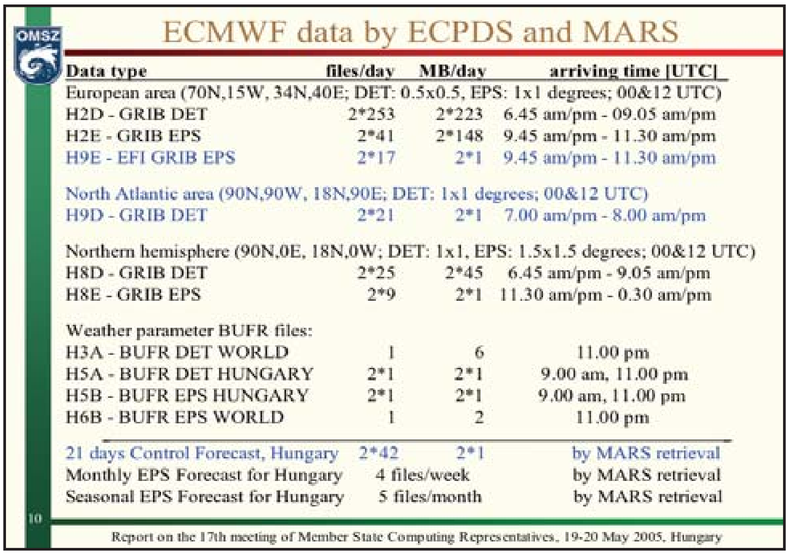
- 60 % were fully satisfied,
- 10 % said: training was good and it was just enough,
- 30 % do not participate on it

Q.11: Experiences of local training course

- 70 % said: it was easy to follow
- 30 % do not participate on it

9

Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



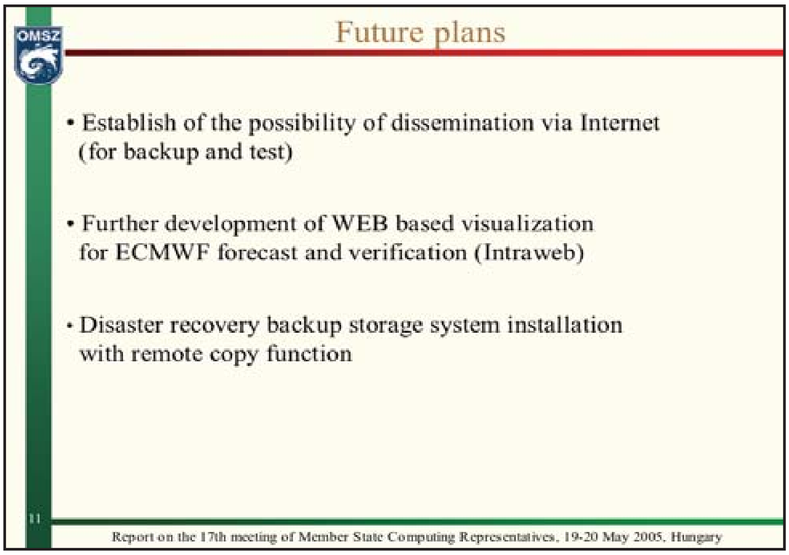
OMSZ

ECMWF data by ECPDS and MARS

Data type	files/day	MB/day	arriving time [UTC]
European area (70N,15W, 34N,40E; DET: 0.5x0.5, EPS: 1x1 degrees; 00&12 UTC)			
H2D - GRIB DET	2*253	2*223	6.45 am/pm - 09.05 am/pm
H2E - GRIB EPS	2*41	2*148	9.45 am/pm - 11.30 am/pm
H9E - EFI GRIB EPS	2*17	2*1	9.45 am/pm - 11.30 am/pm
North Atlantic area (90N,90W, 18N,90E; DET: 1x1 degrees; 00&12 UTC)			
H9D - GRIB DET	2*21	2*1	7.00 am/pm - 8.00 am/pm
Northern hemisphere (90N,0E, 18N,0W; DET: 1x1, EPS: 1.5x1.5 degrees; 00&12 UTC)			
H8D - GRIB DET	2*25	2*45	6.45 am/pm - 9.05 am/pm
H8E - GRIB EPS	2*9	2*1	11.30 am/pm - 0.30 am/pm
Weather parameter BUFR files:			
H3A - BUFR DET WORLD	1	6	11.00 pm
H5A - BUFR DET HUNGARY	2*1	2*1	9.00 am, 11.00 pm
H5B - BUFR EPS HUNGARY	2*1	2*1	9.00 am, 11.00 pm
H6B - BUFR EPS WORLD	1	2	11.00 pm
21 days Control Forecast, Hungary	2*42	2*1	by MARS retrieval
Monthly EPS Forecast for Hungary	4 files/week		by MARS retrieval
Seasonal EPS Forecast for Hungary	5 files/month		by MARS retrieval

10

Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary



Future plans

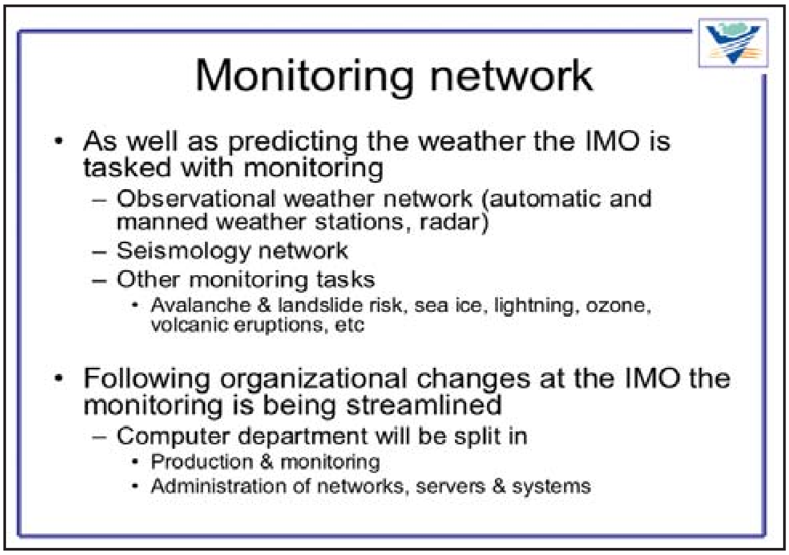
- Establish of the possibility of dissemination via Internet (for backup and test)
- Further development of WEB based visualization for ECMWF forecast and verification (Intraweb)
- Disaster recovery backup storage system installation with remote copy function

11

Report on the 17th meeting of Member State Computing Representatives, 19-20 May 2005, Hungary

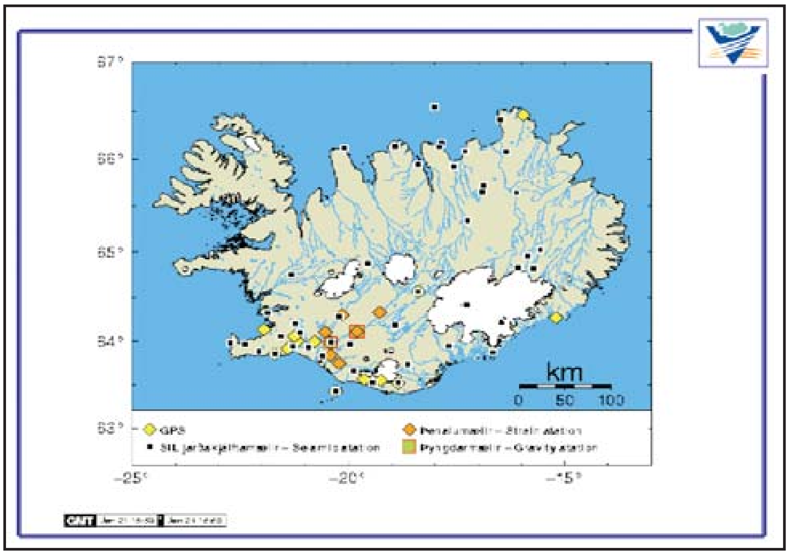
ICELAND ICELAND

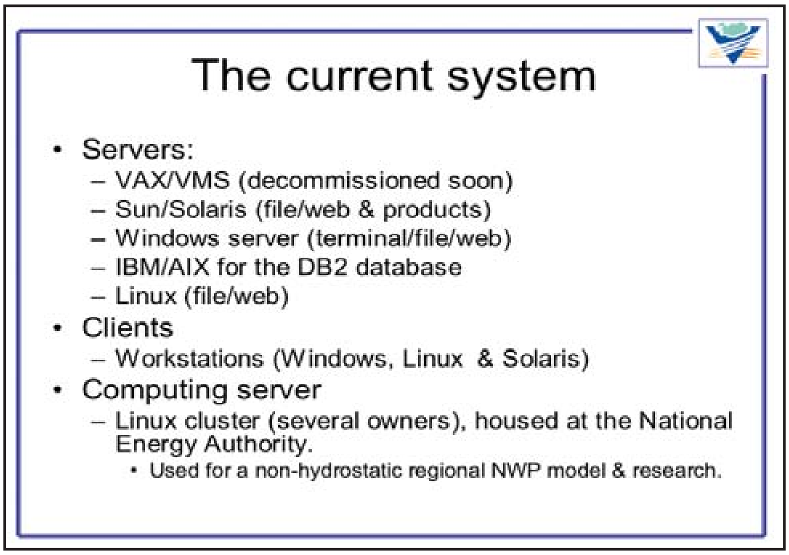
Halldór Björnsson – Iceland Meteorological Office



Monitoring network

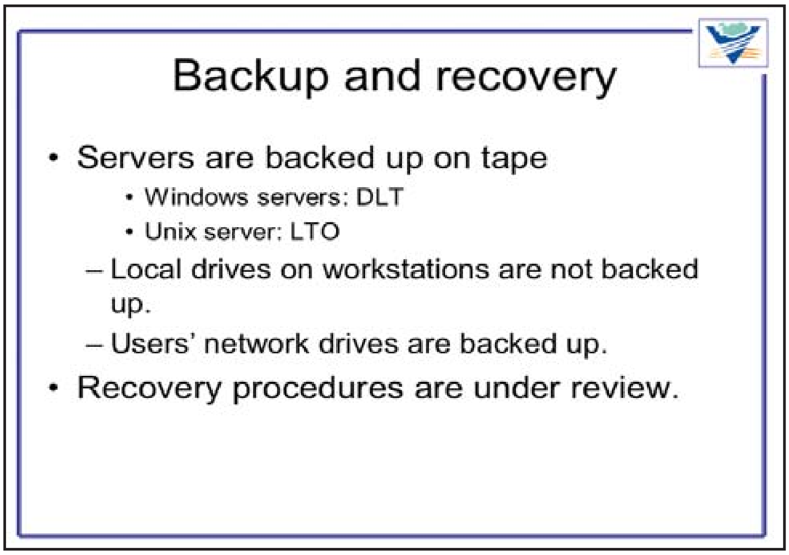
- As well as predicting the weather the IMO is tasked with monitoring
  - Observational weather network (automatic and manned weather stations, radar)
  - Seismology network
  - Other monitoring tasks
    - Avalanche & landslide risk, sea ice, lightning, ozone, volcanic eruptions, etc
- Following organizational changes at the IMO the monitoring is being streamlined
  - Computer department will be split in
    - Production & monitoring
    - Administration of networks, servers & systems


67°
66°
65°
64°
63°
−25°
−20°
−15°
km
0
50
100
GPS
SIL jarðskjálftamælir – Seismic station
Þenslumælir – Strain station
Þyngdarmælir – Gravity station



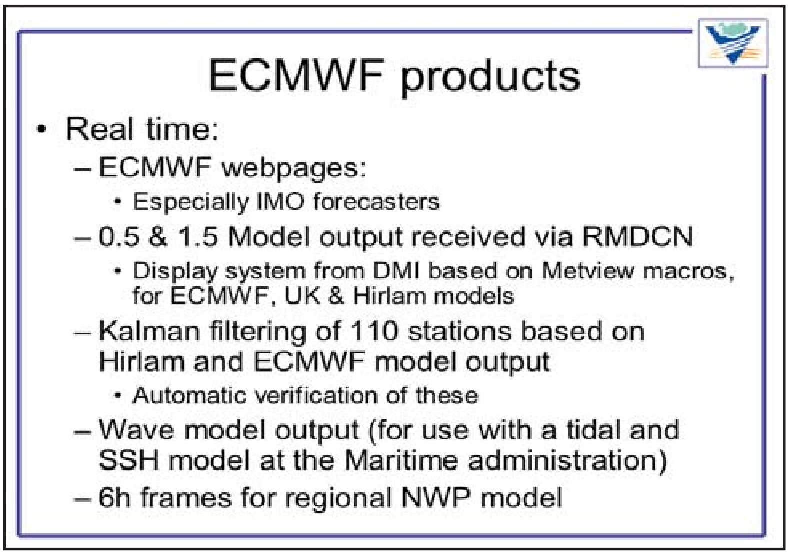
The current system

- Servers:
  - VAX/VMS (decommissioned soon)
  - Sun/Solaris (file/web & products)
  - Windows server (terminal/file/web)
  - IBM/AIX for the DB2 database
  - Linux (file/web)
- Clients
  - Workstations (Windows, Linux & Solaris)
- Computing server
  - Linux cluster (several owners), housed at the National Energy Authority.
    - Used for a non-hydrostatic regional NWP model & research.



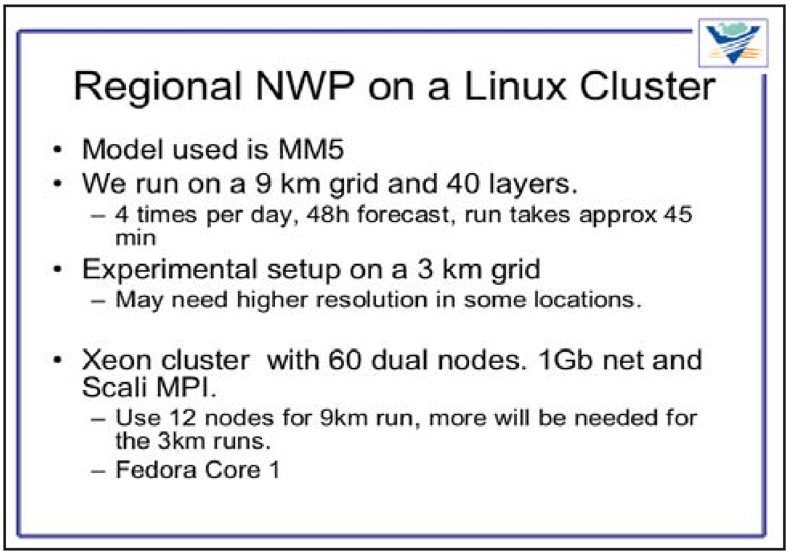
Backup and recovery

- Servers are backed up on tape
    - Windows servers: DLT
    - Unix server: LTO
  - Local drives on workstations are not backed up.
  - Users' network drives are backed up.
- Recovery procedures are under review.



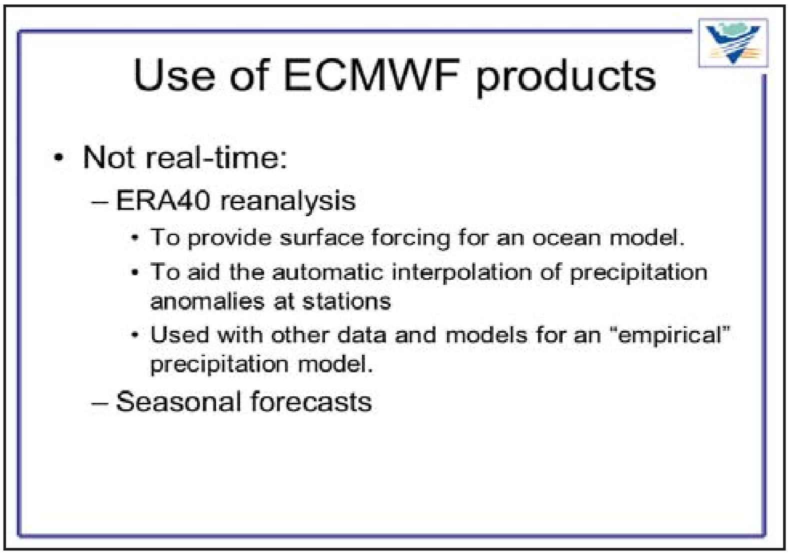
ECMWF products

- Real time:
  - ECMWF webpages:
    - Especially IMO forecasters
  - 0.5 & 1.5 Model output received via RMDCN
    - Display system from DMI based on Metview macros, for ECMWF, UK & Hirlam models
  - Kalman filtering of 110 stations based on Hirlam and ECMWF model output
    - Automatic verification of these
  - Wave model output (for use with a tidal and SSH model at the Maritime administration)
  - 6h frames for regional NWP model



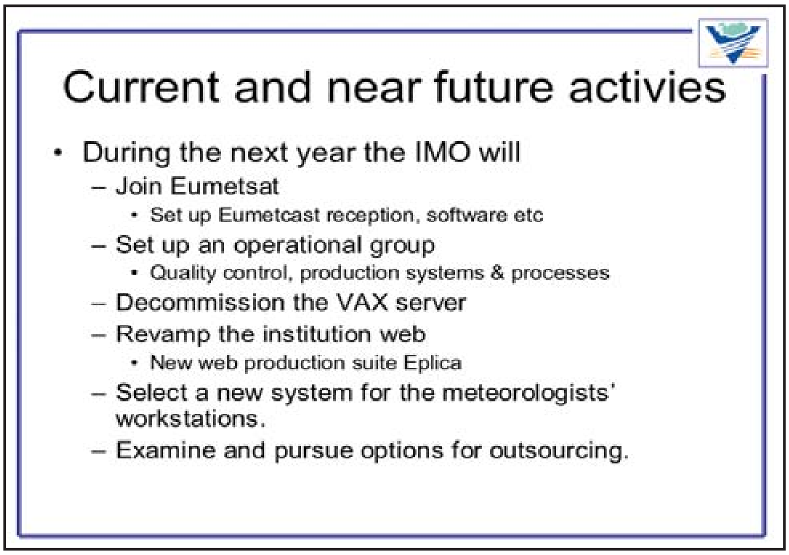
Regional NWP on a Linux Cluster

- Model used is MM5
- We run on a 9 km grid and 40 layers.
  - 4 times per day, 48h forecast, run takes approx 45 min
- Experimental setup on a 3 km grid
  - May need higher resolution in some locations.
- Xeon cluster with 60 dual nodes. 1Gb net and Scali MPI.
  - Use 12 nodes for 9km run, more will be needed for the 3km runs.
  - Fedora Core 1



Use of ECMWF products

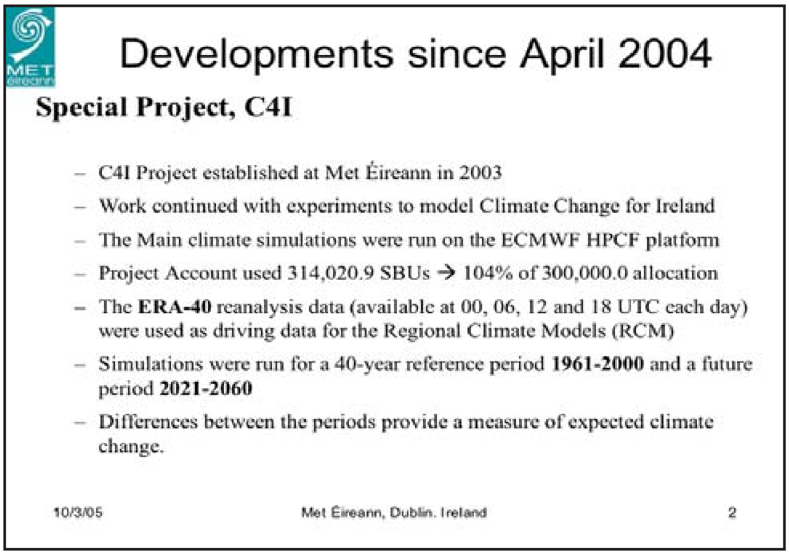
- Not real-time:
  - ERA40 reanalysis
    - To provide surface forcing for an ocean model.
    - To aid the automatic interpolation of precipitation anomalies at stations
    - Used with other data and models for an “empirical” precipitation model.
  - Seasonal forecasts



Current and near future activies

- During the next year the IMO will
  - Join Eumetsat
    - Set up Eumetcast reception, software etc
  - Set up an operational group
    - Quality control, production systems & processes
  - Decommission the VAX server
  - Revamp the institution web
    - New web production suite Eplica
  - Select a new system for the meteorologists’ workstations.
  - Examine and pursue options for outsourcing.

Paul Halton – Met Éireann, Dublin

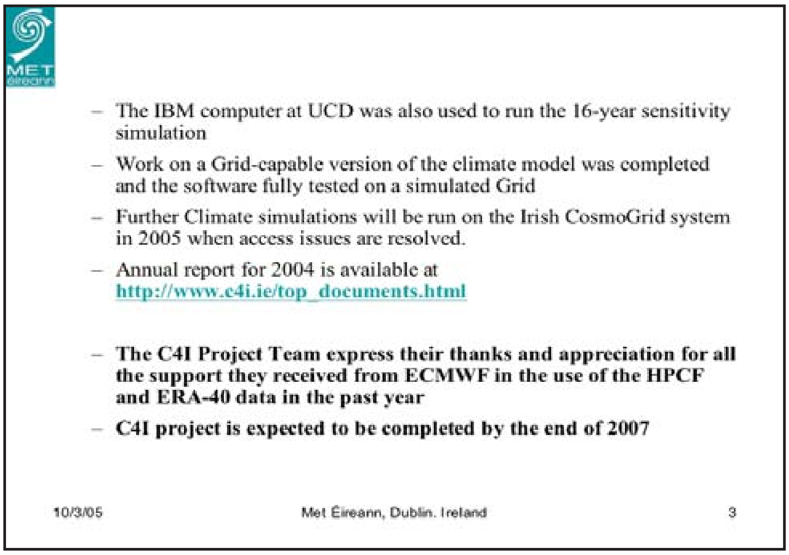


Developments since April 2004

Special Project, C4I

- C4I Project established at Met Éireann in 2003
- Work continued with experiments to model Climate Change for Ireland
- The Main climate simulations were run on the ECMWF HPCF platform
- Project Account used 314,020.9 SBUs → 104% of 300,000.0 allocation
- The **ERA-40** reanalysis data (available at 00, 06, 12 and 18 UTC each day) were used as driving data for the Regional Climate Models (RCM)
- Simulations were run for a 40-year reference period **1961-2000** and a future period **2021-2060**
- Differences between the periods provide a measure of expected climate change.

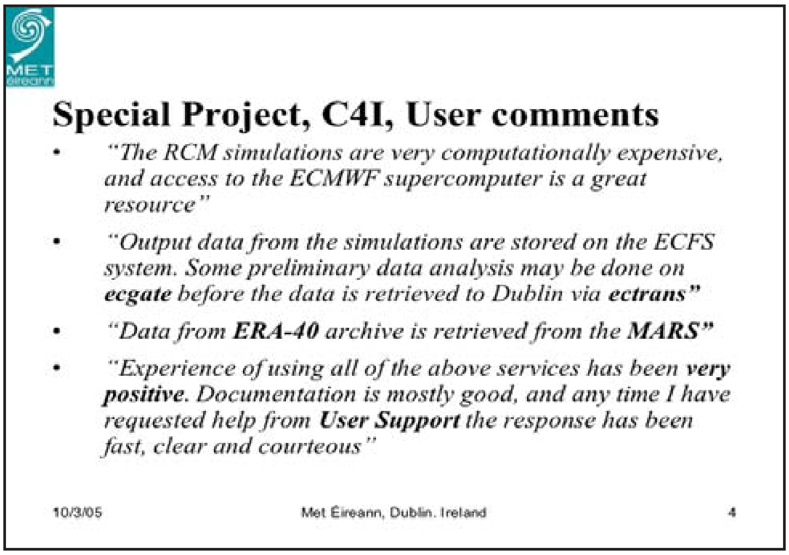
10/3/05 Met Éireann, Dublin. Ireland 2



- The IBM computer at UCD was also used to run the 16-year sensitivity simulation
- Work on a Grid-capable version of the climate model was completed and the software fully tested on a simulated Grid
- Further Climate simulations will be run on the Irish CosmoGrid system in 2005 when access issues are resolved.
- Annual report for 2004 is available at **http://www.c4i.ie/top_documents.html**

- **The C4I Project Team express their thanks and appreciation for all the support they received from ECMWF in the use of the HPCF and ERA-40 data in the past year**
- **C4I project is expected to be completed by the end of 2007**

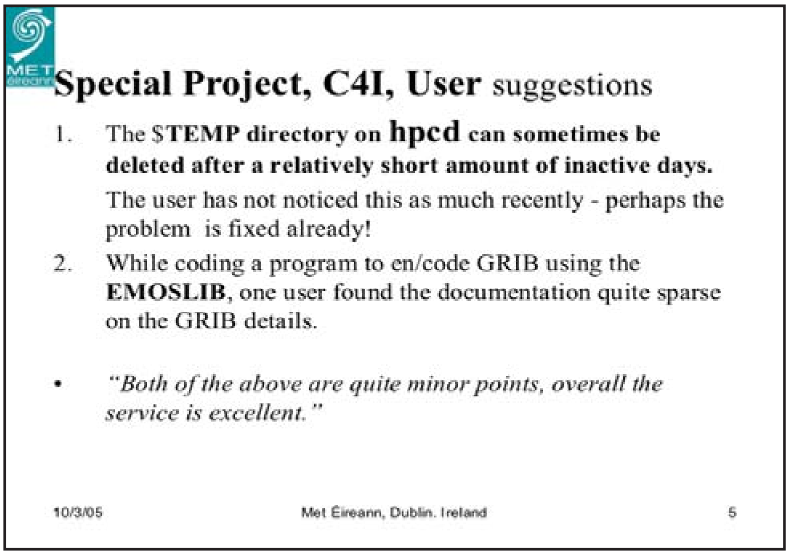
10/3/05 Met Éireann, Dublin. Ireland 3



Special Project, C4I, User comments

- *"The RCM simulations are very computationally expensive, and access to the ECMWF supercomputer is a great resource"*
- *"Output data from the simulations are stored on the ECFS system. Some preliminary data analysis may be done on **ecgate** before the data is retrieved to Dublin via **ectrans**"*
- *"Data from **ERA-40** archive is retrieved from the **MARS**"*
- *"Experience of using all of the above services has been **very positive**. Documentation is mostly good, and any time I have requested help from **User Support** the response has been fast, clear and courteous"*

10/3/05 Met Éireann, Dublin. Ireland 4



Special Project, C4I, User suggestions

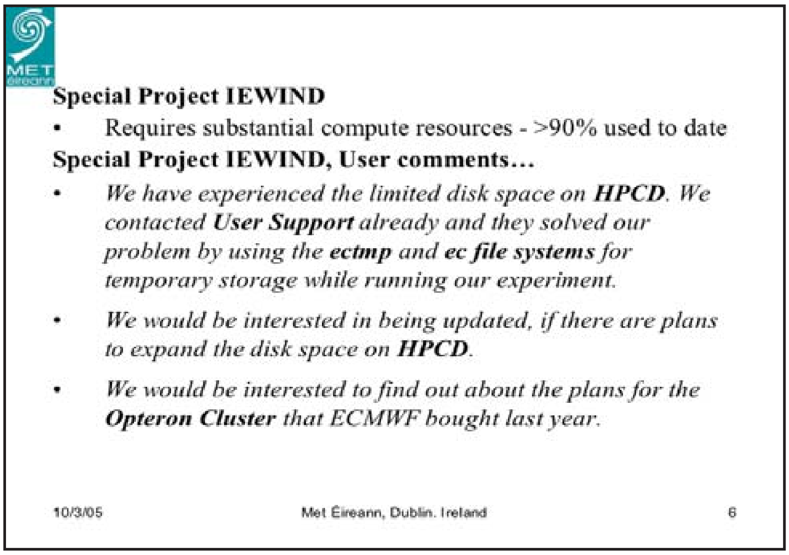
1. **The $TEMP directory on hpcd can sometimes be deleted after a relatively short amount of inactive days.**

   The user has not noticed this as much recently - perhaps the problem is fixed already!

2. While coding a program to en/code GRIB using the **EMOSLIB**, one user found the documentation quite sparse on the GRIB details.

- *"Both of the above are quite minor points, overall the service is excellent."*

10/3/05 Met Éireann, Dublin. Ireland 5



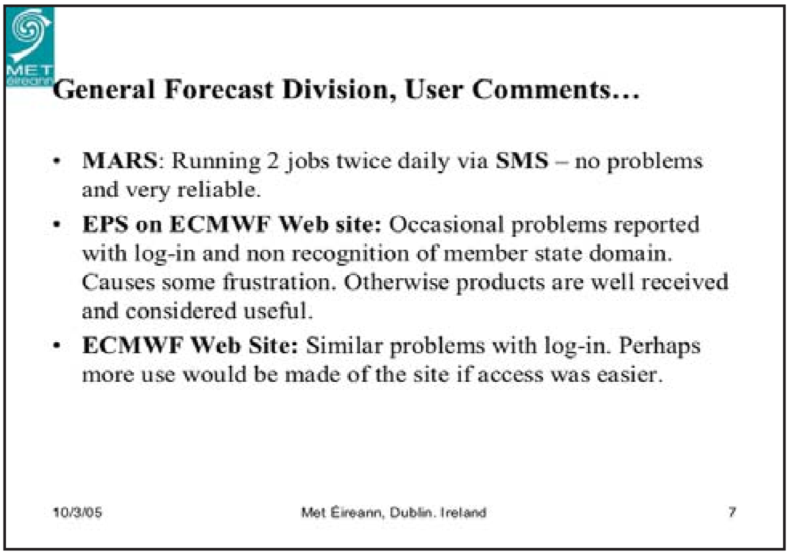
Special Project IEWIND

- Requires substantial compute resources - >90% used to date

Special Project IEWIND, User comments…

- *We have experienced the limited disk space on **HPCD**. We contacted **User Support** already and they solved our problem by using the **ectmp** and **ec file systems** for temporary storage while running our experiment.*
- *We would be interested in being updated, if there are plans to expand the disk space on **HPCD**.*
- *We would be interested to find out about the plans for the **Opteron Cluster** that ECMWF bought last year.*

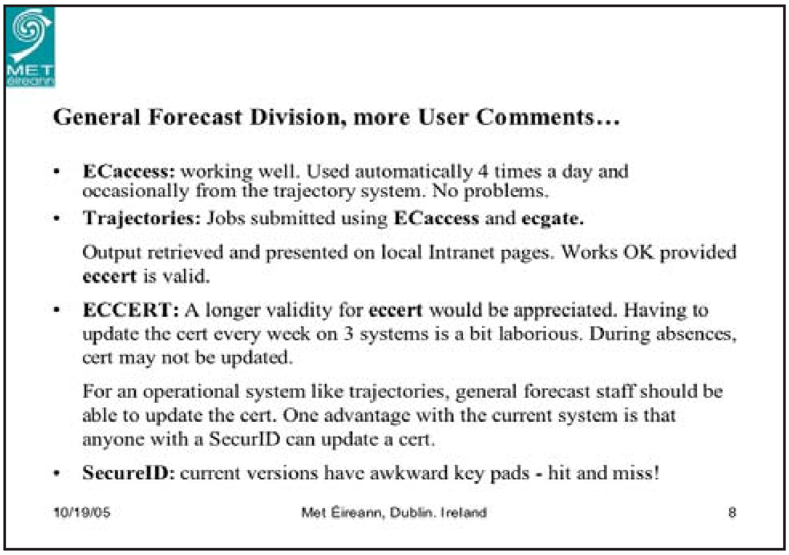
10/3/05 Met Éireann, Dublin. Ireland 6



General Forecast Division, User Comments…

- **MARS**: Running 2 jobs twice daily via **SMS** – no problems and very reliable.
- **EPS on ECMWF Web site:** Occasional problems reported with log-in and non recognition of member state domain. Causes some frustration. Otherwise products are well received and considered useful.
- **ECMWF Web Site:** Similar problems with log-in. Perhaps more use would be made of the site if access was easier.

10/3/05 Met Éireann, Dublin. Ireland 7



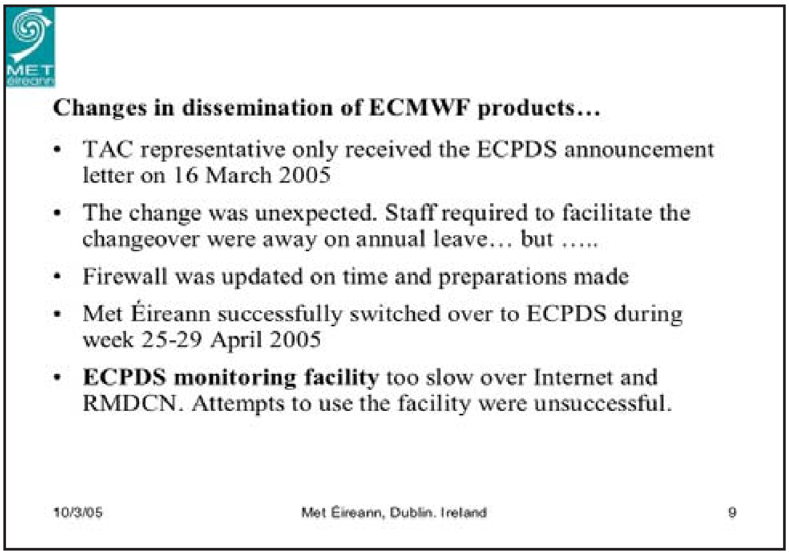
General Forecast Division, more User Comments…

- **ECaccess:** working well. Used automatically 4 times a day and occasionally from the trajectory system. No problems.
- **Trajectories:** Jobs submitted using **ECaccess** and **ecgate.**

  Output retrieved and presented on local Intranet pages. Works OK provided **eccert** is valid.
- **ECCERT:** A longer validity for **eccert** would be appreciated. Having to update the cert every week on 3 systems is a bit laborious. During absences, cert may not be updated.

  For an operational system like trajectories, general forecast staff should be able to update the cert. One advantage with the current system is that anyone with a SecurID can update a cert.
- **SecureID:** current versions have awkward key pads - hit and miss!

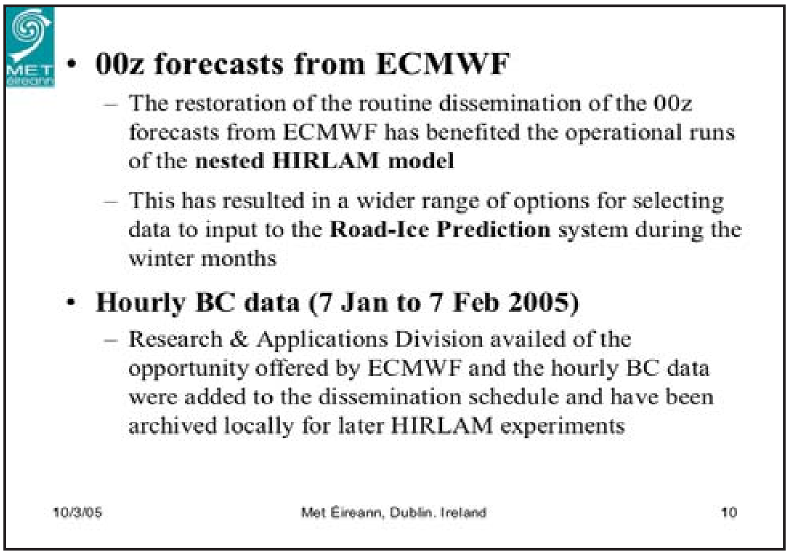
10/19/05 Met Éireann, Dublin. Ireland 8



Changes in dissemination of ECMWF products…

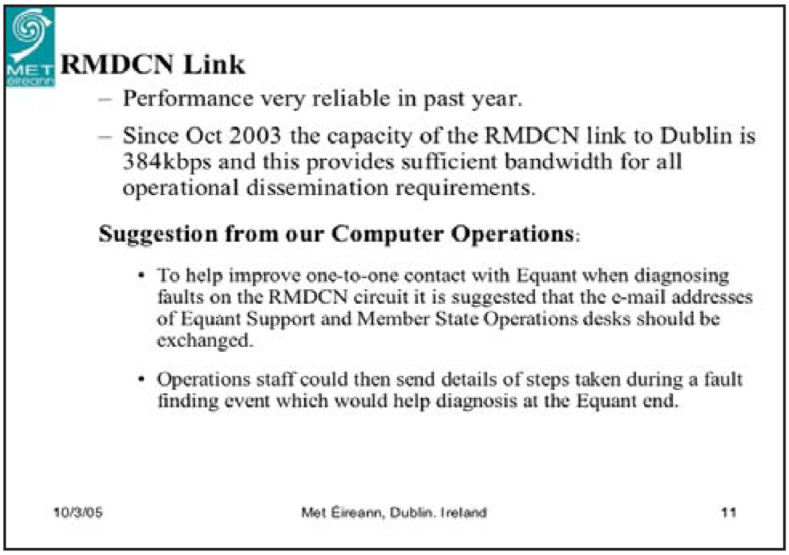
- TAC representative only received the ECPDS announcement letter on 16 March 2005
- The change was unexpected. Staff required to facilitate the changeover were away on annual leave… but …..
- Firewall was updated on time and preparations made
- Met Éireann successfully switched over to ECPDS during week 25-29 April 2005
- **ECPDS monitoring facility** too slow over Internet and RMDCN. Attempts to use the facility were unsuccessful.

10/3/05 Met Éireann, Dublin. Ireland 9



- **00z forecasts from ECMWF**
  - The restoration of the routine dissemination of the 00z forecasts from ECMWF has benefited the operational runs of the **nested HIRLAM model**
  - This has resulted in a wider range of options for selecting data to input to the **Road-Ice Prediction** system during the winter months
- **Hourly BC data (7 Jan to 7 Feb 2005)**
  - Research & Applications Division availed of the opportunity offered by ECMWF and the hourly BC data were added to the dissemination schedule and have been archived locally for later HIRLAM experiments

10/3/05 Met Éireann, Dublin. Ireland 10



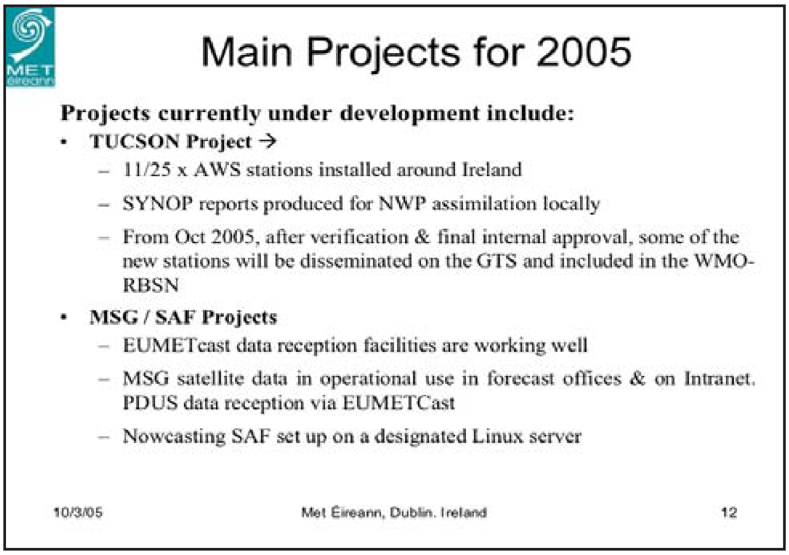
RMDCN Link

- Performance very reliable in past year.
- Since Oct 2003 the capacity of the RMDCN link to Dublin is 384kbps and this provides sufficient bandwidth for all operational dissemination requirements.

Suggestion from our Computer Operations:

- To help improve one-to-one contact with Equant when diagnosing faults on the RMDCN circuit it is suggested that the e-mail addresses of Equant Support and Member State Operations desks should be exchanged.
- Operations staff could then send details of steps taken during a fault finding event which would help diagnosis at the Equant end.

10/3/05 Met Éireann, Dublin. Ireland 11

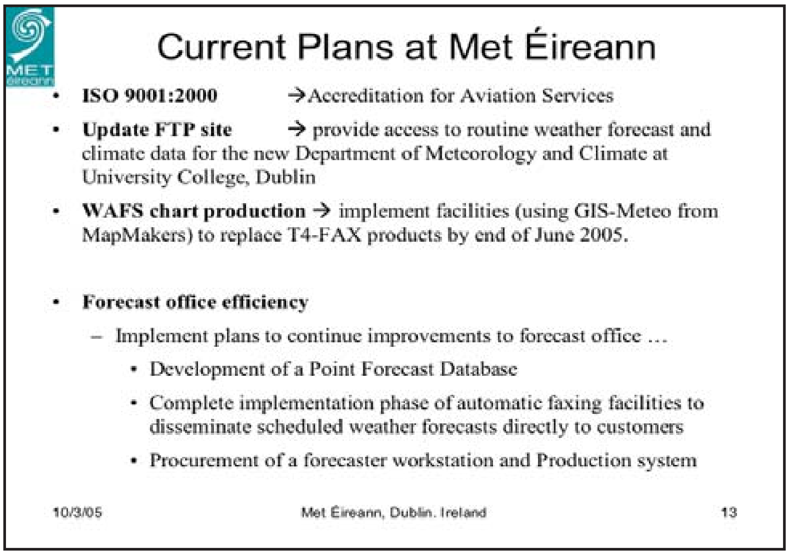


Main Projects for 2005

Projects currently under development include:

- **TUCSON Project →**
  - 11/25 x AWS stations installed around Ireland
  - SYNOP reports produced for NWP assimilation locally
  - From Oct 2005, after verification & final internal approval, some of the new stations will be disseminated on the GTS and included in the WMO-RBSN
- **MSG / SAF Projects**
  - EUMETcast data reception facilities are working well
  - MSG satellite data in operational use in forecast offices & on Intranet. PDUS data reception via EUMETCast
  - Nowcasting SAF set up on a designated Linux server

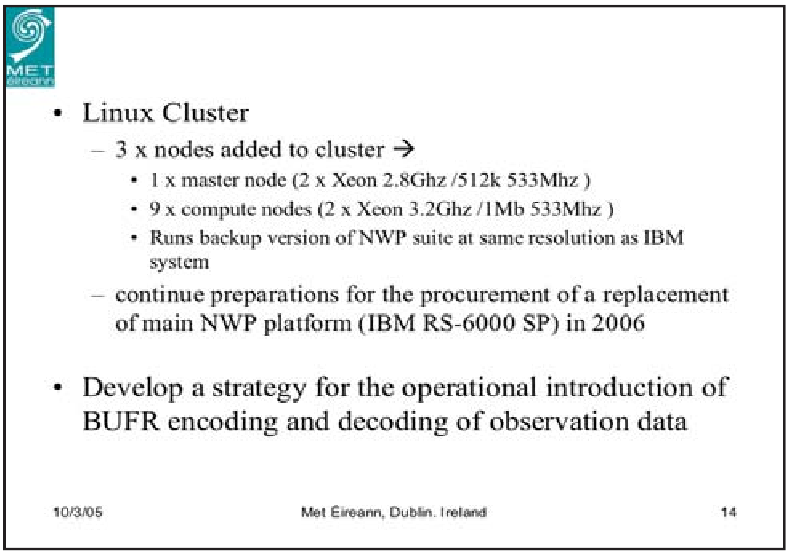
10/3/05 Met Éireann, Dublin. Ireland 12



Current Plans at Met Éireann

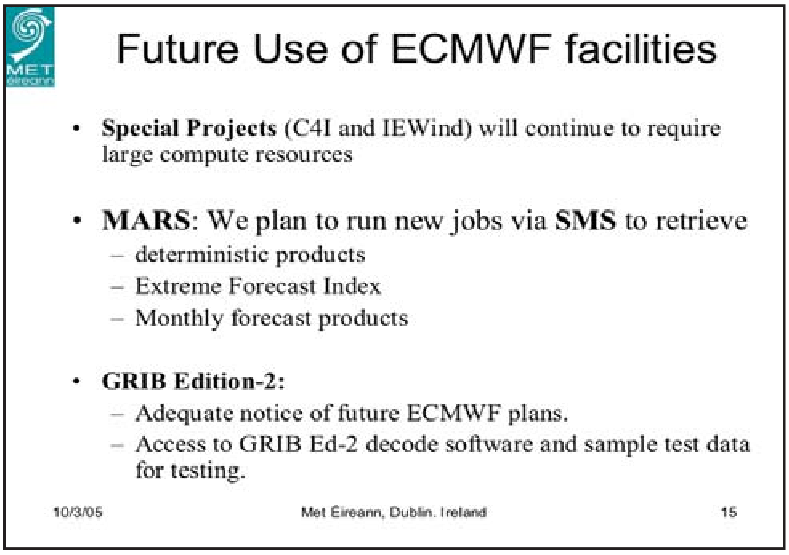
- **ISO 9001:2000** →Accreditation for Aviation Services
- **Update FTP site** → provide access to routine weather forecast and climate data for the new Department of Meteorology and Climate at University College, Dublin
- **WAFS chart production** → implement facilities (using GIS-Meteo from MapMakers) to replace T4-FAX products by end of June 2005.
- **Forecast office efficiency**
  - Implement plans to continue improvements to forecast office …
    - Development of a Point Forecast Database
    - Complete implementation phase of automatic faxing facilities to disseminate scheduled weather forecasts directly to customers
    - Procurement of a forecaster workstation and Production system

10/3/05 Met Éireann, Dublin. Ireland 13



- Linux Cluster
  - 3 x nodes added to cluster →
    - 1 x master node (2 x Xeon 2.8Ghz /512k 533Mhz)
    - 9 x compute nodes (2 x Xeon 3.2Ghz /1Mb 533Mhz)
    - Runs backup version of NWP suite at same resolution as IBM system
  - continue preparations for the procurement of a replacement of main NWP platform (IBM RS-6000 SP) in 2006
- Develop a strategy for the operational introduction of BUFR encoding and decoding of observation data

10/3/05 Met Éireann, Dublin. Ireland 14

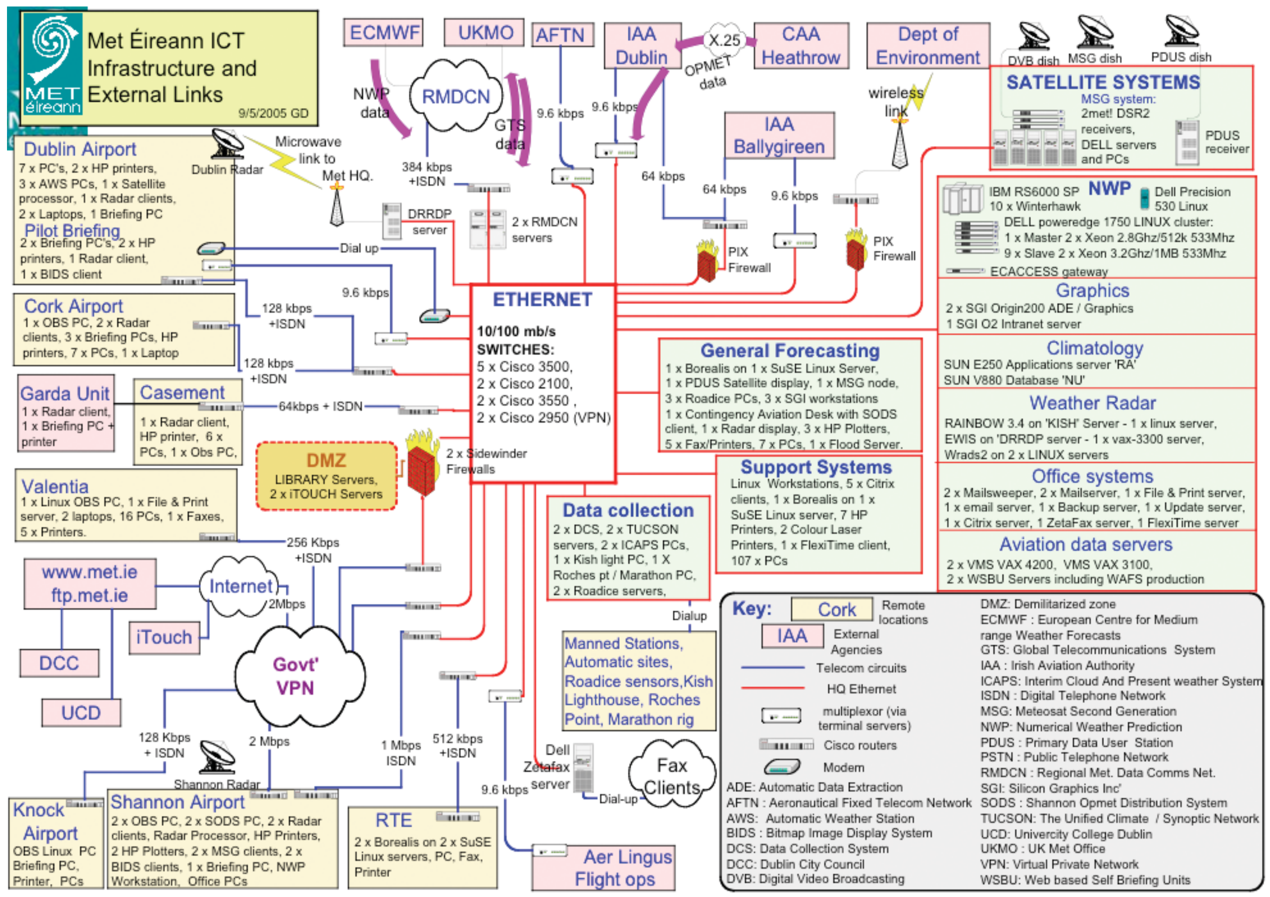


Future Use of ECMWF facilities

- **Special Projects** (C4I and IEWind) will continue to require large compute resources
- **MARS**: We plan to run new jobs via **SMS** to retrieve
  - deterministic products
  - Extreme Forecast Index
  - Monthly forecast products
- **GRIB Edition-2:**
  - Adequate notice of future ECMWF plans.
  - Access to GRIB Ed-2 decode software and sample test data for testing.

10/3/05 Met Éireann, Dublin. Ireland 15

IRELAND



Met Éireann ICT Infrastructure and External Links

9/5/2005 GD

Dublin Airport
7 x PC's, 2 x HP printers, 3 x AWS PCs, 1 x Satellite processor, 1 x Radar clients, 2 x Laptops, 1 Briefing PC

Pilot Briefing
2 x Briefing PC's, 2 x HP printers, 1 Radar client, 1 x BIDS client

Cork Airport
1 x OBS PC, 2 x Radar clients, 3 x Briefing PCs, HP printers, 7 x PCs, 1 x Laptop

Garda Unit
1 x Radar client, 1 x Briefing PC + printer

Casement
1 x Radar client, HP printer, 6 x PCs, 1 x Obs PC,

Valentia
1 x Linux OBS PC, 1 x File & Print server, 2 laptops, 16 PCs, 1 x Faxes, 5 x Printers.

DMZ
LIBRARY Servers, 2 x iTOUCH Servers

ETHERNET
10/100 mb/s
SWITCHES:
5 x Cisco 3500,
2 x Cisco 2100,
2 x Cisco 3550 ,
2 x Cisco 2950 (VPN)

SATELLITE SYSTEMS
MSG system: 2met! DSR2 receivers, DELL servers and PCs
PDUS receiver

NWP
IBM RS6000 SP 10 x Winterhawk
Dell Precision 530 Linux
DELL poweredge 1750 LINUX cluster:
1 x Master 2 x Xeon 2.8Ghz/512k 533Mhz
9 x Slave 2 x Xeon 3.2Ghz/1MB 533Mhz
ECACCESS gateway

Graphics
2 x SGI Origin200 ADE / Graphics
1 SGI O2 Intranet server

Climatology
SUN E250 Applications server 'RA'
SUN V880 Database 'NU'

Weather Radar
RAINBOW 3.4 on 'KISH' Server - 1 x linux server,
EWIS on 'DRRDP server - 1 x vax-3300 server,
Wrads2 on 2 x LINUX servers

General Forecasting
1 x Borealis on 1 x SuSE Linux Server, 1 x PDUS Satellite display, 1 x MSG node, 3 x Roadice PCs, 3 x SGI workstations 1 x Contingency Aviation Desk with SODS client, 1 x Radar display, 3 x HP Plotters, 5 x Fax/Printers, 7 x PCs, 1 x Flood Server.

Office systems
2 x Mailsweeper, 2 x Mailserver, 1 x File & Print server,
1 x email server, 1 x Backup server, 1 x Update server,
1 x Citrix server, 1 ZetaFax server, 1 FlexiTime server

Support Systems
Linux Workstations, 5 x Citrix clients, 1 x Borealis on 1 x SuSE Linux server, 7 HP Printers, 2 Colour Laser Printers, 1 x FlexiTime client, 107 x PCs

Aviation data servers
2 x VMS VAX 4200, VMS VAX 3100,
2 x WSBU Servers including WAFS production

Data collection
2 x DCS, 2 x TUCSON servers, 2 x ICAPS PCs, 1 x Kish light PC, 1 X Roches pt / Marathon PC, 2 x Roadice servers,

Manned Stations, Automatic sites, Roadice sensors,Kish Lighthouse, Roches Point, Marathon rig

Shannon Airport
2 x OBS PC, 2 x SODS PC, 2 x Radar clients, Radar Processor, HP Printers, 2 HP Plotters, 2 x MSG clients, 2 x BIDS clients, 1 x Briefing PC, NWP Workstation, Office PCs

Knock Airport
OBS Linux PC Briefing PC, Printer, PCs

RTE
2 x Borealis on 2 x SuSE Linux servers, PC, Fax, Printer

Key:
Cork Remote locations
IAA External Agencies
Telecom circuits
HQ Ethernet
multiplexor (via terminal servers)
Cisco routers
Modem

ADE: Automatic Data Extraction
AFTN : Aeronautical Fixed Telecom Network
AWS: Automatic Weather Station
BIDS : Bitmap Image Display System
DCS: Data Collection System
DCC: Dublin City Council
DVB: Digital Video Broadcasting
DMZ: Demilitarized zone
ECMWF : European Centre for Medium range Weather Forecasts
GTS: Global Telecommunications System
IAA : Irish Aviation Authority
ICAPS: Interim Cloud And Present weather System
ISDN : Digital Telephone Network
MSG: Meteosat Second Generation
NWP: Numerical Weather Prediction
PDUS : Primary Data User Station
PSTN : Public Telephone Network
RMDCN : Regional Met. Data Comms Net.
SGI: Silicon Graphics Inc'
SODS : Shannon Opmet Distribution System
TUCSON: The Unified Climate / Synoptic Network
UCD: Univercity College Dublin
UKMO : UK Met Office
VPN: Virtual Private Network
WSBU: Web based Self Briefing Units

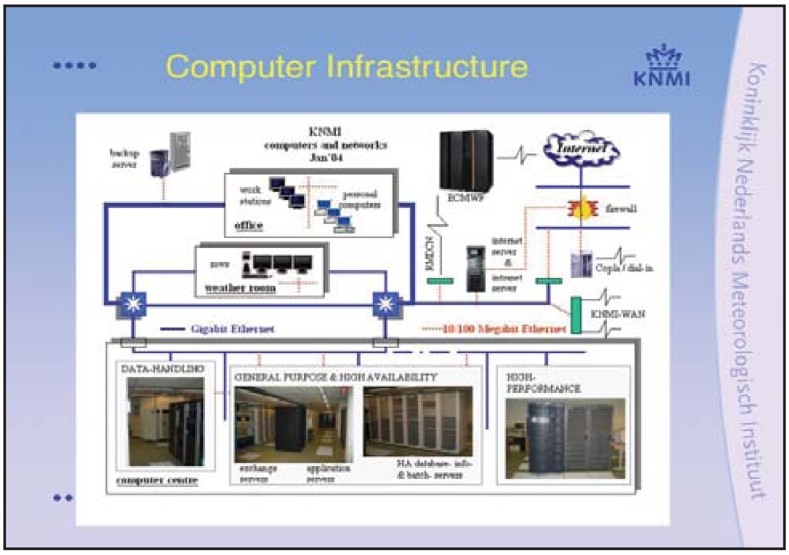
10/3/05 Met Éireann, Dublin. Ireland 16

L. Gougeon explained that certificates with longer validity could be set up for operational tasks. A request should be sent to the ECMWF Security Officer (M. Dell'Acqua), explaining the purpose for which the extended validity certificate is required.

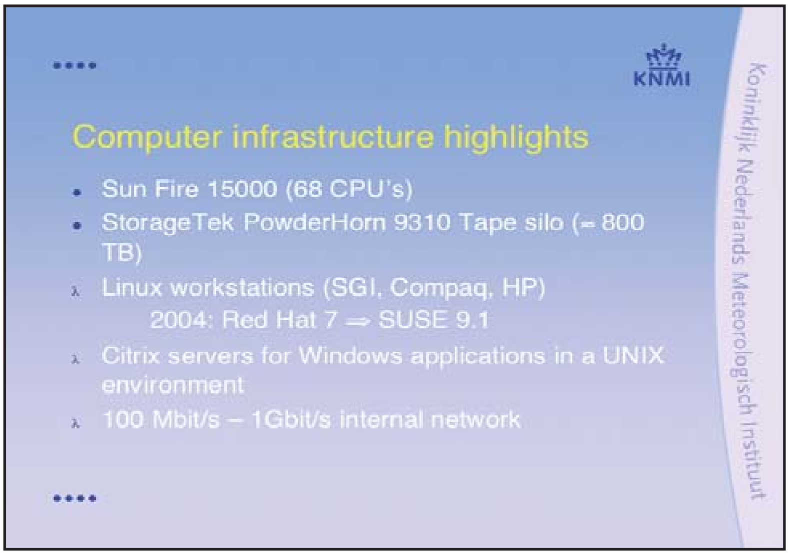
In regard to the reported problems accessing ECPDS monitoring tools, L. Gougeon asked whether the user might have been trying to gain access during the dissemination, when the network bandwidth was fully utilised. P. Halton replied that the user was at Shannon airport, so network problems might well have been to blame.

R. Rudsar asked why they used jobs submitted by SMS for data retrieval, rather than the dissemination. P.Halton replied that they had encouraged ECaccess use to keep additional, experimental products separate from the routine dissemination. U. Modigliani noted that new products were not immediately available in the dissemination, so users could obtain them initially via SMS and ECaccess, until they became part of the dissemination.

Gert-Jan Marseille – KNMI, The Netherlands



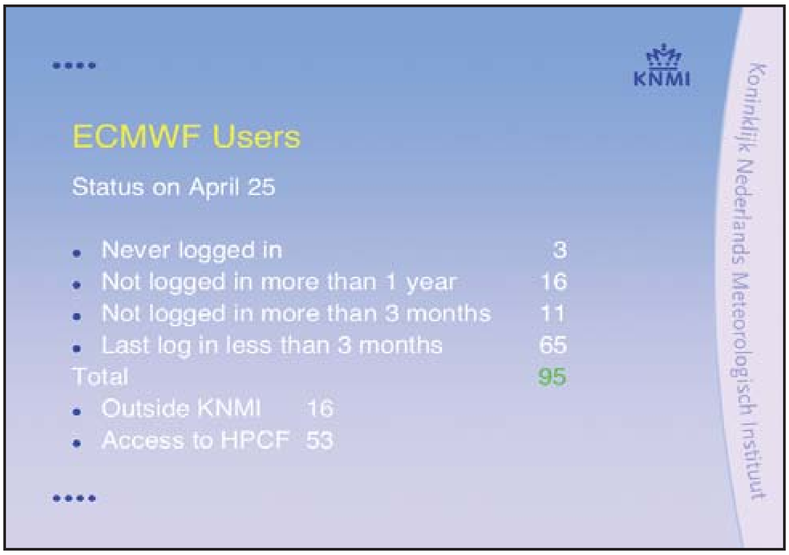
Computer Infrastructure



Computer infrastructure highlights

- Sun Fire 15000 (68 CPU's)
- StorageTek PowderHorn 9310 Tape silo (≈ 800 TB)
- Linux workstations (SGI, Compaq, HP)
  2004: Red Hat 7 ⇒ SUSE 9.1
- Citrix servers for Windows applications in a UNIX environment
- 100 Mbit/s – 1Gbit/s internal network

KNMI
Koninklijk Nederlands Meteorologisch Instituut

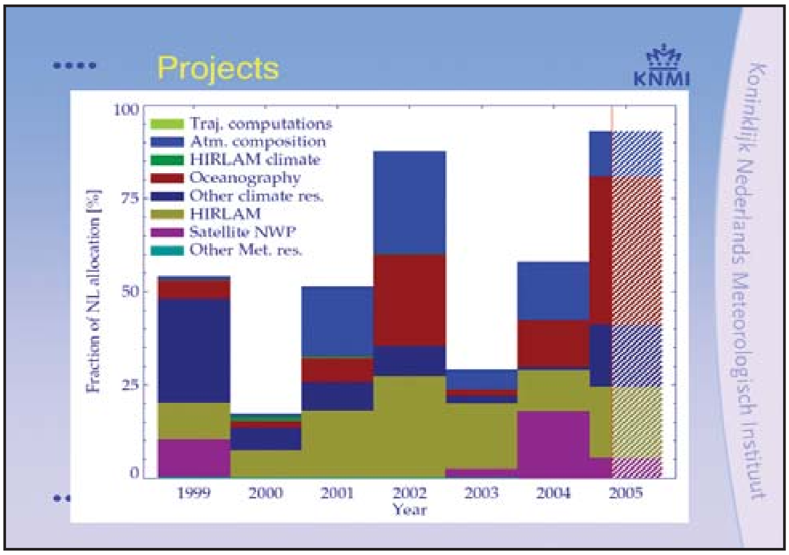


ECMWF Users

Status on April 25

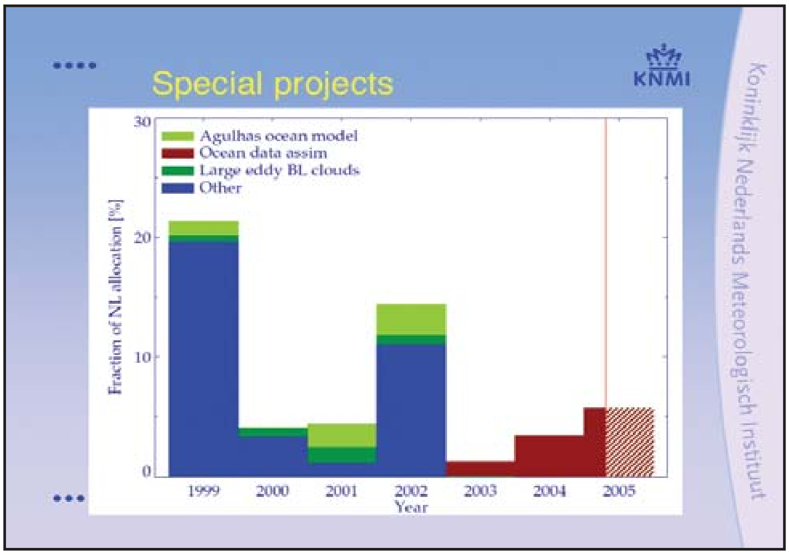
• Never logged in	3
• Not logged in more than 1 year	16
• Not logged in more than 3 months	11
• Last log in less than 3 months	65
Total	95
• Outside KNMI	16
• Access to HPCF	53

KNMI
Koninklijk Nederlands Meteorologisch Instituut



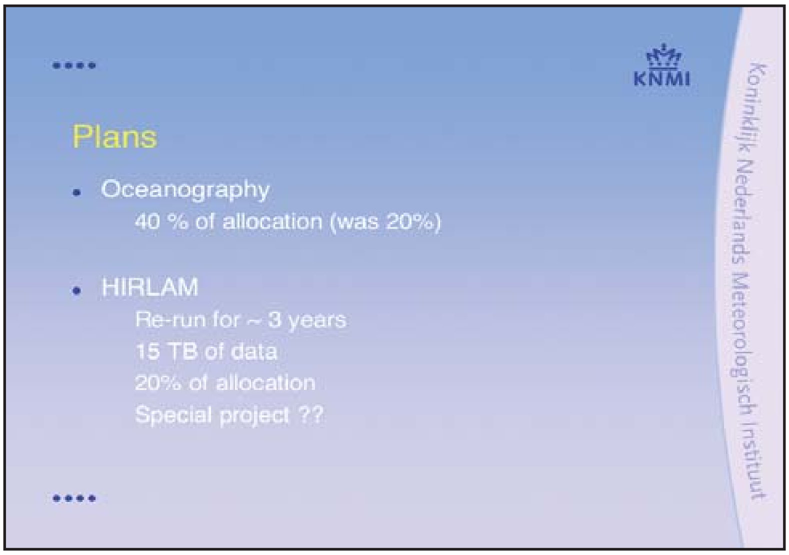
Projects

Koninklijk Nederlands Meteorologisch Instituut



Special projects

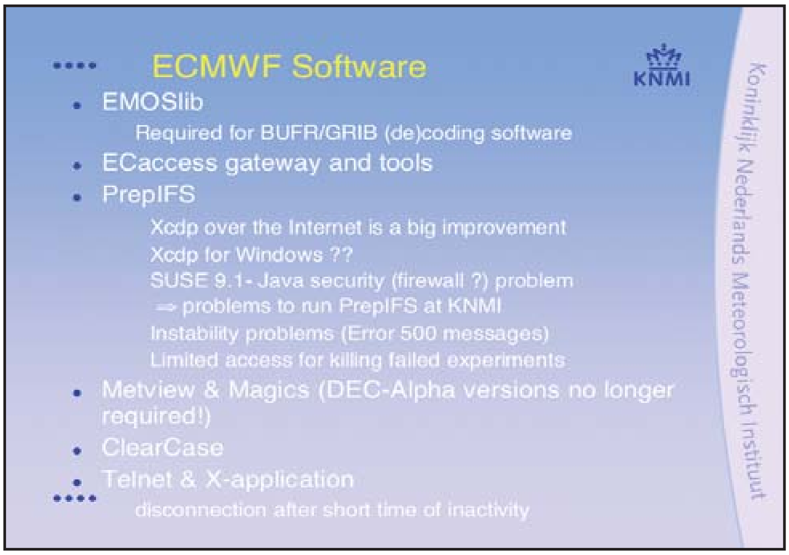
Koninklijk Nederlands Meteorologisch Instituut



Plans

- Oceanography
  40 % of allocation (was 20%)
- HIRLAM
  Re-run for ~ 3 years
  15 TB of data
  20% of allocation
  Special project ??

Koninklijk Nederlands Meteorologisch Instituut

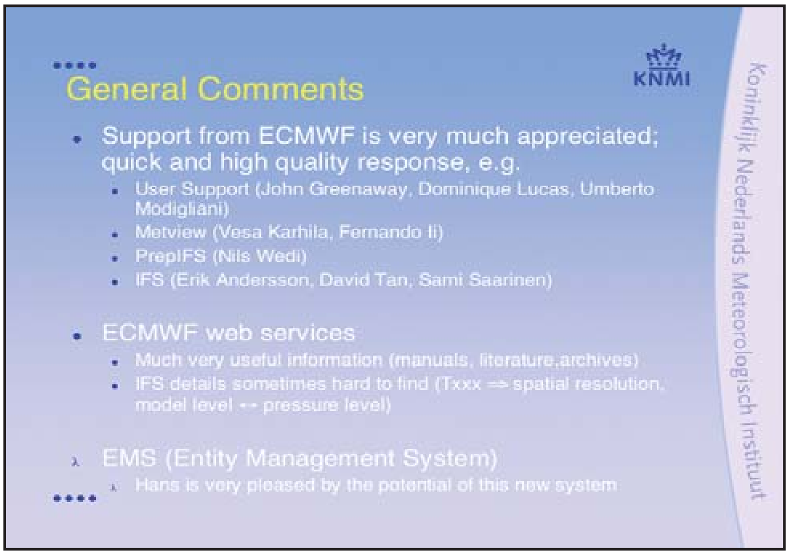


ECMWF Software

KNMI

Koninklijk Nederlands Meteorologisch Instituut

- EMOSlib
  Required for BUFR/GRIB (de)coding software
- ECaccess gateway and tools
- PrepIFS
  Xcdp over the Internet is a big improvement
  Xcdp for Windows ??
  SUSE 9.1- Java security (firewall ?) problem
  ⇒ problems to run PrepIFS at KNMI
  Instability problems (Error 500 messages)
  Limited access for killing failed experiments
- Metview & Magics (DEC-Alpha versions no longer required!)
- ClearCase
- Telnet & X-application
  disconnection after short time of inactivity



General Comments

KNMI

Koninklijk Nederlands Meteorologisch Instituut

- Support from ECMWF is very much appreciated; quick and high quality response, e.g.
  - User Support (John Greenaway, Dominique Lucas, Umberto Modigliani)
  - Metview (Vesa Karhila, Fernando Ii)
  - PrepIFS (Nils Wedi)
  - IFS (Erik Andersson, David Tan, Sami Saarinen)
- ECMWF web services
  - Much very useful information (manuals, literature,archives)
  - IFS details sometimes hard to find (Txxx ⇒ spatial resolution, model level ↔ pressure level)
- EMS (Entity Management System)
  - Hans is very pleased by the potential of this new system

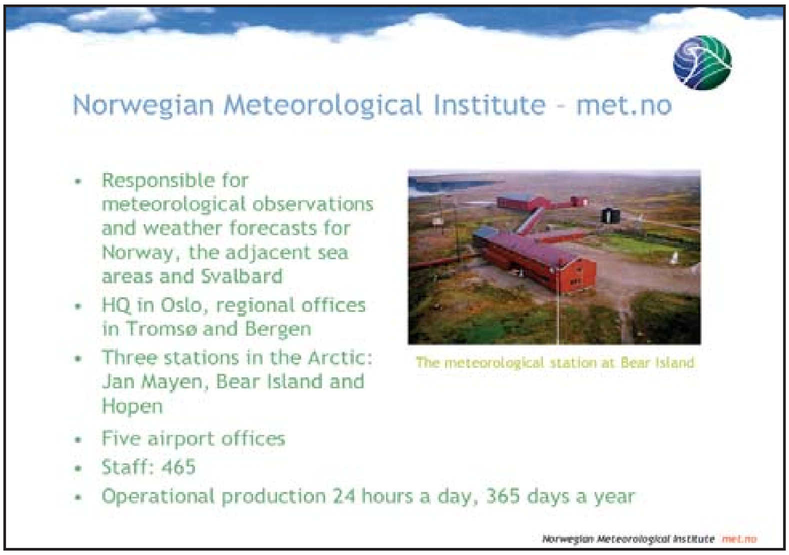
In reply to comments made during the presentation J. Greenaway noted that the instability problems would need in depth investigation. A new version of PrepIFs is now available and, it is hoped, will resolve the problems experienced at KNMI.

L. Gougeon reported that the problem of ssh sessions being disconnected after a short period of inactivity had been resolved for some users by increasing the timeout period. The disconnection of x-sessions is linked to Firewall inactivity timeouts: users connect to their remote ECaccess gateway, which is connected to the ECMWF ECaccess server by a non-standard port and Firewalls tend to disconnect after very short periods of inactivity. These periods can be increased to avoid unnecessary timeouts.

R. Fisker noted that Xcdp was run with Windows on ECMWF laptops using Public Domain software CYGWIN, which provides an x-server under Windows. It is not planned to port Xcdp to Windows.

NORWAY

Rebecca Rudsar – Norwegian Meteorological Institute, met.no

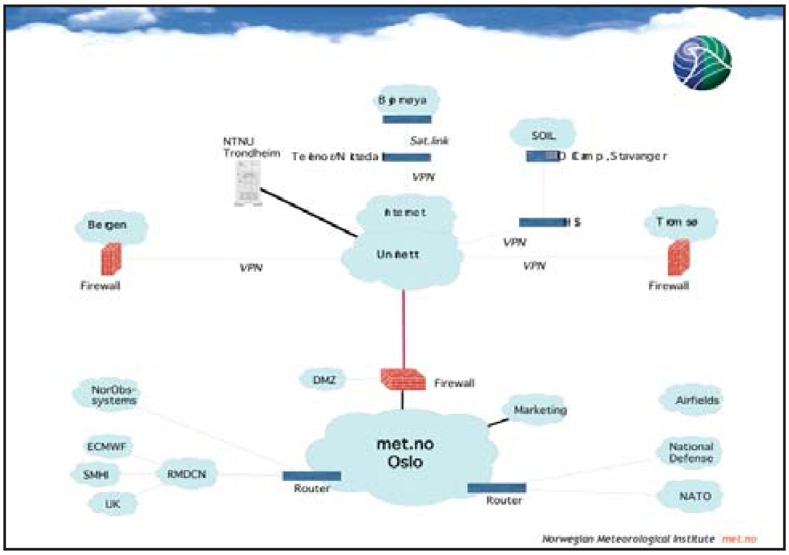


Norwegian Meteorological Institute - met.no

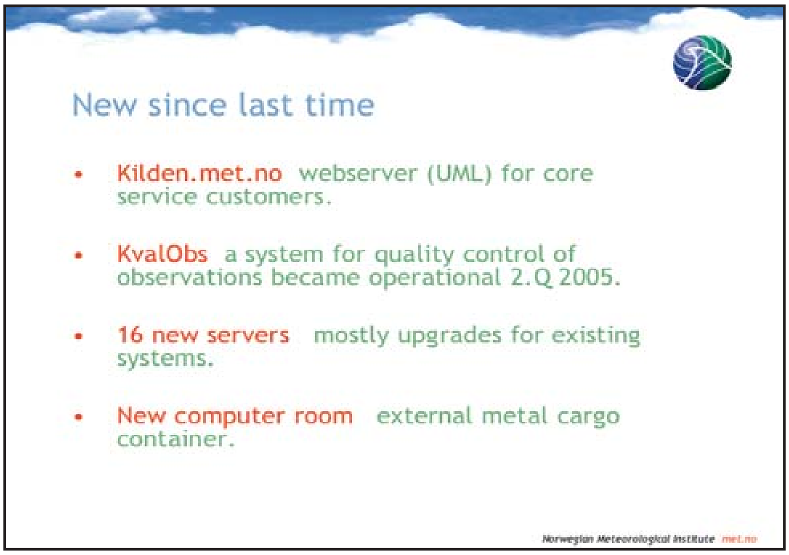
- Responsible for meteorological observations and weather forecasts for Norway, the adjacent sea areas and Svalbard
- HQ in Oslo, regional offices in Tromsø and Bergen
- Three stations in the Arctic: Jan Mayen, Bear Island and Hopen
- Five airport offices
- Staff: 465
- Operational production 24 hours a day, 365 days a year

The meteorological station at Bear Island

Norwegian Meteorological Institute met.no



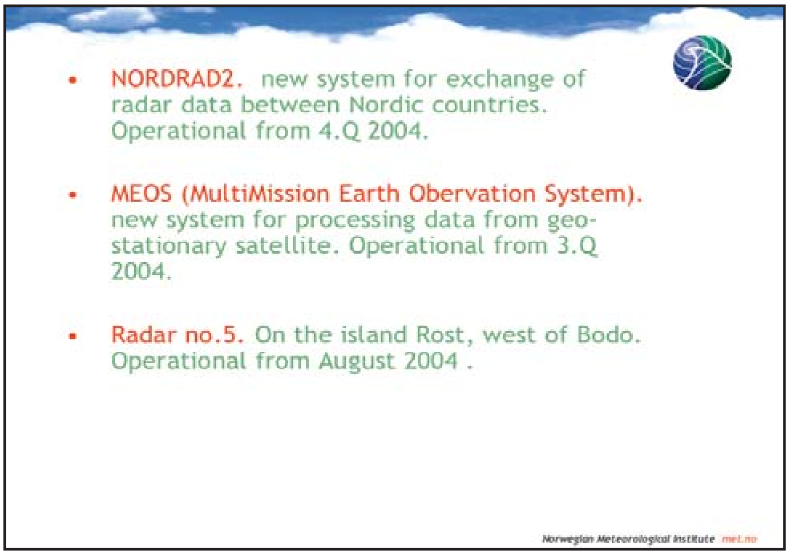
Norwegian Meteorological Institute met.no



New since last time

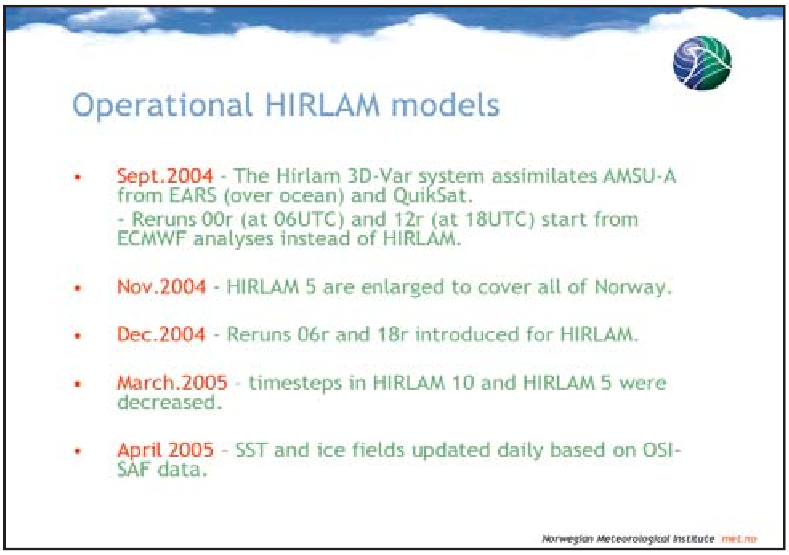
- Kilden.met.no webserver (UML) for core service customers.
- KvalObs a system for quality control of observations became operational 2.Q 2005.
- 16 new servers mostly upgrades for existing systems.
- New computer room external metal cargo container.

Norwegian Meteorological Institute met.no



- NORDRAD2. new system for exchange of radar data between Nordic countries. Operational from 4.Q 2004.
- MEOS (MultiMission Earth Obervation System). new system for processing data from geo-stationary satellite. Operational from 3.Q 2004.
- Radar no.5. On the island Rost, west of Bodo. Operational from August 2004 .

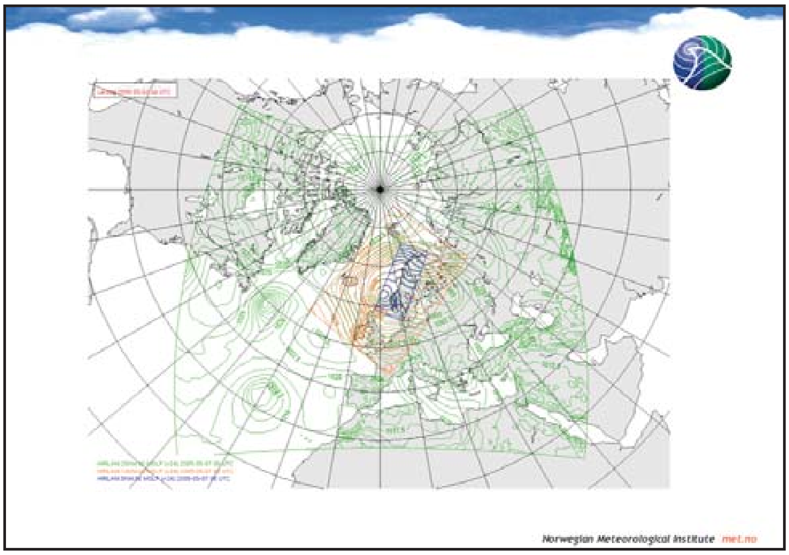
Norwegian Meteorological Institute met.no



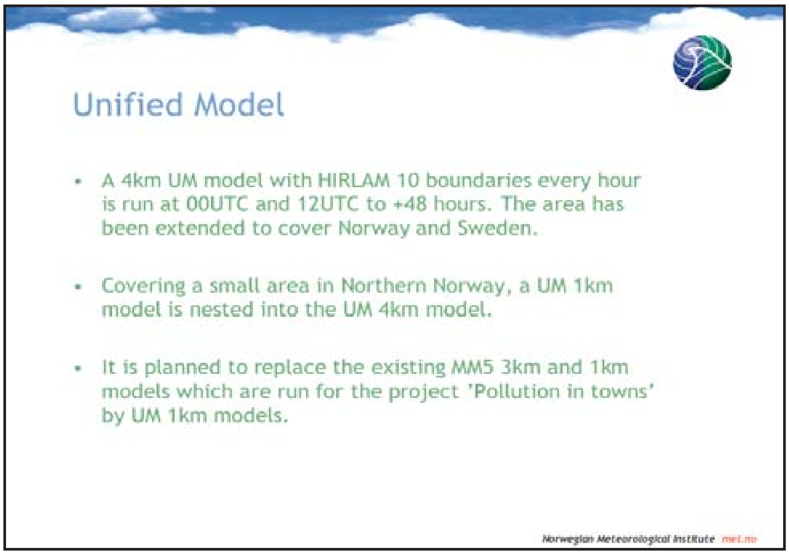
Operational HIRLAM models

- Sept.2004 - The Hirlam 3D-Var system assimilates AMSU-A from EARS (over ocean) and QuikSat.
  - Reruns 00r (at 06UTC) and 12r (at 18UTC) start from ECMWF analyses instead of HIRLAM.
- Nov.2004 - HIRLAM 5 are enlarged to cover all of Norway.
- Dec.2004 - Reruns 06r and 18r introduced for HIRLAM.
- March.2005 - timesteps in HIRLAM 10 and HIRLAM 5 were decreased.
- April 2005 - SST and ice fields updated daily based on OSI-SAF data.

Norwegian Meteorological Institute met.no



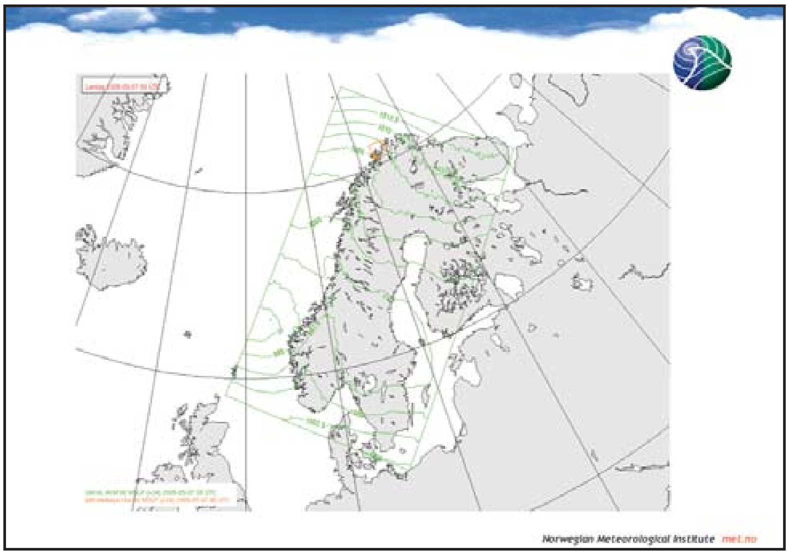
Norwegian Meteorological Institute met.no



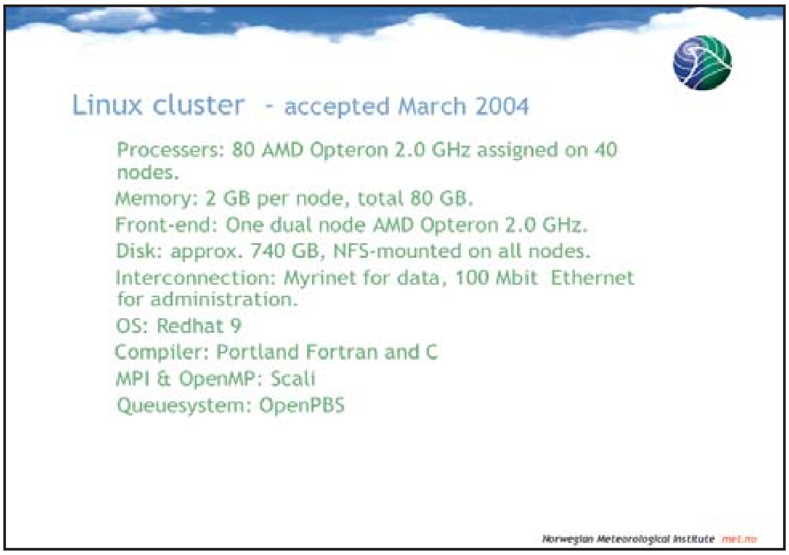
Unified Model

- A 4km UM model with HIRLAM 10 boundaries every hour is run at 00UTC and 12UTC to +48 hours. The area has been extended to cover Norway and Sweden.
- Covering a small area in Northern Norway, a UM 1km model is nested into the UM 4km model.
- It is planned to replace the existing MM5 3km and 1km models which are run for the project 'Pollution in towns' by UM 1km models.

Norwegian Meteorological Institute met.no



Norwegian Meteorological Institute met.no



Linux cluster - accepted March 2004

Processers: 80 AMD Opteron 2.0 GHz assigned on 40 nodes.

Memory: 2 GB per node, total 80 GB.

Front-end: One dual node AMD Opteron 2.0 GHz.

Disk: approx. 740 GB, NFS-mounted on all nodes.

Interconnection: Myrinet for data, 100 Mbit Ethernet for administration.

OS: Redhat 9

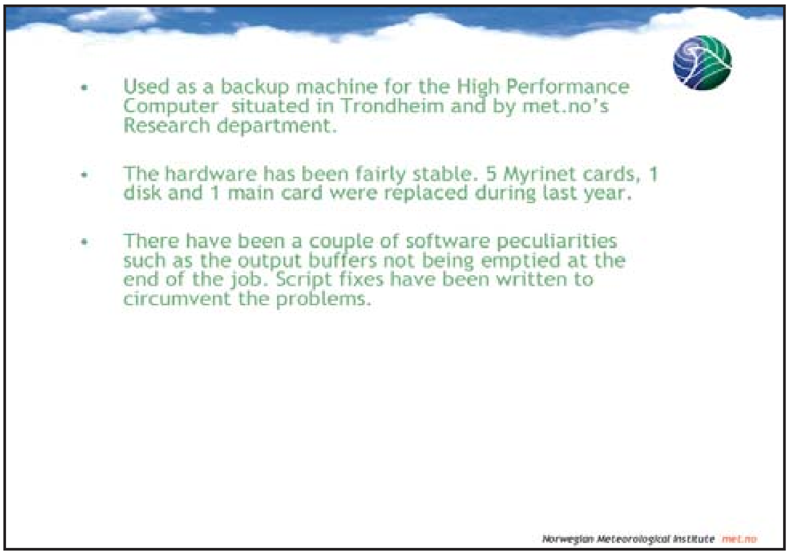
Compiler: Portland Fortran and C

MPI & OpenMP: Scali

Queuesystem: OpenPBS

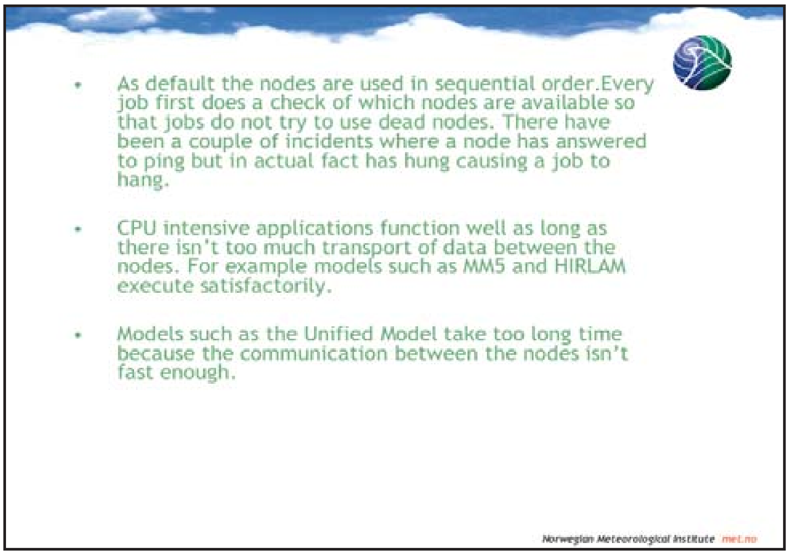
Norwegian Meteorological Institute met.no

NORWAY



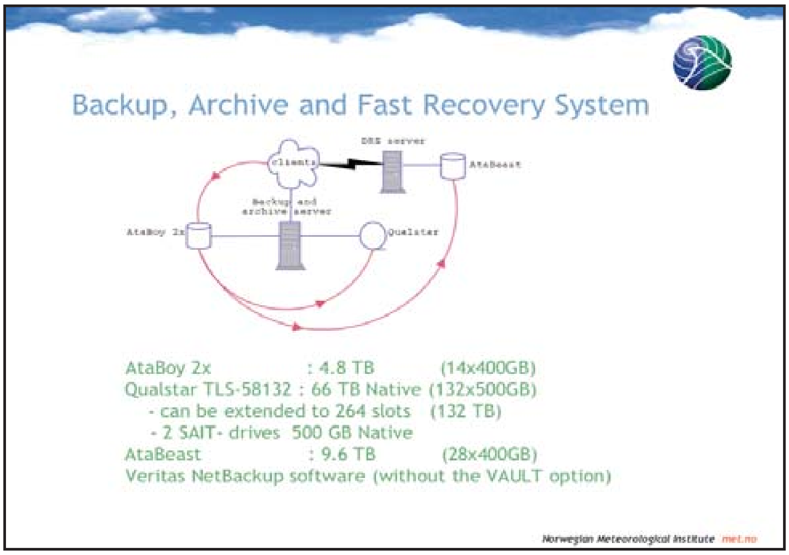
- Used as a backup machine for the High Performance Computer situated in Trondheim and by met.no's Research department.
- The hardware has been fairly stable. 5 Myrinet cards, 1 disk and 1 main card were replaced during last year.
- There have been a couple of software peculiarities such as the output buffers not being emptied at the end of the job. Script fixes have been written to circumvent the problems.

Norwegian Meteorological Institute met.no



- As default the nodes are used in sequential order.Every job first does a check of which nodes are available so that jobs do not try to use dead nodes. There have been a couple of incidents where a node has answered to ping but in actual fact has hung causing a job to hang.
- CPU intensive applications function well as long as there isn't too much transport of data between the nodes. For example models such as MM5 and HIRLAM execute satisfactorily.
- Models such as the Unified Model take too long time because the communication between the nodes isn't fast enough.

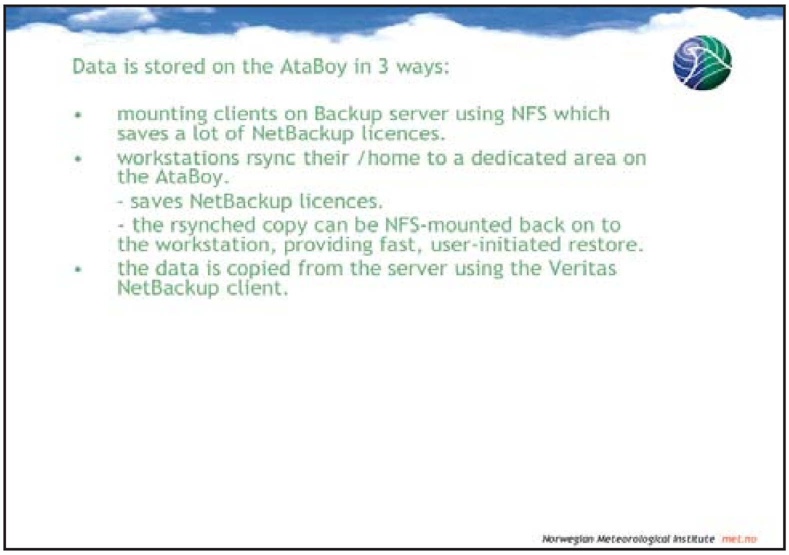
Norwegian Meteorological Institute met.no



Backup, Archive and Fast Recovery System

AtaBoy 2x : 4.8 TB (14x400GB)
Qualstar TLS-58132 : 66 TB Native (132x500GB)
- can be extended to 264 slots (132 TB)
- 2 SAIT- drives 500 GB Native
AtaBeast : 9.6 TB (28x400GB)
Veritas NetBackup software (without the VAULT option)

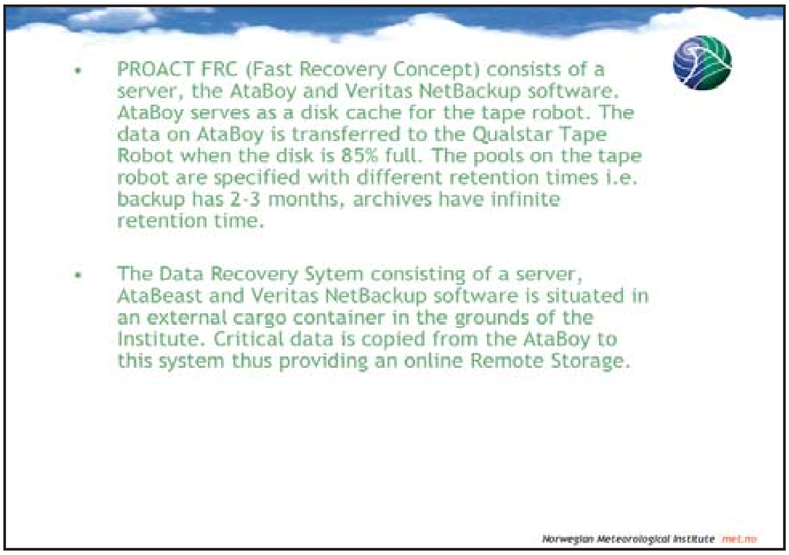
Norwegian Meteorological Institute met.no



Data is stored on the AtaBoy in 3 ways:

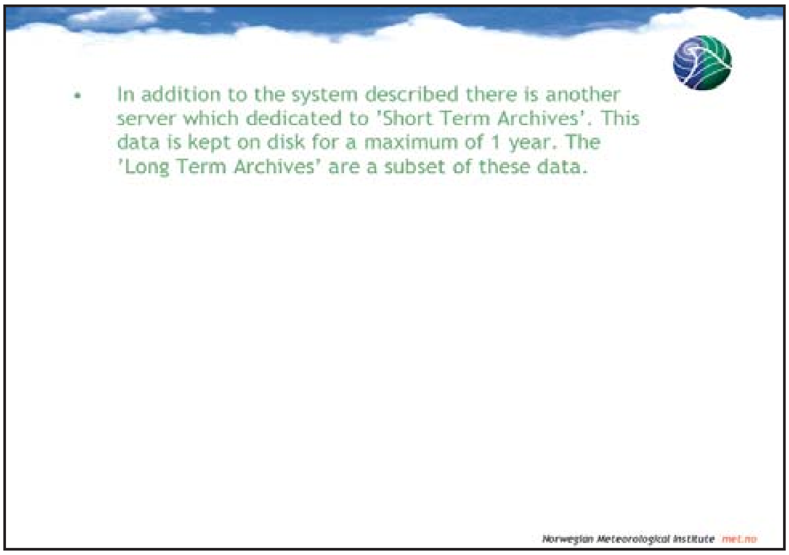
- mounting clients on Backup server using NFS which saves a lot of NetBackup licences.
- workstations rsync their /home to a dedicated area on the AtaBoy.
  - saves NetBackup licences.
  - the rsynched copy can be NFS-mounted back on to the workstation, providing fast, user-initiated restore.
- the data is copied from the server using the Veritas NetBackup client.

Norwegian Meteorological Institute met.no



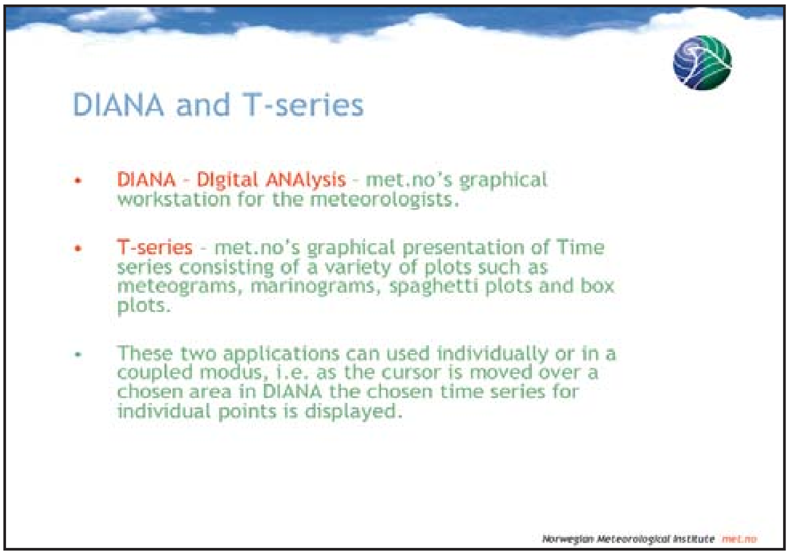
- PROACT FRC (Fast Recovery Concept) consists of a server, the AtaBoy and Veritas NetBackup software. AtaBoy serves as a disk cache for the tape robot. The data on AtaBoy is transferred to the Qualstar Tape Robot when the disk is 85% full. The pools on the tape robot are specified with different retention times i.e. backup has 2-3 months, archives have infinite retention time.
- The Data Recovery Sytem consisting of a server, AtaBeast and Veritas NetBackup software is situated in an external cargo container in the grounds of the Institute. Critical data is copied from the AtaBoy to this system thus providing an online Remote Storage.

Norwegian Meteorological Institute met.no



- In addition to the system described there is another server which dedicated to 'Short Term Archives'. This data is kept on disk for a maximum of 1 year. The 'Long Term Archives' are a subset of these data.

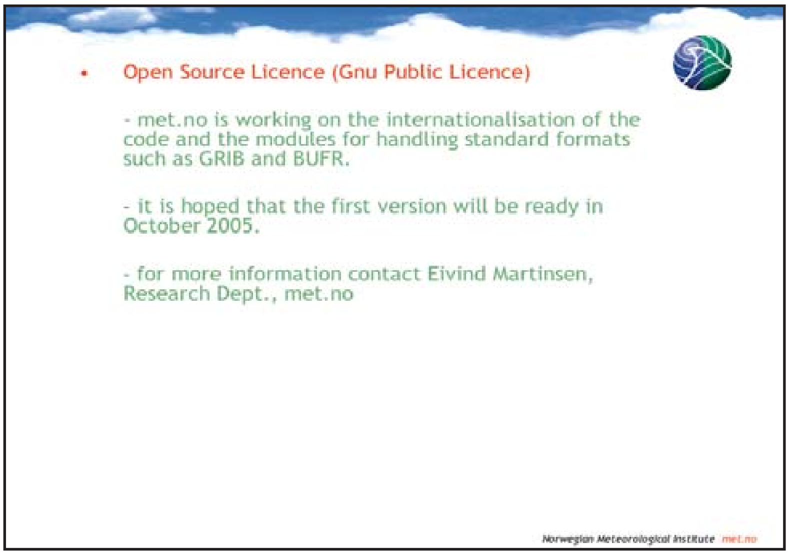
Norwegian Meteorological Institute met.no



DIANA and T-series

- DIANA - DIgital ANAlysis - met.no's graphical workstation for the meteorologists.
- T-series - met.no's graphical presentation of Time series consisting of a variety of plots such as meteograms, marinograms, spaghetti plots and box plots.
- These two applications can used individually or in a coupled modus, i.e. as the cursor is moved over a chosen area in DIANA the chosen time series for individual points is displayed.

Norwegian Meteorological Institute met.no



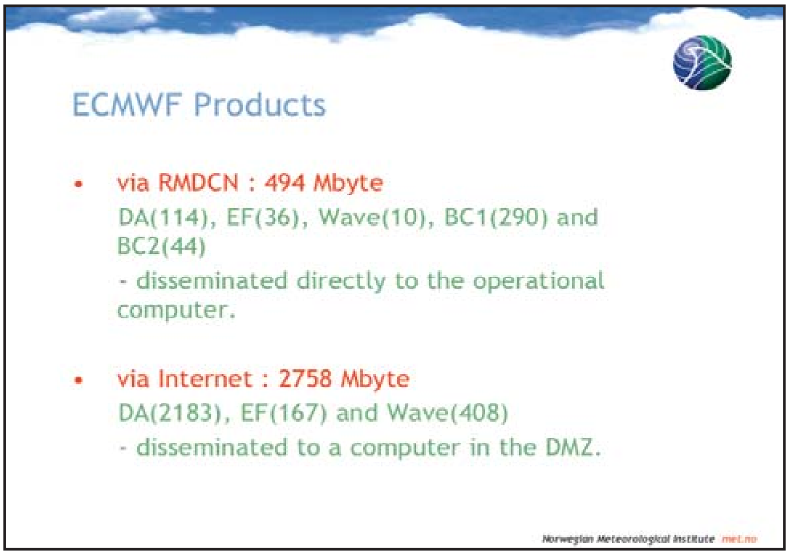
- Open Source Licence (Gnu Public Licence)

  - met.no is working on the internationalisation of the code and the modules for handling standard formats such as GRIB and BUFR.

  - it is hoped that the first version will be ready in October 2005.

  - for more information contact Eivind Martinsen, Research Dept., met.no

Norwegian Meteorological Institute met.no



ECMWF Products

- via RMDCN : 494 Mbyte

  DA(114), EF(36), Wave(10), BC1(290) and BC2(44)

  - disseminated directly to the operational computer.

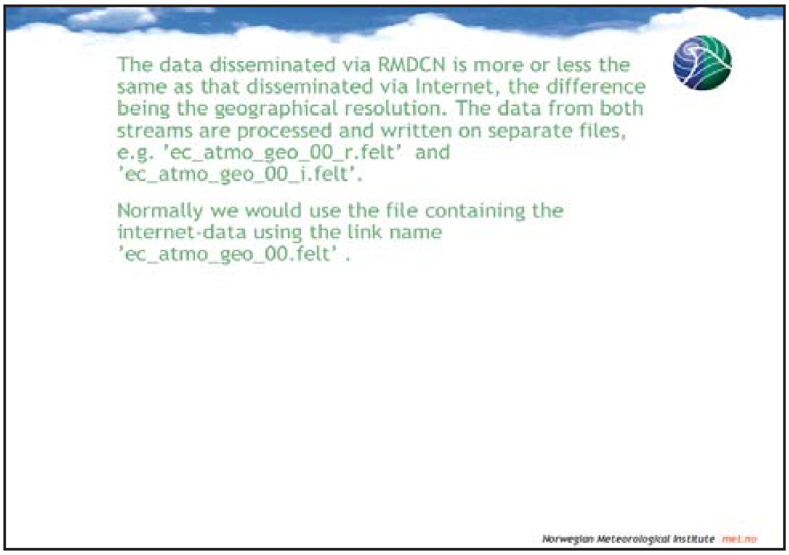
- via Internet : 2758 Mbyte

  DA(2183), EF(167) and Wave(408)

  - disseminated to a computer in the DMZ.

Norwegian Meteorological Institute met.no

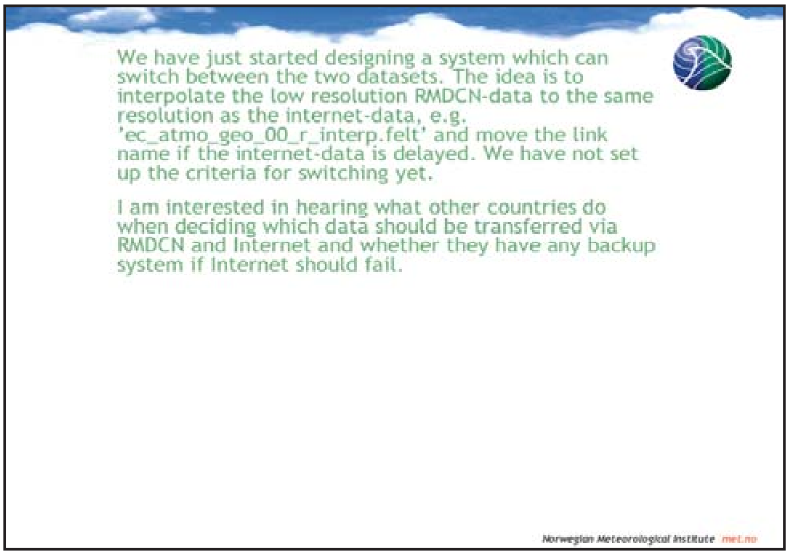
NORWAY NORWAY



The data disseminated via RMDCN is more or less the same as that disseminated via Internet, the difference being the geographical resolution. The data from both streams are processed and written on separate files, e.g. 'ec_atmo_geo_00_r.felt' and 'ec_atmo_geo_00_i.felt'.

Normally we would use the file containing the internet-data using the link name 'ec_atmo_geo_00.felt' .

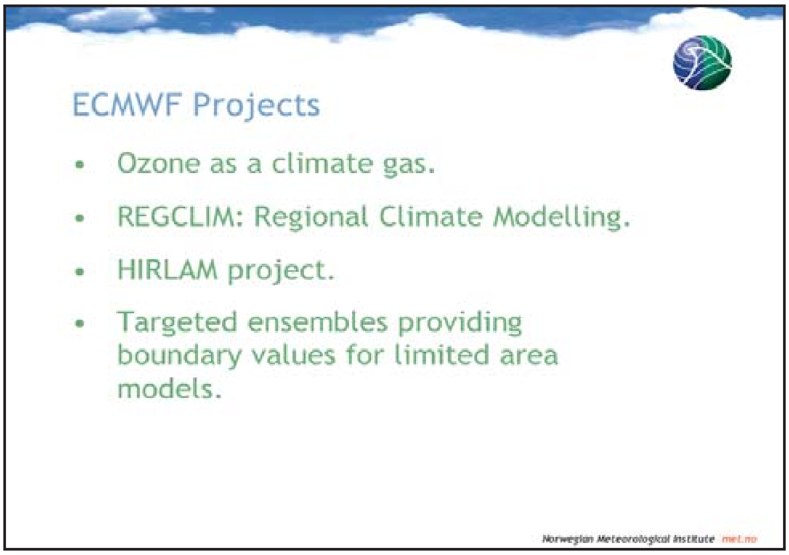
Norwegian Meteorological Institute met.no



We have just started designing a system which can switch between the two datasets. The idea is to interpolate the low resolution RMDCN-data to the same resolution as the internet-data, e.g. 'ec_atmo_geo_00_r_interp.felt' and move the link name if the internet-data is delayed. We have not set up the criteria for switching yet.

I am interested in hearing what other countries do when deciding which data should be transferred via RMDCN and Internet and whether they have any backup system if Internet should fail.

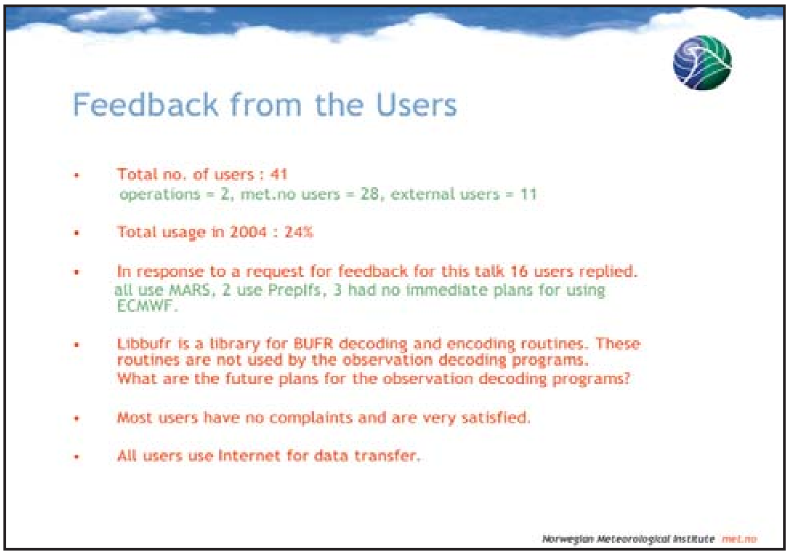
Norwegian Meteorological Institute met.no



ECMWF Projects

- Ozone as a climate gas.
- REGCLIM: Regional Climate Modelling.
- HIRLAM project.
- Targeted ensembles providing boundary values for limited area models.

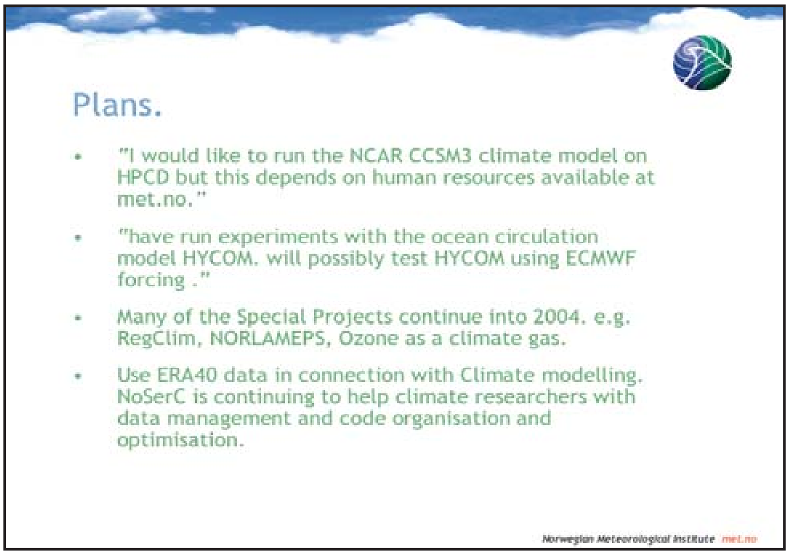
Norwegian Meteorological Institute met.no



Feedback from the Users

- Total no. of users : 41
  operations = 2, met.no users = 28, external users = 11
- Total usage in 2004 : 24%
- In response to a request for feedback for this talk 16 users replied.
  all use MARS, 2 use Preplfs, 3 had no immediate plans for using ECMWF.
- Libbufr is a library for BUFR decoding and encoding routines. These routines are not used by the observation decoding programs.
  What are the future plans for the observation decoding programs?
- Most users have no complaints and are very satisfied.
- All users use Internet for data transfer.

Norwegian Meteorological Institute met.no



Plans.

- "I would like to run the NCAR CCSM3 climate model on HPCD but this depends on human resources available at met.no."
- "have run experiments with the ocean circulation model HYCOM. will possibly test HYCOM using ECMWF forcing ."
- Many of the Special Projects continue into 2004. e.g. RegClim, NORLAMEPS, Ozone as a climate gas.
- Use ERA40 data in connection with Climate modelling. NoSerC is continuing to help climate researchers with data management and code organisation and optimisation.

Norwegian Meteorological Institute met.no

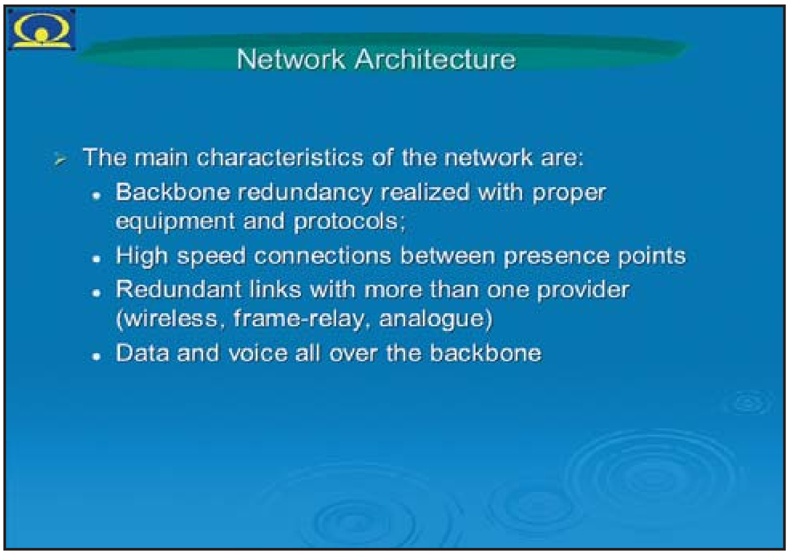
In reply to R. Rudsar's question about plans for observation decoding programs, U. Modigliani stated that he was unaware of any plans to rewrite the preprocessing software. R. Rudsar had detailed discussions with A.Hofstadler after the meeting.

M. Pithon, referring to mention of slow communications within applications on Norway's Linux cluster, asked if the source of the problems — hardware or software — was known. P. Dando, speaking as a former U.K. met service member, replied that the delay was likely to have been caused by model communications: there is much swapping of haloes. Previously, buffered MPI was used; a recent upgrade dispensed with the use of buffers and this seems to cause the delays.

H. Bjornsson asked why Norway ran two high resolution, non-hydrostatic models (MM5 and UM). R. Rudsar replied that the MM5 model had been run for approximately five years, but only in conjunction with the Pollution in Towns project, for very small areas over towns. They are now running the Unified Model (UM), with the agreement of the UK met. service, and are able to run smaller resolutions too. They do not have the resources to maintain both models; the MM5 will be discontinued.

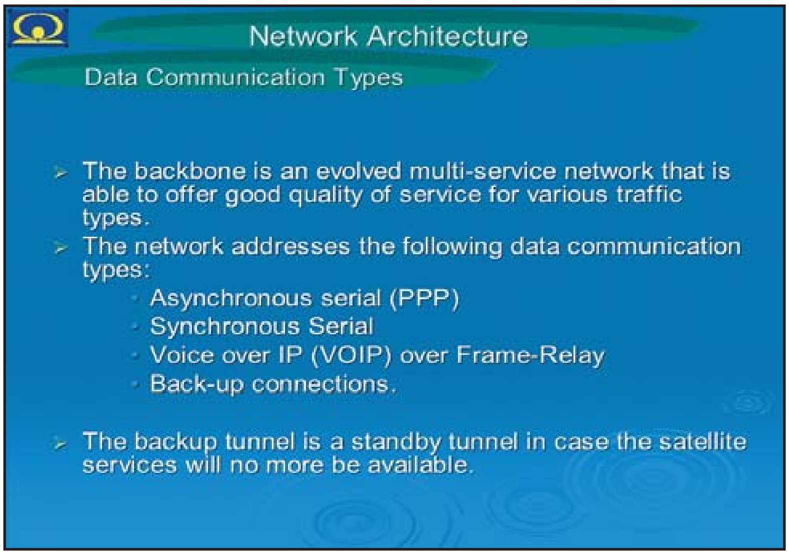
ROMANIA ROMANIA

Catalin Ostroveanu – National Meteorological Administration, Romania



Network Architecture

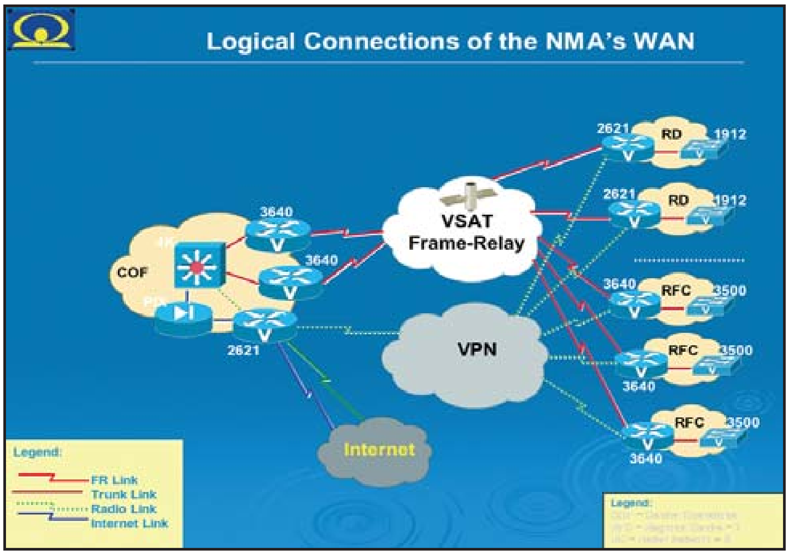
- The main characteristics of the network are:
  - Backbone redundancy realized with proper equipment and protocols;
  - High speed connections between presence points
  - Redundant links with more than one provider (wireless, frame-relay, analogue)
  - Data and voice all over the backbone



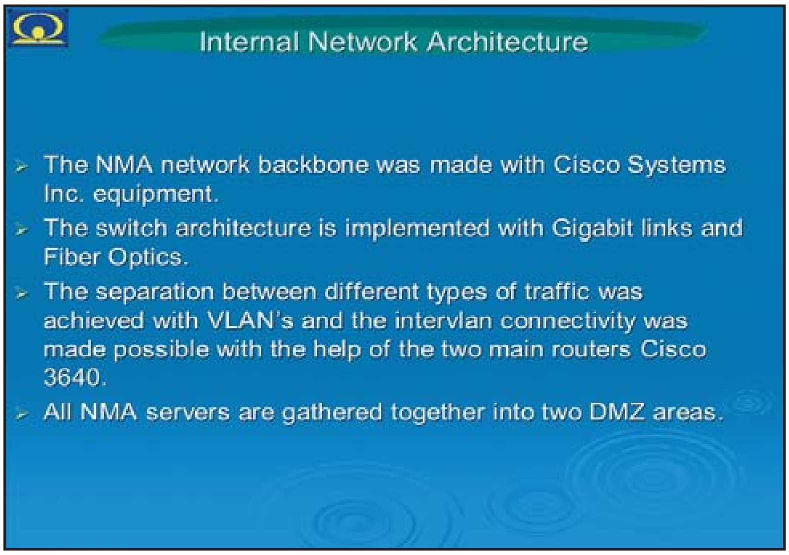
Network Architecture

Data Communication Types

- The backbone is an evolved multi-service network that is able to offer good quality of service for various traffic types.
- The network addresses the following data communication types:
  - Asynchronous serial (PPP)
  - Synchronous Serial
  - Voice over IP (VOIP) over Frame-Relay
  - Back-up connections.
- The backup tunnel is a standby tunnel in case the satellite services will no more be available.

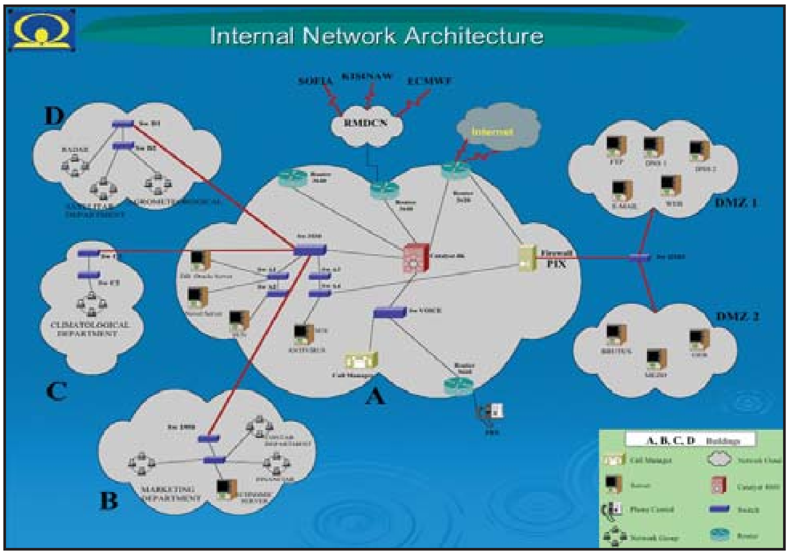


Logical Connections of the NMA's WAN

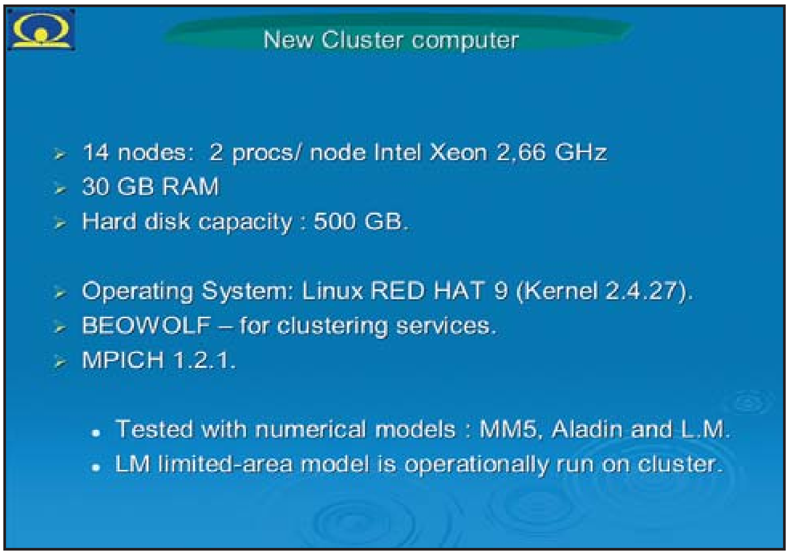


Internal Network Architecture

- The NMA network backbone was made with Cisco Systems Inc. equipment.
- The switch architecture is implemented with Gigabit links and Fiber Optics.
- The separation between different types of traffic was achieved with VLAN's and the intervlan connectivity was made possible with the help of the two main routers Cisco 3640.
- All NMA servers are gathered together into two DMZ areas.



Internal Network Architecture

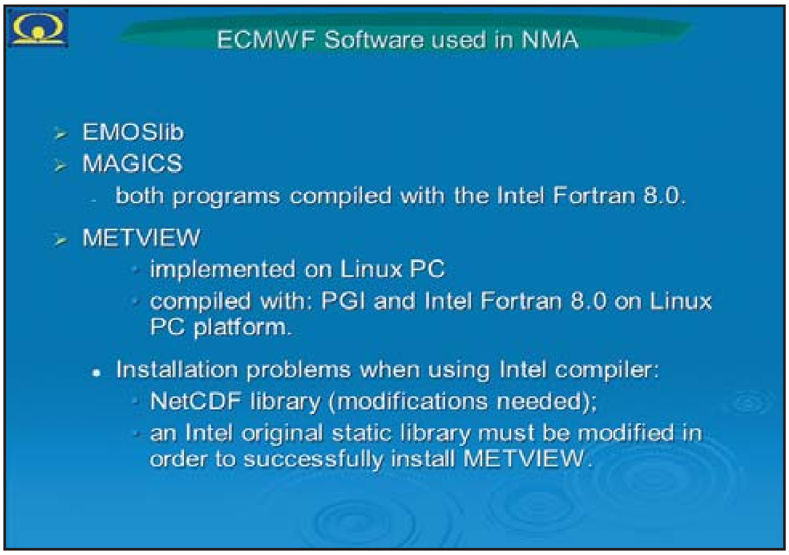


New Cluster computer

- 14 nodes: 2 procs/ node Intel Xeon 2,66 GHz
- 30 GB RAM
- Hard disk capacity : 500 GB.

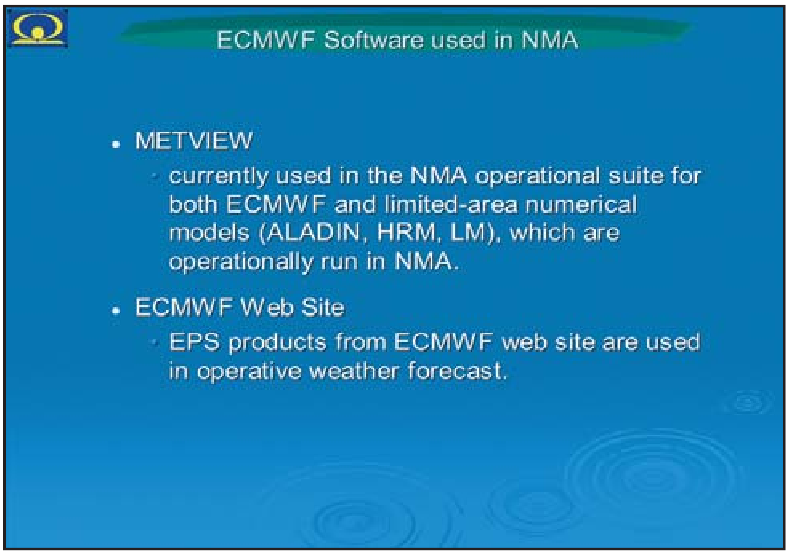
- Operating System: Linux RED HAT 9 (Kernel 2.4.27).
- BEOWOLF – for clustering services.
- MPICH 1.2.1.

  - Tested with numerical models : MM5, Aladin and L.M.
  - LM limited-area model is operationally run on cluster.



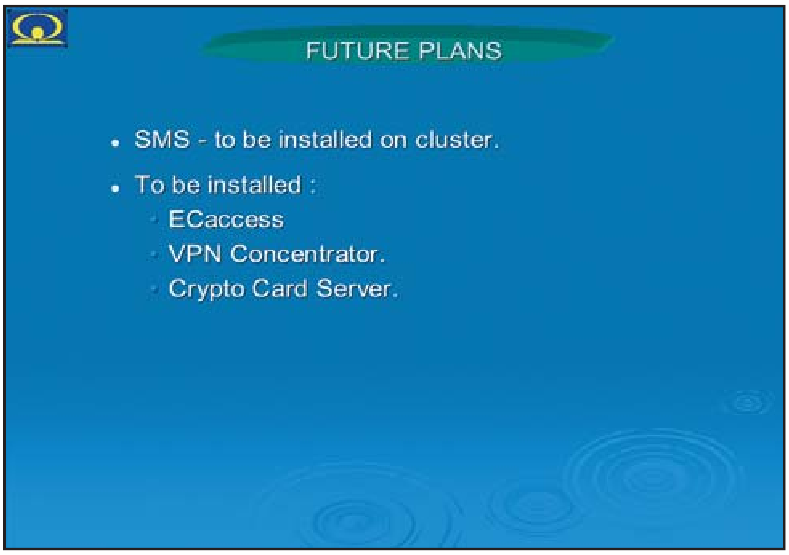
ECMWF Software used in NMA

- EMOSlib
- MAGICS
  - both programs compiled with the Intel Fortran 8.0.
- METVIEW
  - implemented on Linux PC
  - compiled with: PGI and Intel Fortran 8.0 on Linux PC platform.
- Installation problems when using Intel compiler:
  - NetCDF library (modifications needed);
  - an Intel original static library must be modified in order to successfully install METVIEW.



ECMWF Software used in NMA

- METVIEW
  - currently used in the NMA operational suite for both ECMWF and limited-area numerical models (ALADIN, HRM, LM), which are operationally run in NMA.
- ECMWF Web Site
  - EPS products from ECMWF web site are used in operative weather forecast.

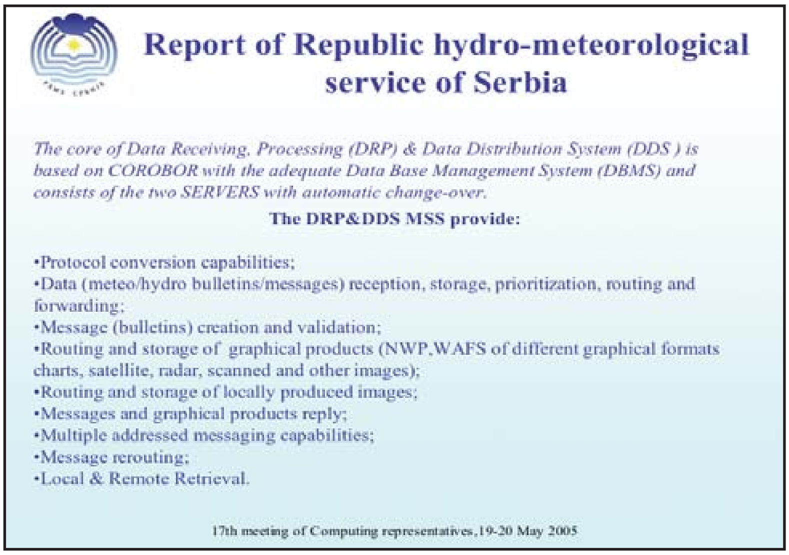


FUTURE PLANS

- SMS - to be installed on cluster.
- To be installed :
  - ECaccess
  - VPN Concentrator.
  - Crypto Card Server.

SERBIA & MONTENEGRO SERBIA & MONTENEGRO

Vladimir M. Dimitrijevic – Republic Hydro-Meteorological Service of Serbia



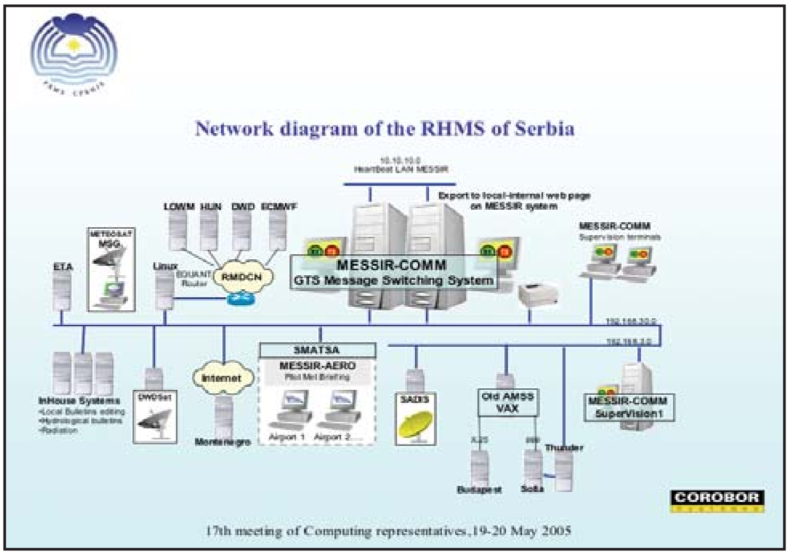
Report of Republic hydro-meteorological service of Serbia

The core of Data Receiving, Processing (DRP) & Data Distribution System (DDS) is based on COROBOR with the adequate Data Base Management System (DBMS) and consists of the two SERVERS with automatic change-over.

The DRP&DDS MSS provide:

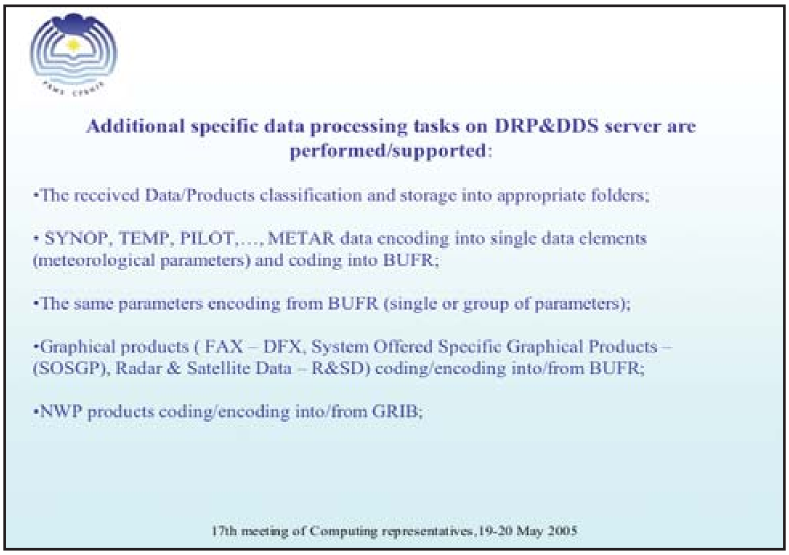
- •Protocol conversion capabilities;
- •Data (meteo/hydro bulletins/messages) reception, storage, prioritization, routing and forwarding;
- •Message (bulletins) creation and validation;
- •Routing and storage of graphical products (NWP,WAFS of different graphical formats charts, satellite, radar, scanned and other images);
- •Routing and storage of locally produced images;
- •Messages and graphical products reply;
- •Multiple addressed messaging capabilities;
- •Message rerouting;
- •Local & Remote Retrieval.

17th meeting of Computing representatives,19-20 May 2005



Network diagram of the RHMS of Serbia

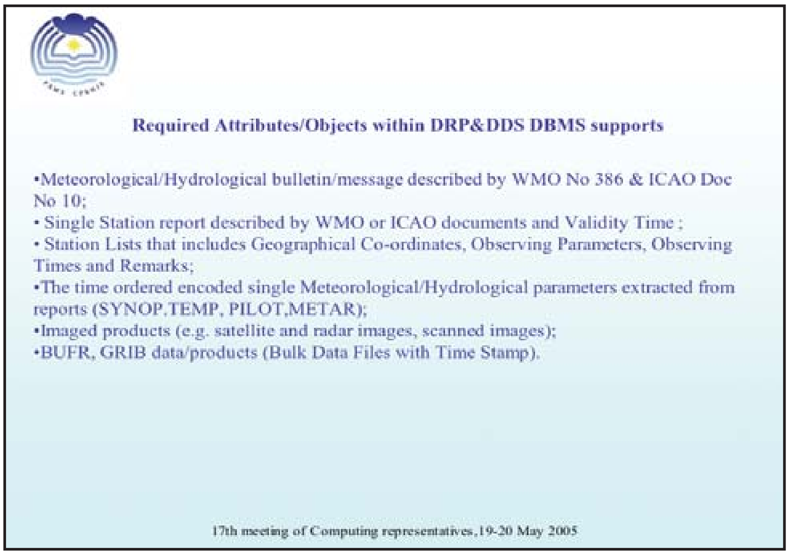
17th meeting of Computing representatives,19-20 May 2005



Additional specific data processing tasks on DRP&DDS server are performed/supported:

- •The received Data/Products classification and storage into appropriate folders;
- • SYNOP, TEMP, PILOT,…, METAR data encoding into single data elements (meteorological parameters) and coding into BUFR;
- •The same parameters encoding from BUFR (single or group of parameters);
- •Graphical products (FAX – DFX, System Offered Specific Graphical Products – (SOSGP), Radar & Satellite Data – R&SD) coding/encoding into/from BUFR;
- •NWP products coding/encoding into/from GRIB;

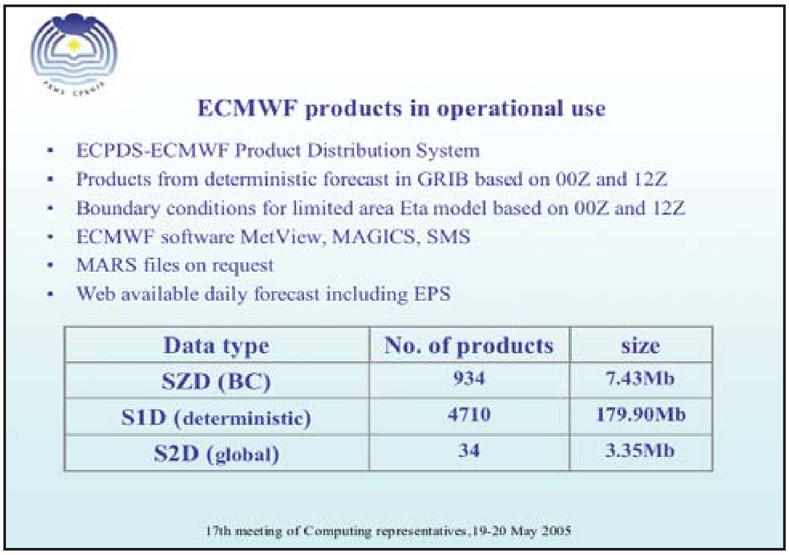
17th meeting of Computing representatives,19-20 May 2005



Required Attributes/Objects within DRP&DDS DBMS supports

•Meteorological/Hydrological bulletin/message described by WMO No 386 & ICAO Doc No 10;
• Single Station report described by WMO or ICAO documents and Validity Time ;
• Station Lists that includes Geographical Co-ordinates, Observing Parameters, Observing Times and Remarks;
•The time ordered encoded single Meteorological/Hydrological parameters extracted from reports (SYNOP,TEMP, PILOT,METAR);
•Imaged products (e.g. satellite and radar images, scanned images);
•BUFR, GRIB data/products (Bulk Data Files with Time Stamp).

17th meeting of Computing representatives,19-20 May 2005



ECMWF products in operational use

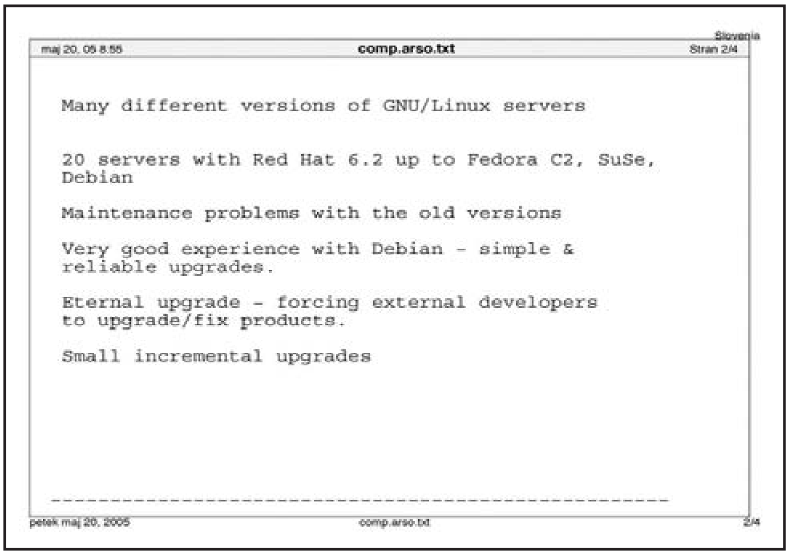
- ECPDS-ECMWF Product Distribution System
- Products from deterministic forecast in GRIB based on 00Z and 12Z
- Boundary conditions for limited area Eta model based on 00Z and 12Z
- ECMWF software MetView, MAGICS, SMS
- MARS files on request
- Web available daily forecast including EPS

Data type	No. of products	size
SZD (BC)	934	7.43Mb
S1D (deterministic)	4710	179.90Mb
S2D (global)	34	3.35Mb

17th meeting of Computing representatives,19-20 May 2005

SLOVENIA SLOVENIA

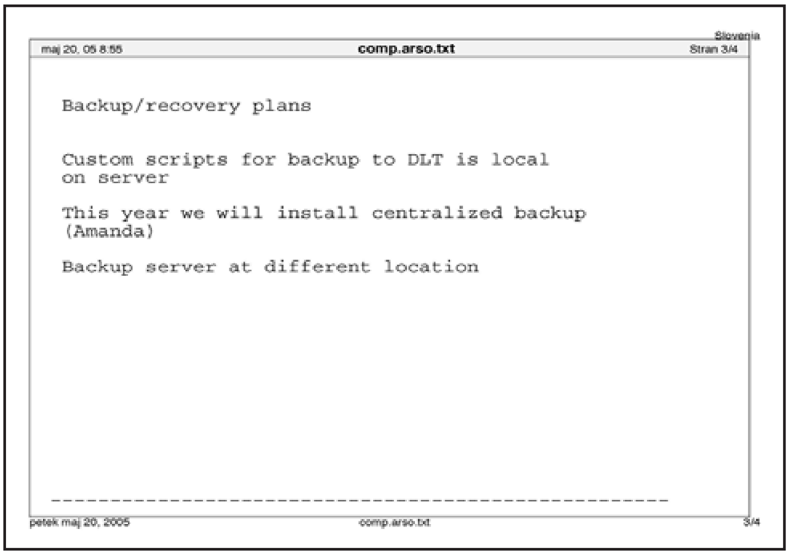
Petar Hitij – Environmental Agency of the Republic of Slovenia (EARS)


Slovenia

maj 20, 05 8:55 **comp.arso.txt** Stran 2/4

```
Many different versions of GNU/Linux servers

20 servers with Red Hat 6.2 up to Fedora C2, SuSe,
Debian

Maintenance problems with the old versions

Very good experience with Debian - simple &
reliable upgrades.

Eternal upgrade - forcing external developers
to upgrade/fix products.

Small incremental upgrades



----------------------------------------------------
```

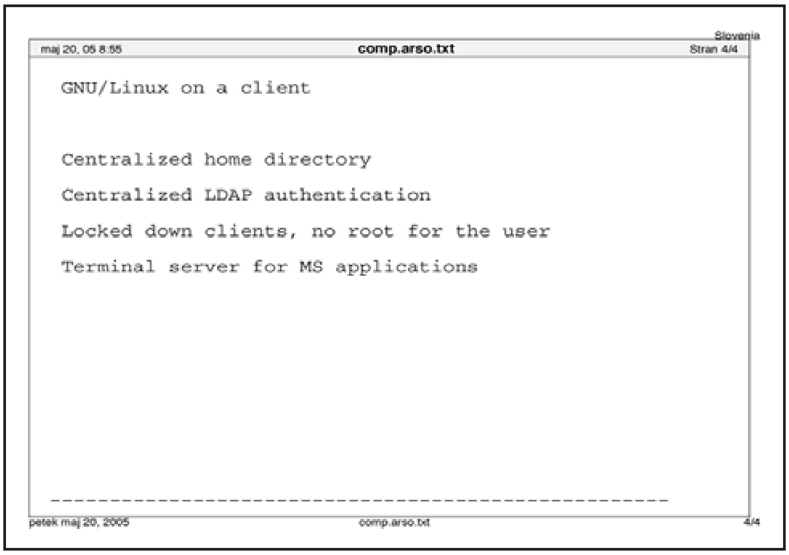
petek maj 20, 2005 comp.arso.txt 2/4


Slovenia

maj 20, 05 8:55 **comp.arso.txt** Stran 3/4

```
Backup/recovery plans

Custom scripts for backup to DLT is local
on server

This year we will install centralized backup
(Amanda)

Backup server at different location



----------------------------------------------------
```

petek maj 20, 2005 comp.arso.txt 3/4

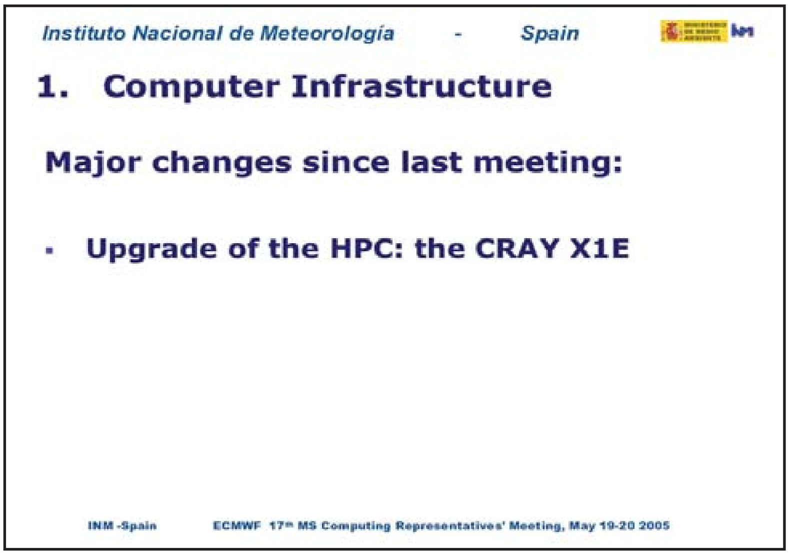

Slovenia

maj 20, 05 8:55 **comp.arso.txt** Stran 4/4

```
GNU/Linux on a client


Centralized home directory

Centralized LDAP authentication

Locked down clients, no root for the user

Terminal server for MS applications



----------------------------------------------------
```

petek maj 20, 2005 comp.arso.txt 4/4

Eduardo Monreal – Instituto Nacional de Meteorología



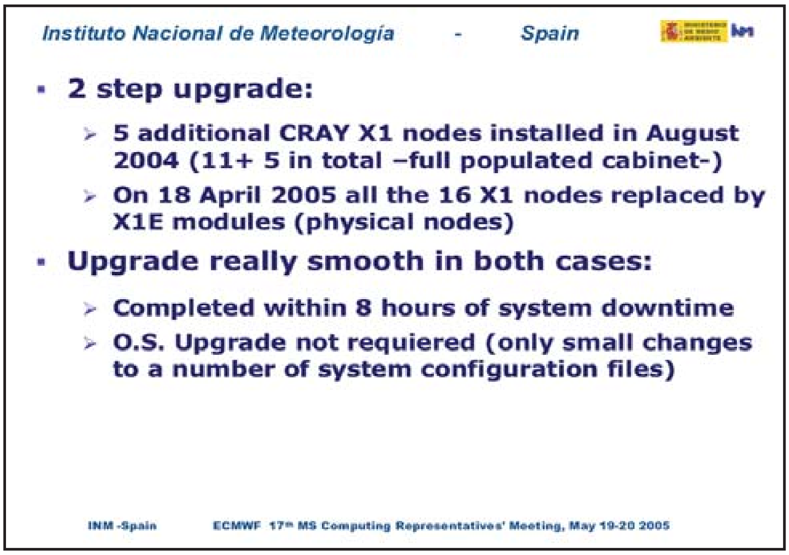
Instituto Nacional de Meteorología - Spain

1. Computer Infrastructure

Major changes since last meeting:

- Upgrade of the HPC: the CRAY X1E

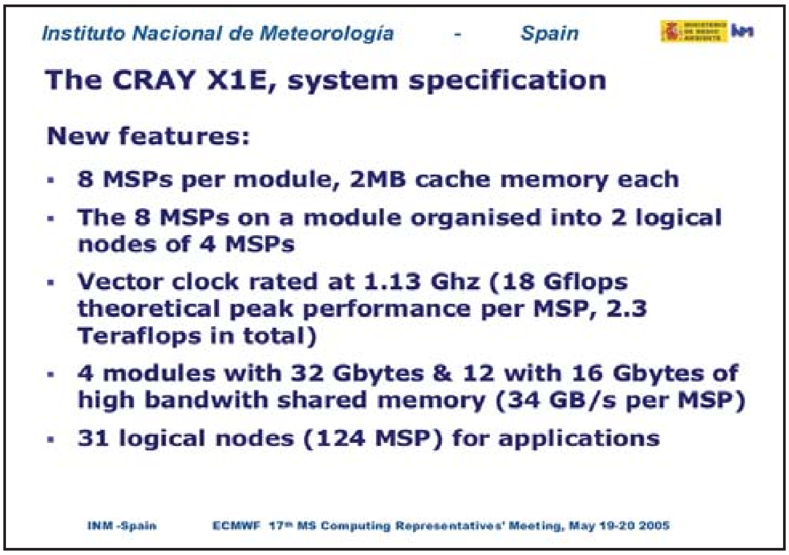
INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

- 2 step upgrade:
  - 5 additional CRAY X1 nodes installed in August 2004 (11+ 5 in total –full populated cabinet-)
  - On 18 April 2005 all the 16 X1 nodes replaced by X1E modules (physical nodes)
- Upgrade really smooth in both cases:
  - Completed within 8 hours of system downtime
  - O.S. Upgrade not requiered (only small changes to a number of system configuration files)

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



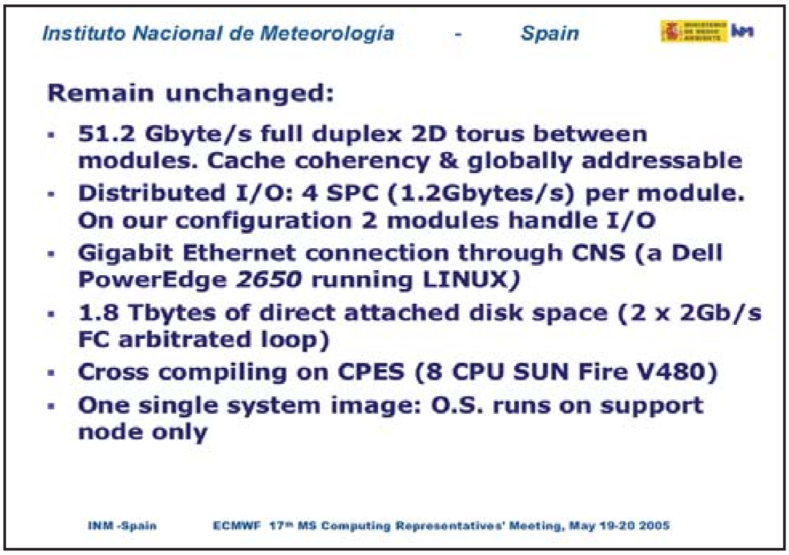
Instituto Nacional de Meteorología - Spain

The CRAY X1E, system specification

New features:

- 8 MSPs per module, 2MB cache memory each
- The 8 MSPs on a module organised into 2 logical nodes of 4 MSPs
- Vector clock rated at 1.13 Ghz (18 Gflops theoretical peak performance per MSP, 2.3 Teraflops in total)
- 4 modules with 32 Gbytes & 12 with 16 Gbytes of high bandwith shared memory (34 GB/s per MSP)
- 31 logical nodes (124 MSP) for applications

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

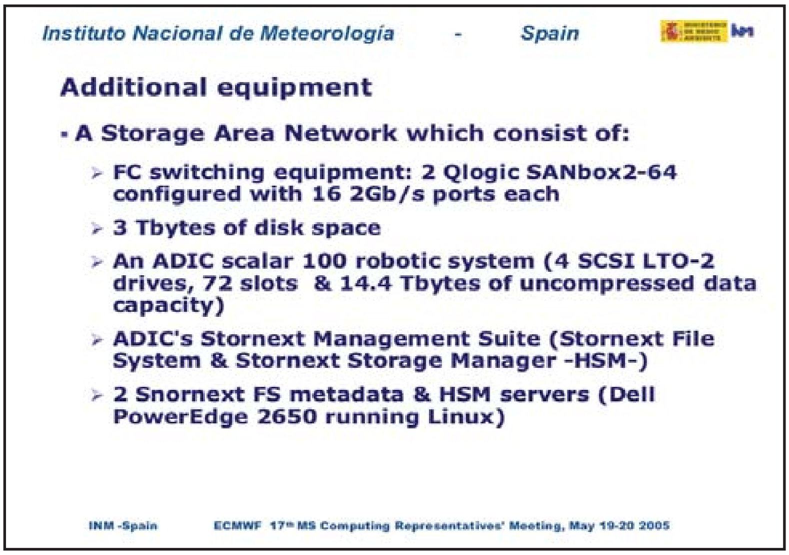


Instituto Nacional de Meteorología - Spain

Remain unchanged:

- 51.2 Gbyte/s full duplex 2D torus between modules. Cache coherency & globally addressable
- Distributed I/O: 4 SPC (1.2Gbytes/s) per module. On our configuration 2 modules handle I/O
- Gigabit Ethernet connection through CNS (a Dell PowerEdge *2650* running LINUX*)*
- 1.8 Tbytes of direct attached disk space (2 x 2Gb/s FC arbitrated loop)
- Cross compiling on CPES (8 CPU SUN Fire V480)
- One single system image: O.S. runs on support node only

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

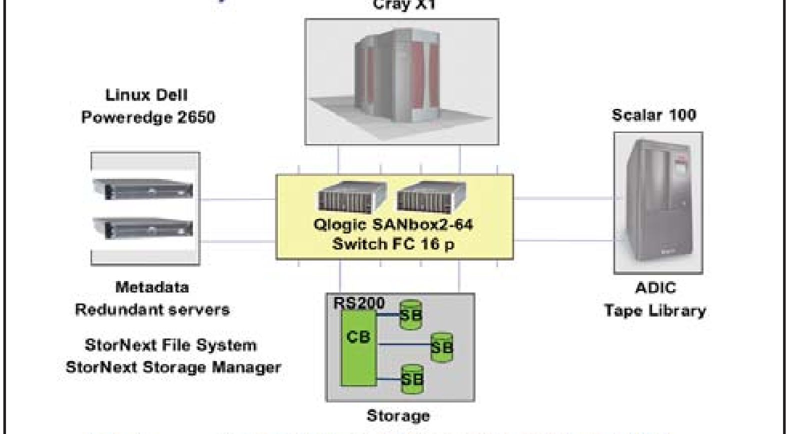
Additional equipment

- A Storage Area Network which consist of:
  - FC switching equipment: 2 Qlogic SANbox2-64 configured with 16 2Gb/s ports each
  - 3 Tbytes of disk space
  - An ADIC scalar 100 robotic system (4 SCSI LTO-2 drives, 72 slots & 14.4 Tbytes of uncompressed data capacity)
  - ADIC's Stornext Management Suite (Stornext File System & Stornext Storage Manager -HSM-)
  - 2 Snornext FS metadata & HSM servers (Dell PowerEdge 2650 running Linux)

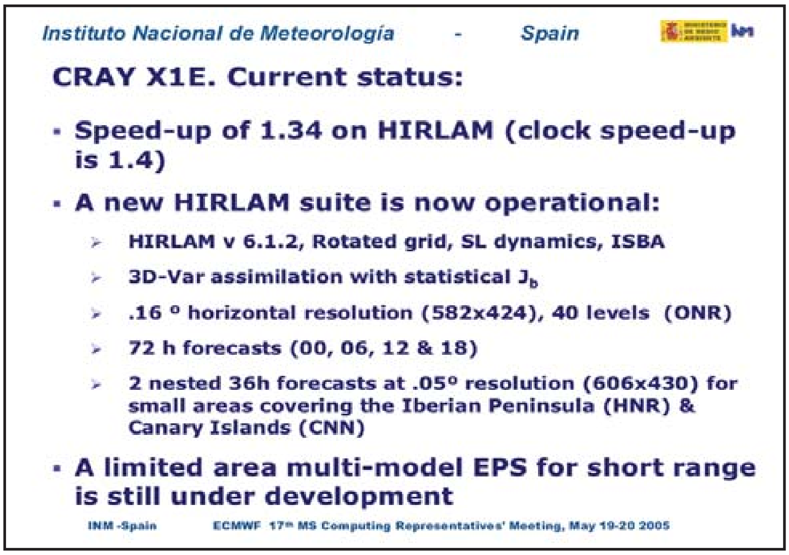
INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

Instituto Nacional de Meteorología - Spain

SAN layout



INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

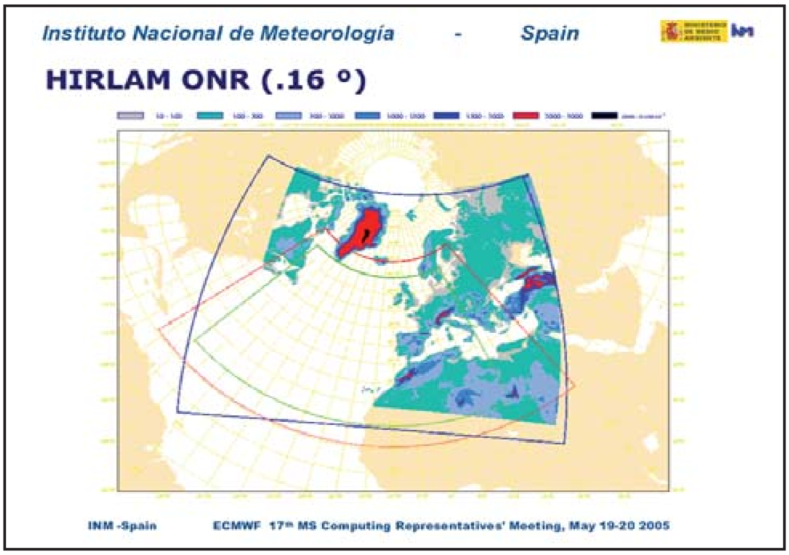


Instituto Nacional de Meteorología - Spain

CRAY X1E. Current status:

- Speed-up of 1.34 on HIRLAM (clock speed-up is 1.4)
- A new HIRLAM suite is now operational:
  - HIRLAM v 6.1.2, Rotated grid, SL dynamics, ISBA
  - 3D-Var assimilation with statistical J_b
  - .16 º horizontal resolution (582x424), 40 levels (ONR)
  - 72 h forecasts (00, 06, 12 & 18)
  - 2 nested 36h forecasts at .05º resolution (606x430) for small areas covering the Iberian Peninsula (HNR) & Canary Islands (CNN)
- A limited area multi-model EPS for short range is still under development

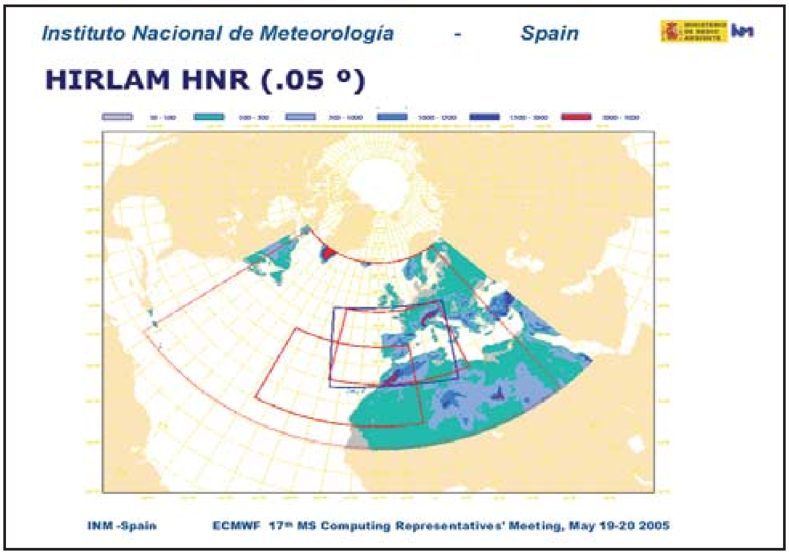
INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

HIRLAM ONR (.16 º)

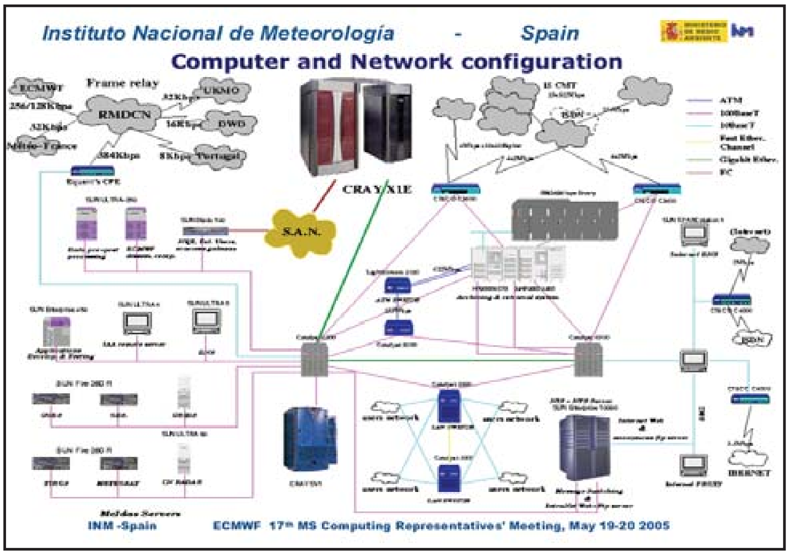
INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

HIRLAM HNR (.05 º)

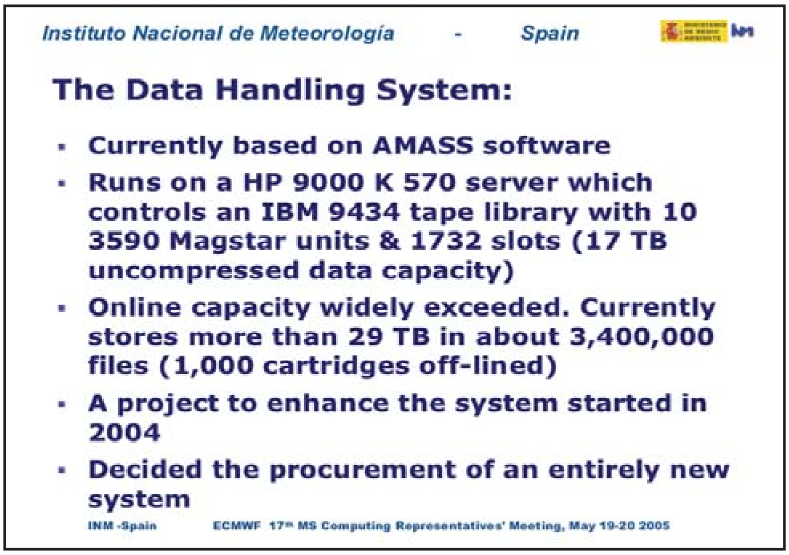
INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

Computer and Network configuration

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

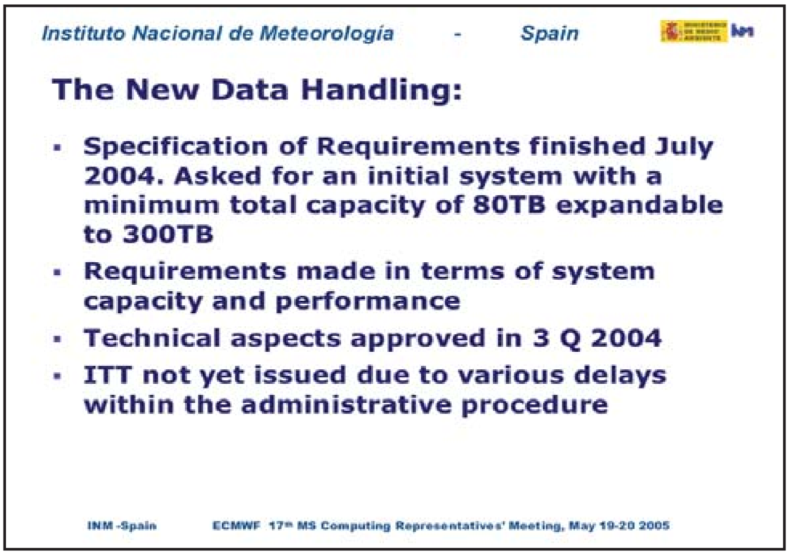


Instituto Nacional de Meteorología - Spain

The Data Handling System:

- Currently based on AMASS software
- Runs on a HP 9000 K 570 server which controls an IBM 9434 tape library with 10 3590 Magstar units & 1732 slots (17 TB uncompressed data capacity)
- Online capacity widely exceeded. Currently stores more than 29 TB in about 3,400,000 files (1,000 cartridges off-lined)
- A project to enhance the system started in 2004
- Decided the procurement of an entirely new system

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

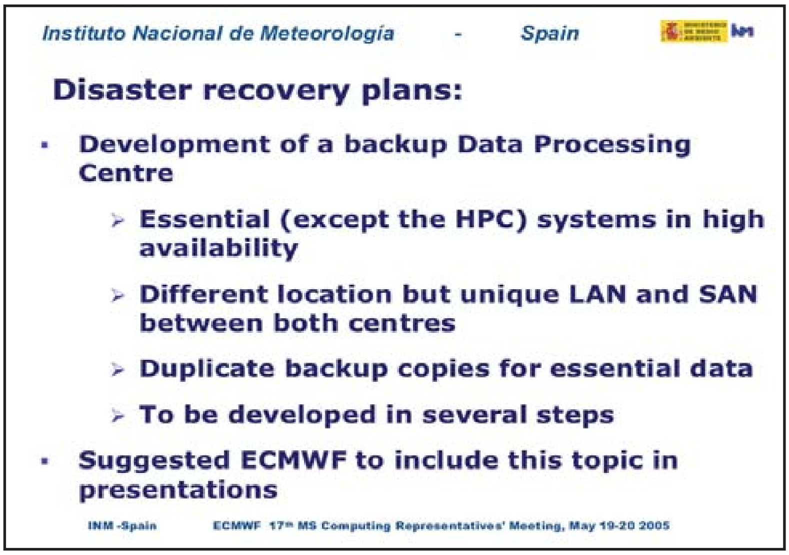


Instituto Nacional de Meteorología - Spain

The New Data Handling:

- Specification of Requirements finished July 2004. Asked for an initial system with a minimum total capacity of 80TB expandable to 300TB
- Requirements made in terms of system capacity and performance
- Technical aspects approved in 3 Q 2004
- ITT not yet issued due to various delays within the administrative procedure

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

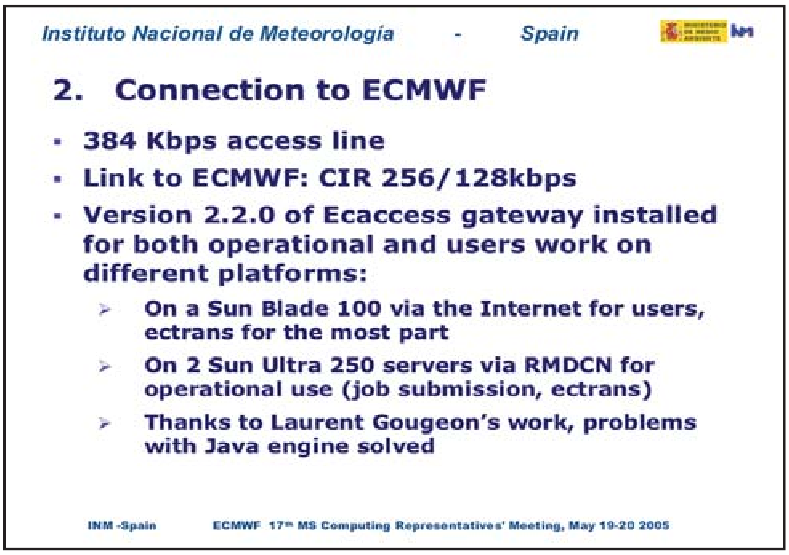


Instituto Nacional de Meteorología - Spain

Disaster recovery plans:

- Development of a backup Data Processing Centre
  - Essential (except the HPC) systems in high availability
  - Different location but unique LAN and SAN between both centres
  - Duplicate backup copies for essential data
  - To be developed in several steps
- Suggested ECMWF to include this topic in presentations

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

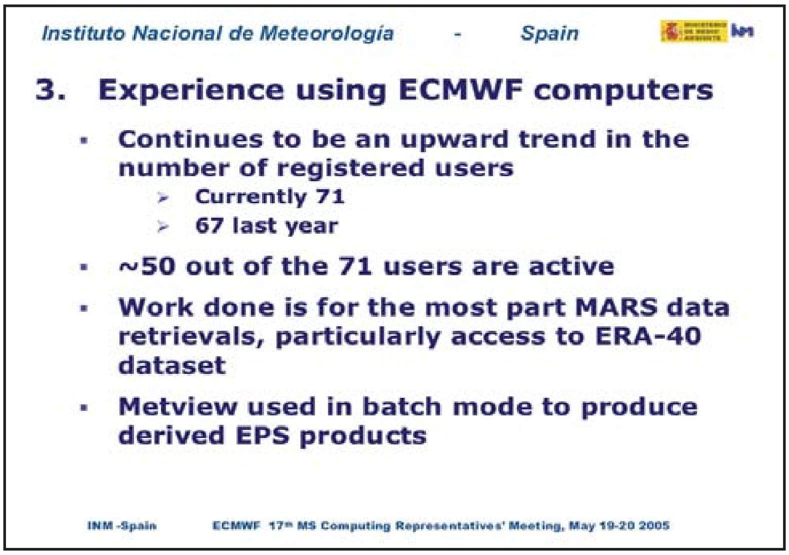


Instituto Nacional de Meteorología - Spain

2. Connection to ECMWF

- 384 Kbps access line
- Link to ECMWF: CIR 256/128kbps
- Version 2.2.0 of Ecaccess gateway installed for both operational and users work on different platforms:
  - On a Sun Blade 100 via the Internet for users, ectrans for the most part
  - On 2 Sun Ultra 250 servers via RMDCN for operational use (job submission, ectrans)
  - Thanks to Laurent Gougeon's work, problems with Java engine solved

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

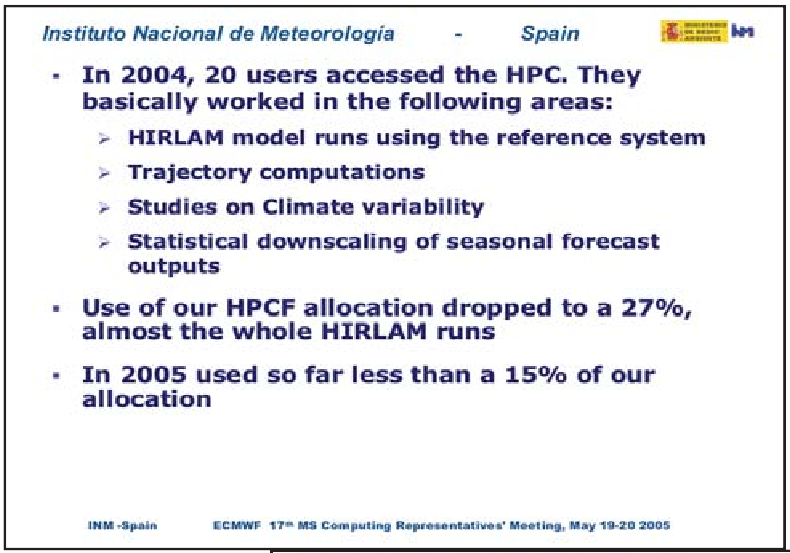


Instituto Nacional de Meteorología - Spain

3. Experience using ECMWF computers

- Continues to be an upward trend in the number of registered users
  - Currently 71
  - 67 last year
- ~50 out of the 71 users are active
- Work done is for the most part MARS data retrievals, particularly access to ERA-40 dataset
- Metview used in batch mode to produce derived EPS products

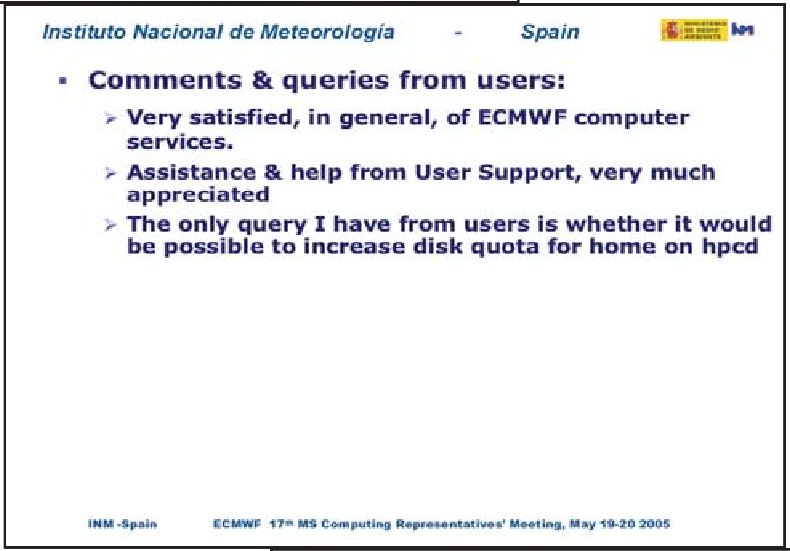
INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

- **In 2004, 20 users accessed the HPC. They basically worked in the following areas:**
  - **HIRLAM model runs using the reference system**
  - **Trajectory computations**
  - **Studies on Climate variability**
  - **Statistical downscaling of seasonal forecast outputs**
- **Use of our HPCF allocation dropped to a 27%, almost the whole HIRLAM runs**
- **In 2005 used so far less than a 15% of our allocation**

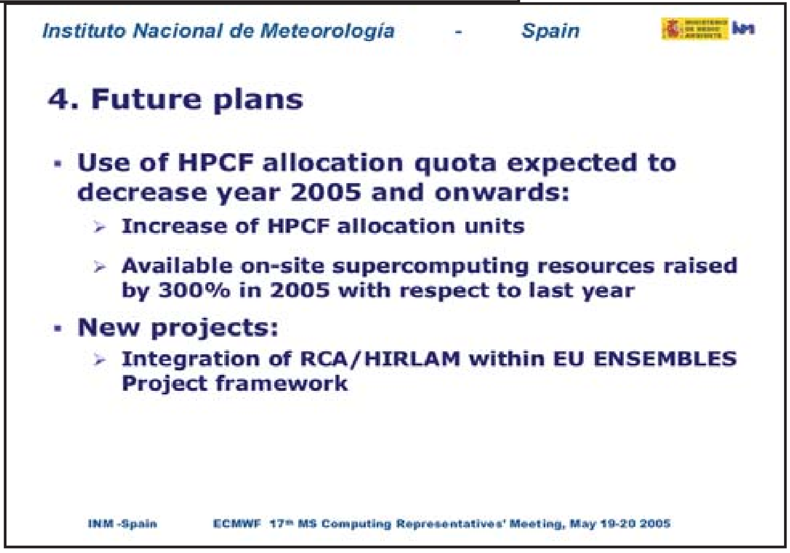
INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

- **Comments & queries from users:**
  - **Very satisfied, in general, of ECMWF computer services.**
  - **Assistance & help from User Support, very much appreciated**
  - **The only query I have from users is whether it would be possible to increase disk quota for home on hpcd**

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005



Instituto Nacional de Meteorología - Spain

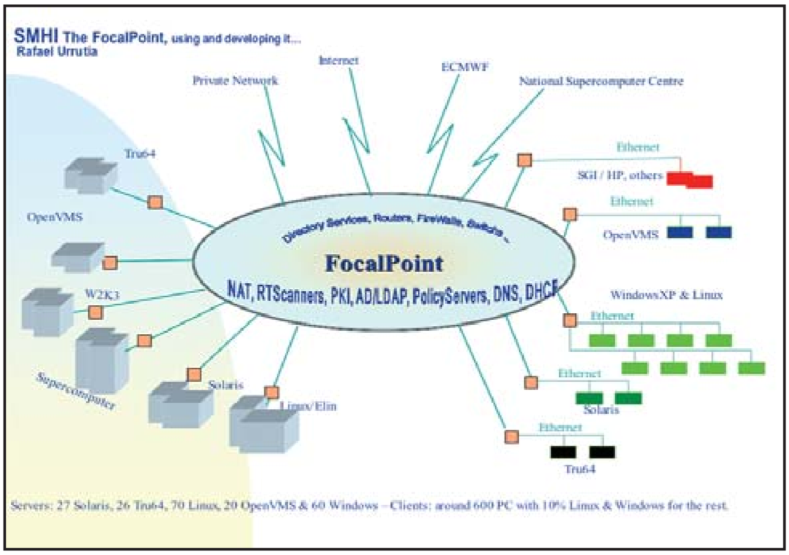
4. Future plans

- **Use of HPCF allocation quota expected to decrease year 2005 and onwards:**
  - **Increase of HPCF allocation units**
  - **Available on-site supercomputing resources raised by 300% in 2005 with respect to last year**
- **New projects:**
  - **Integration of RCA/HIRLAM within EU ENSEMBLES Project framework**

INM -Spain ECMWF 17th MS Computing Representatives' Meeting, May 19-20 2005

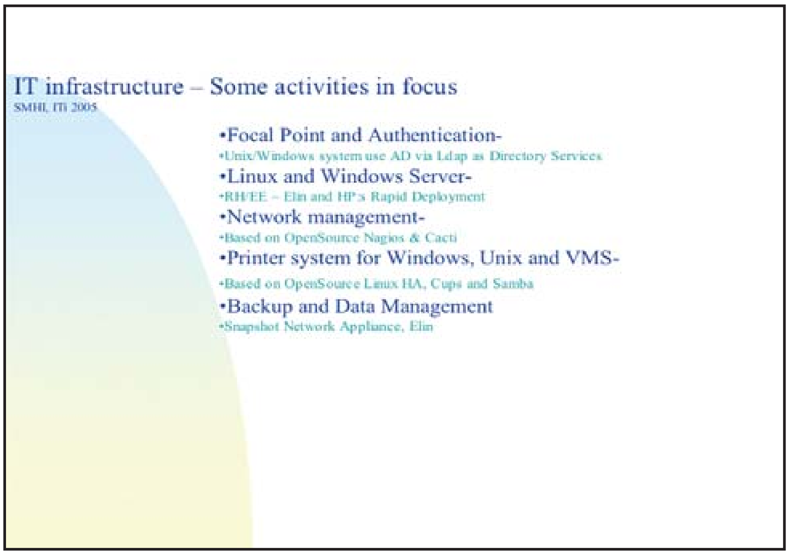
In response to the request for more disk space in /home, for instance for the maintenance of source code, U. Modigliani pointed out that larger quotas were available under /ms_perm, though users must be aware that there are no automatic back ups of this space and must make their own backup arrangements.

Rafael Urrutia – Swedish Meteorological and Hydrological Institute (SMHI)



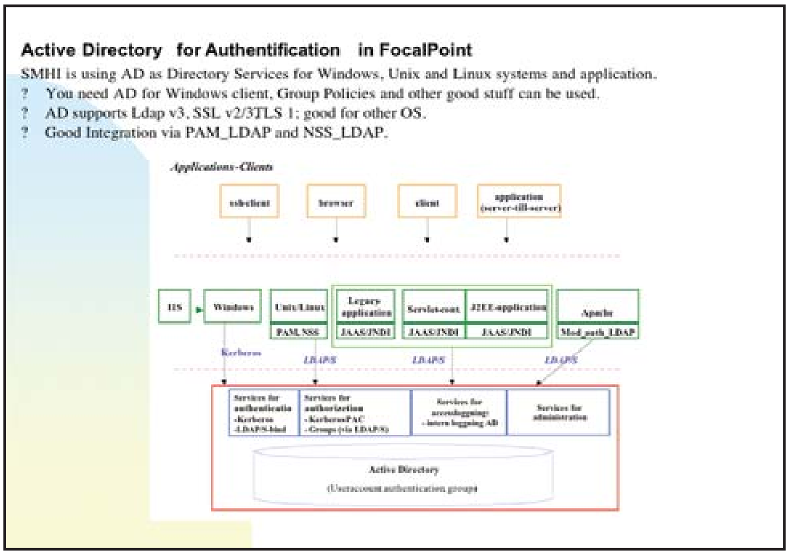
SMHI The FocalPoint, using and developing it...
Rafael Urrutia

Servers: 27 Solaris, 26 Tru64, 70 Linux, 20 OpenVMS & 60 Windows – Clients: around 600 PC with 10% Linux & Windows for the rest.



IT infrastructure – Some activities in focus
SMHI, ITi 2005

- •Focal Point and Authentication-
  - •Unix/Windows system use AD via Ldap as Directory Services
- •Linux and Windows Server-
  - •RH/EE – Elin and HP:s Rapid Deployment
- •Network management-
  - •Based on OpenSource Nagios & Cacti
- •Printer system for Windows, Unix and VMS-
  - •Based on OpenSource Linux HA, Cups and Samba
- •Backup and Data Management
  - •Snapshot Network Appliance, Elin



Active Directory for Authentification in FocalPoint

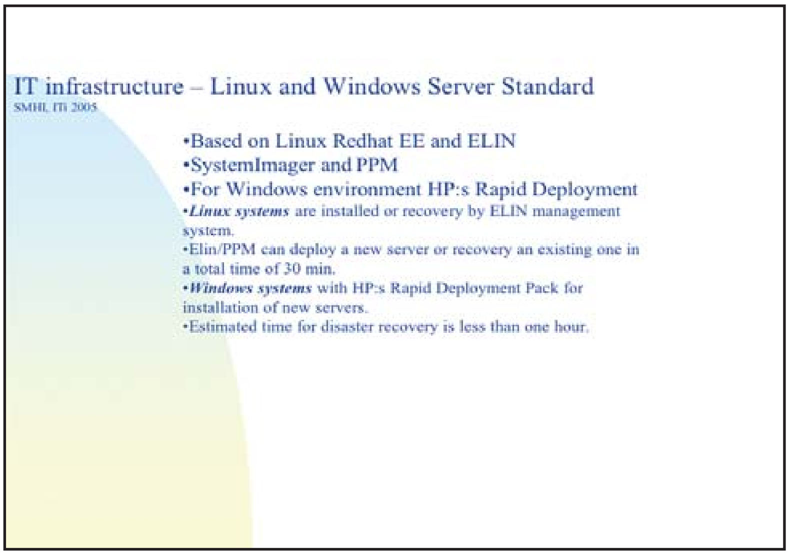
SMHI is using AD as Directory Services for Windows, Unix and Linux systems and application.

? You need AD for Windows client, Group Policies and other good stuff can be used.

? AD supports Ldap v3, SSL v2/3TLS 1; good for other OS.

? Good Integration via PAM_LDAP and NSS_LDAP.

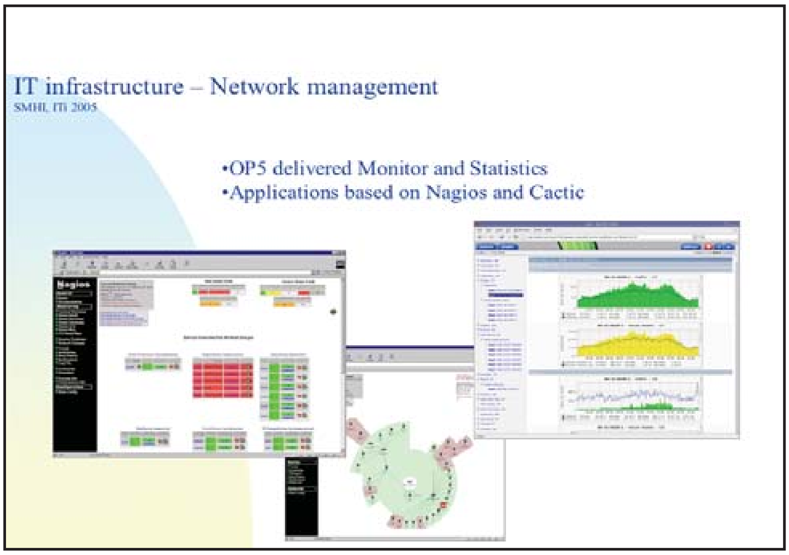
Applications-Clients
sshclient
browser
client
application (server-till-server)
IIS
Windows
Unix/Linux
PAM, NSS
Legacy application
JAAS/JNDI
Servlet-cont.
JAAS/JNDI
J2EE-application
JAAS/JNDI
Apache
Mod_auth_LDAP
Kerberos
LDAP/S
LDAP/S
LDAP/S
Services for authenticatio -Kerberos -LDAP/S-bind
Services for authorization -KerberosPAC -Groups (via LDAP/S)
Services for access/logging: - intern logging AD
Services for administration
Active Directory
(Useraccount authentication groups)



IT infrastructure – Linux and Windows Server Standard

SMHI, ITi 2005

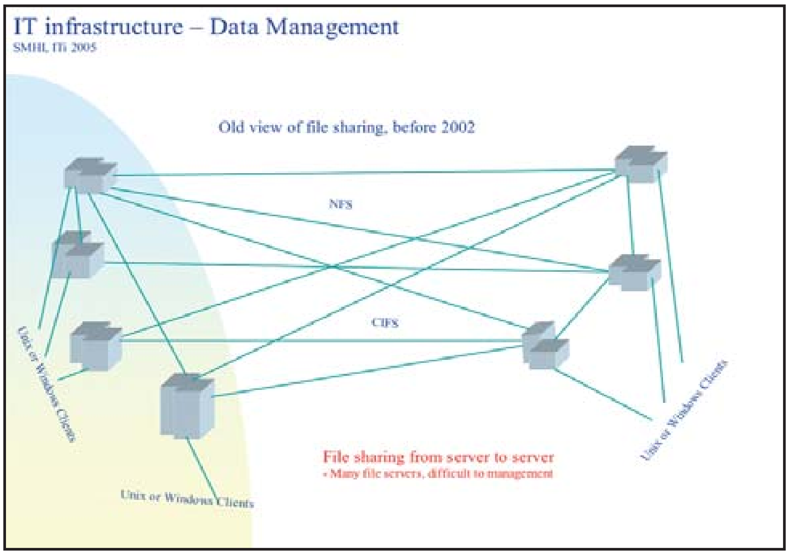
- •Based on Linux Redhat EE and ELIN
- •SystemImager and PPM
- •For Windows environment HP:s Rapid Deployment
- •***Linux systems*** are installed or recovery by ELIN management system.
- •Elin/PPM can deploy a new server or recovery an existing one in a total time of 30 min.
- •***Windows systems*** with HP:s Rapid Deployment Pack for installation of new servers.
- •Estimated time for disaster recovery is less than one hour.



IT infrastructure – Network management

SMHI, ITi 2005

- •OP5 delivered Monitor and Statistics
- •Applications based on Nagios and Cactic



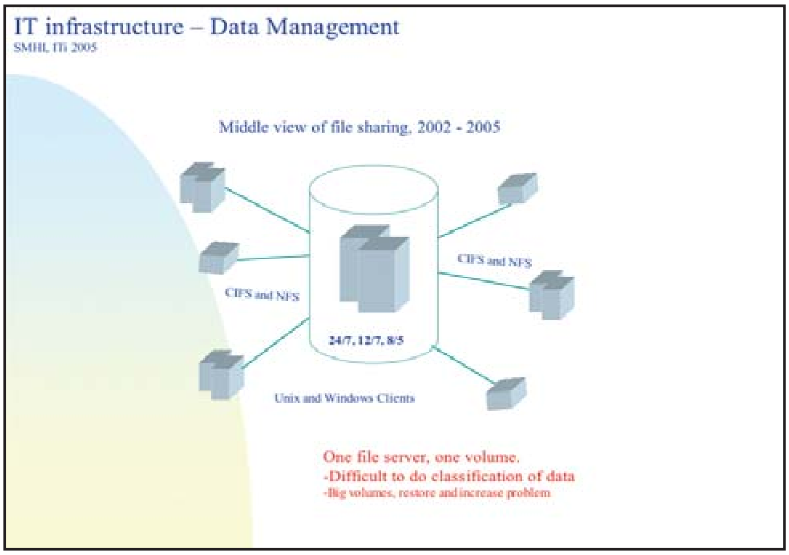
IT infrastructure – Data Management

SMHI, ITi 2005

Old view of file sharing, before 2002

File sharing from server to server

• Many file servers, difficult to management



IT infrastructure – Data Management

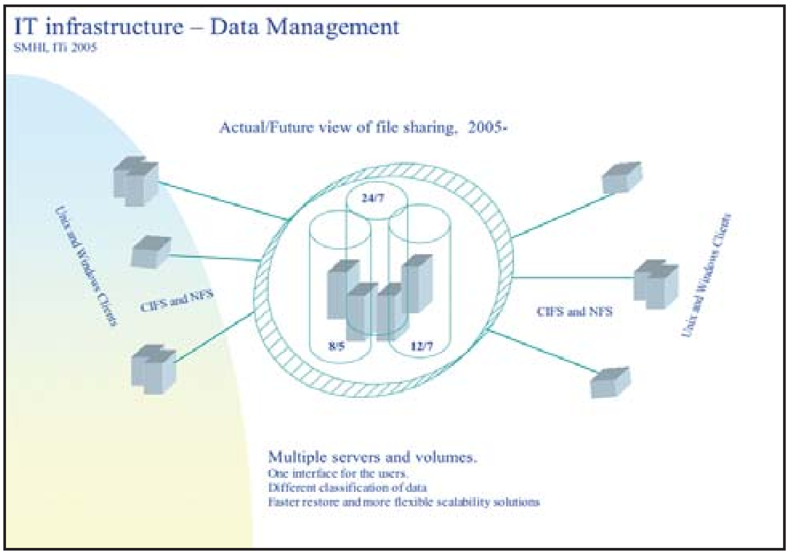
SMHI, ITi 2005

Middle view of file sharing, 2002 - 2005

One file server, one volume.

-Difficult to do classification of data

-Big volumes, restore and increase problem



IT infrastructure – Data Management

SMHI, ITi 2005

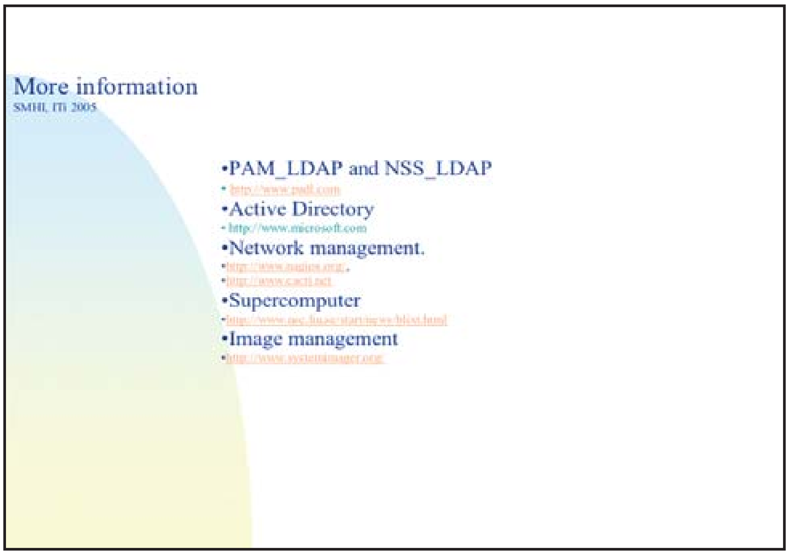
Actual/Future view of file sharing, 2005-

Multiple servers and volumes.

One interface for the users.

Different classification of data

Faster restore and more flexible scalability solutions



More information

SMHI, ITi 2005

•PAM_LDAP and NSS_LDAP

• http://www.padl.com

•Active Directory

• http://www.microsoft.com

•Network management.

•http://www.nagios.org/,

•http://www.cacti.net

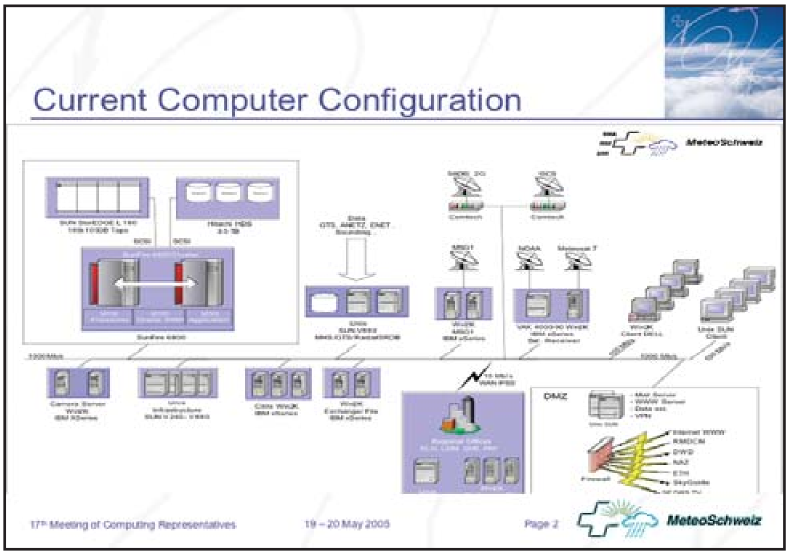
•Supercomputer

•http://www.nsc.liu.se/start/news/blixt.html

•Image management

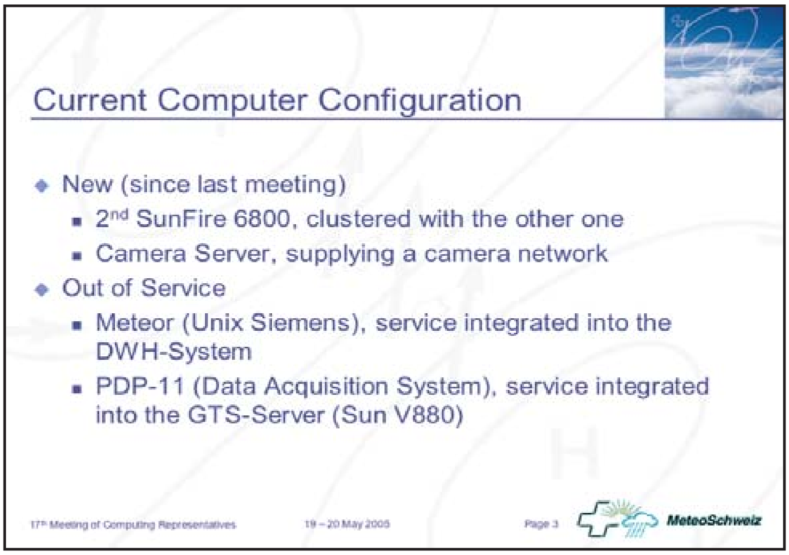
•http://www.systemimager.org

Peter Roth – MeteoSwiss



Current Computer Configuration

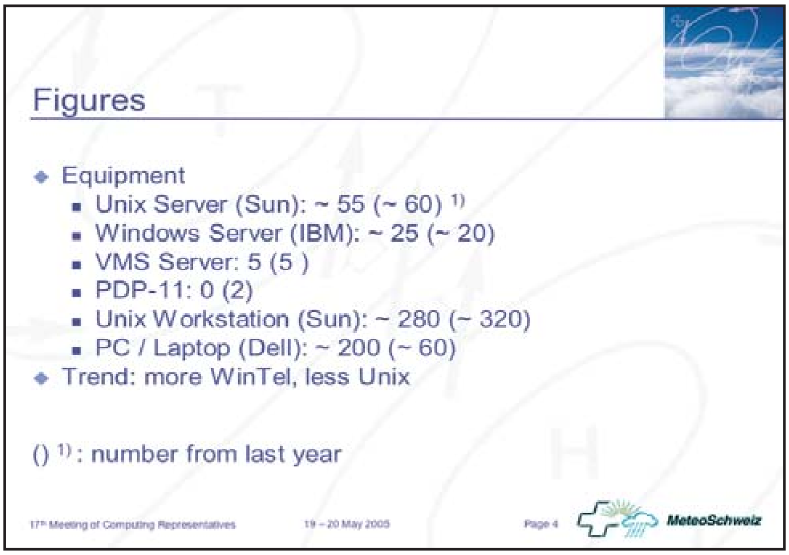
17th Meeting of Computing Representatives 19 – 20 May 2005 Page 2 MeteoSchweiz



Current Computer Configuration

- New (since last meeting)
  - 2nd SunFire 6800, clustered with the other one
  - Camera Server, supplying a camera network
- Out of Service
  - Meteor (Unix Siemens), service integrated into the DWH-System
  - PDP-11 (Data Acquisition System), service integrated into the GTS-Server (Sun V880)

17th Meeting of Computing Representatives 19 – 20 May 2005 Page 3 MeteoSchweiz

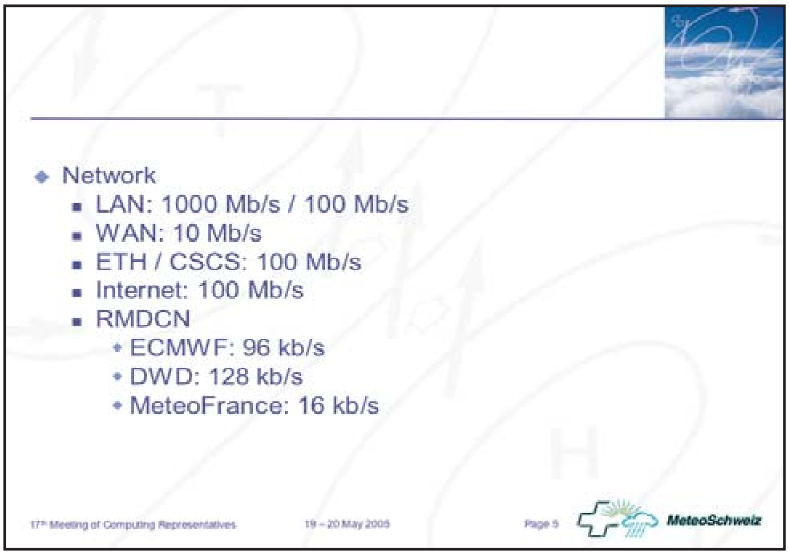


Figures

- Equipment
  - Unix Server (Sun): ~ 55 (~ 60) 1)
  - Windows Server (IBM): ~ 25 (~ 20)
  - VMS Server: 5 (5)
  - PDP-11: 0 (2)
  - Unix Workstation (Sun): ~ 280 (~ 320)
  - PC / Laptop (Dell): ~ 200 (~ 60)
- Trend: more WinTel, less Unix

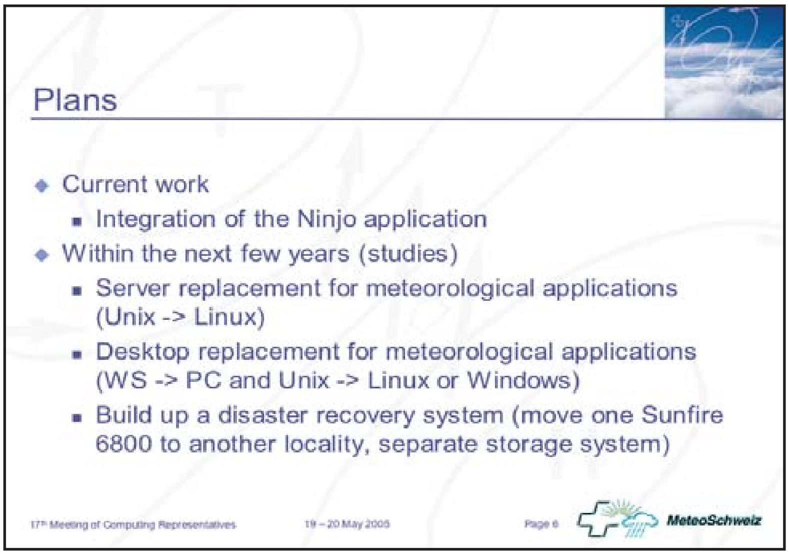
() 1) : number from last year

17th Meeting of Computing Representatives 19 – 20 May 2005 Page 4 MeteoSchweiz



- Network
  - LAN: 1000 Mb/s / 100 Mb/s
  - WAN: 10 Mb/s
  - ETH / CSCS: 100 Mb/s
  - Internet: 100 Mb/s
  - RMDCN
    - ECMWF: 96 kb/s
    - DWD: 128 kb/s
    - MeteoFrance: 16 kb/s

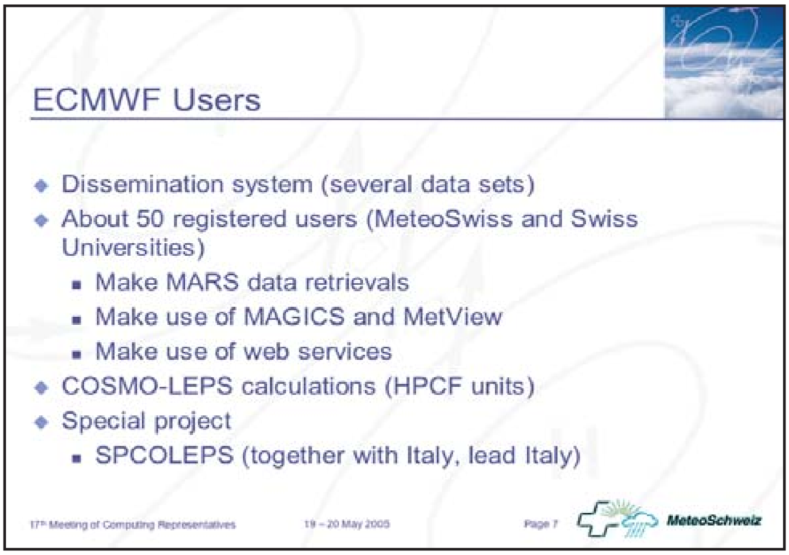
17th Meeting of Computing Representatives 19 – 20 May 2005 Page 5 MeteoSchweiz



Plans

- Current work
  - Integration of the Ninjo application
- Within the next few years (studies)
  - Server replacement for meteorological applications (Unix -> Linux)
  - Desktop replacement for meteorological applications (WS -> PC and Unix -> Linux or Windows)
  - Build up a disaster recovery system (move one Sunfire 6800 to another locality, separate storage system)

17th Meeting of Computing Representatives 19 – 20 May 2005 Page 6 MeteoSchweiz

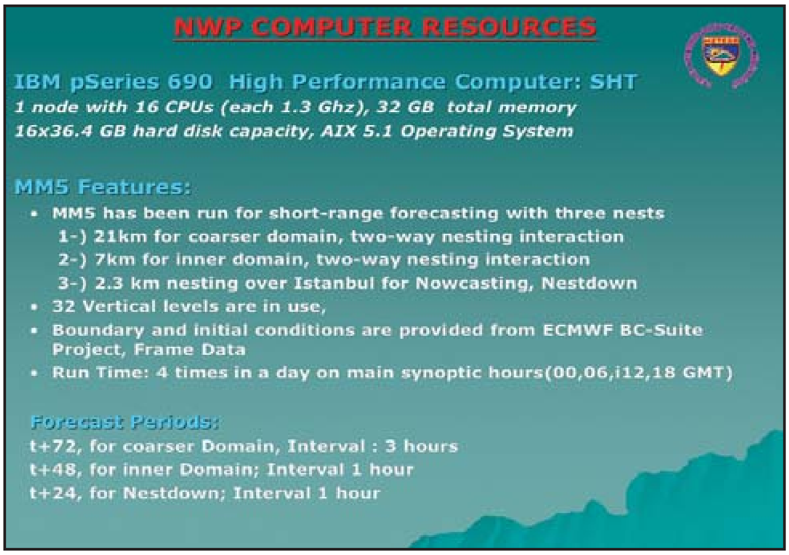


ECMWF Users

- Dissemination system (several data sets)
- About 50 registered users (MeteoSwiss and Swiss Universities)
  - Make MARS data retrievals
  - Make use of MAGICS and MetView
  - Make use of web services
- COSMO-LEPS calculations (HPCF units)
- Special project
  - SPCOLEPS (together with Italy, lead Italy)

17th Meeting of Computing Representatives 19 – 20 May 2005 Page 7 MeteoSchweiz

Ahmet Erturk – Turkish State Meteorological Service (TSMS)



NWP COMPUTER RESOURCES

IBM pSeries 690 High Performance Computer: SHT

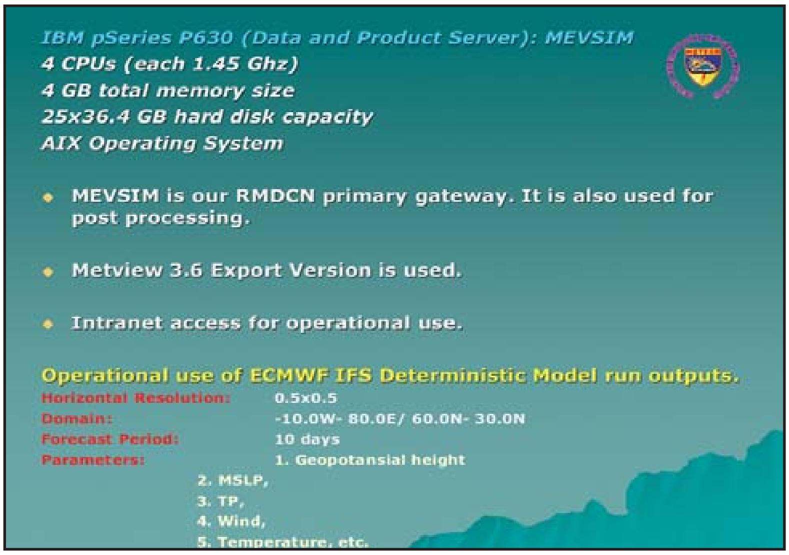
1 node with 16 CPUs (each 1.3 Ghz), 32 GB total memory
16x36.4 GB hard disk capacity, AIX 5.1 Operating System

MM5 Features:

- MM5 has been run for short-range forecasting with three nests
  1-) 21km for coarser domain, two-way nesting interaction
  2-) 7km for inner domain, two-way nesting interaction
  3-) 2.3 km nesting over Istanbul for Nowcasting, Nestdown
- 32 Vertical levels are in use,
- Boundary and initial conditions are provided from ECMWF BC-Suite Project, Frame Data
- Run Time: 4 times in a day on main synoptic hours(00,06,i12,18 GMT)

Forecast Periods:

t+72, for coarser Domain, Interval : 3 hours
t+48, for inner Domain; Interval 1 hour
t+24, for Nestdown; Interval 1 hour

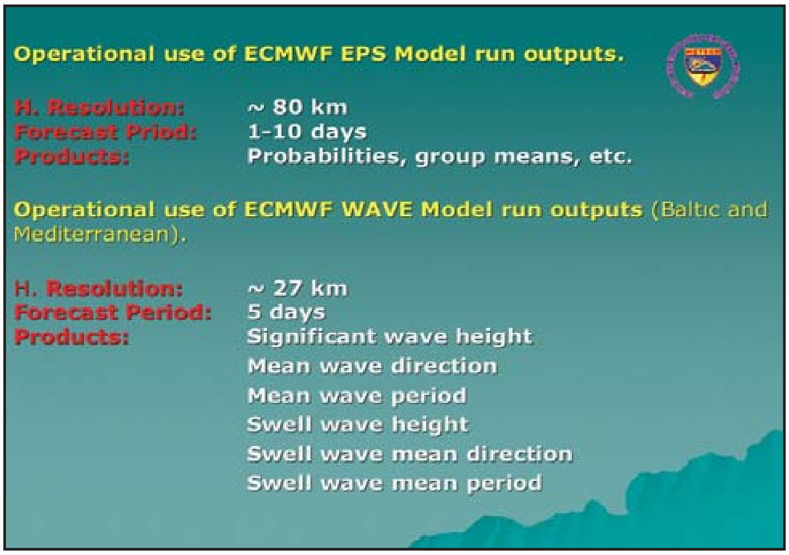


IBM pSeries P630 (Data and Product Server): MEVSIM
4 CPUs (each 1.45 Ghz)
4 GB total memory size
25x36.4 GB hard disk capacity
AIX Operating System

- MEVSIM is our RMDCN primary gateway. It is also used for post processing.
- Metview 3.6 Export Version is used.
- Intranet access for operational use.

Operational use of ECMWF IFS Deterministic Model run outputs.

Horizontal Resolution:	0.5x0.5
Domain:	-10.0W- 80.0E/ 60.0N- 30.0N
Forecast Period:	10 days
Parameters:	1. Geopotansial height 2. MSLP, 3. TP, 4. Wind, 5. Temperature, etc.

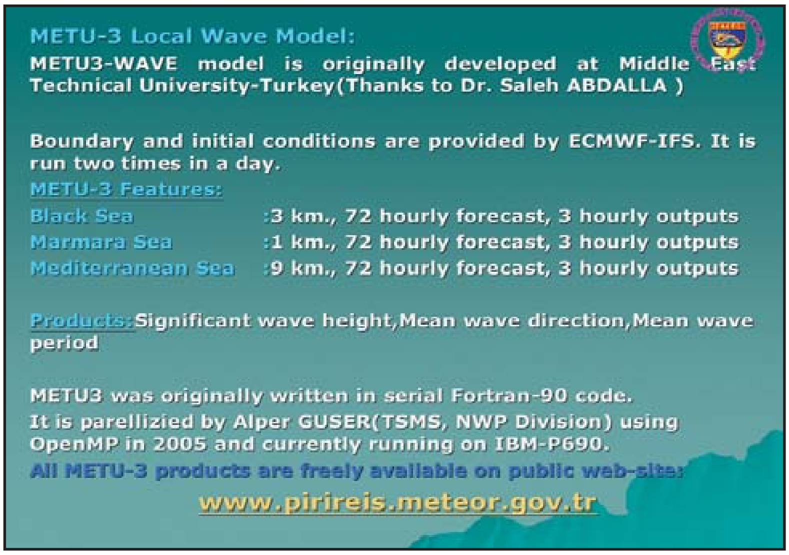


Operational use of ECMWF EPS Model run outputs.

H. Resolution:	~ 80 km
Forecast Priod:	1-10 days
Products:	Probabilities, group means, etc.

Operational use of ECMWF WAVE Model run outputs (Baltic and Mediterranean).

H. Resolution:	~ 27 km
Forecast Period:	5 days
Products:	Significant wave height Mean wave direction Mean wave period Swell wave height Swell wave mean direction Swell wave mean period



METU-3 Local Wave Model:

METU3-WAVE model is originally developed at Middle East Technical University-Turkey(Thanks to Dr. Saleh ABDALLA)

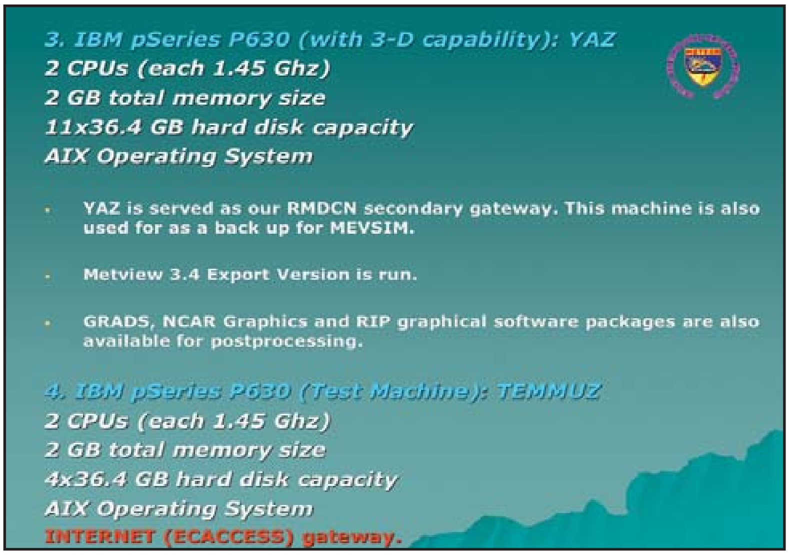
Boundary and initial conditions are provided by ECMWF-IFS. It is run two times in a day.

METU-3 Features:

Black Sea	:3 km., 72 hourly forecast, 3 hourly outputs
Marmara Sea	:1 km., 72 hourly forecast, 3 hourly outputs
Mediterranean Sea	:9 km., 72 hourly forecast, 3 hourly outputs

Products:Significant wave height,Mean wave direction,Mean wave period

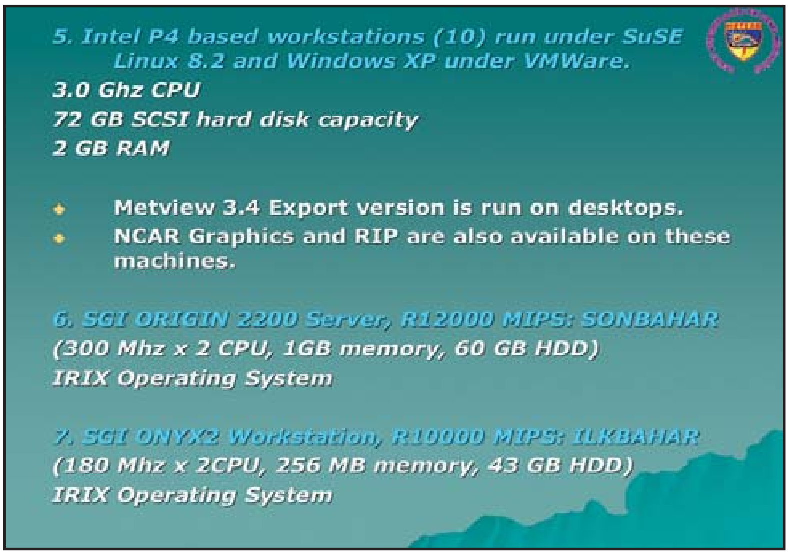
METU3 was originally written in serial Fortran-90 code.
It is parellizied by Alper GUSER(TSMS, NWP Division) using OpenMP in 2005 and currently running on IBM-P690.
All METU-3 products are freely available on public web-site:
www.pirireis.meteor.gov.tr



3. IBM pSeries P630 (with 3-D capability): YAZ
2 CPUs (each 1.45 Ghz)
2 GB total memory size
11x36.4 GB hard disk capacity
AIX Operating System

- YAZ is served as our RMDCN secondary gateway. This machine is also used for as a back up for MEVSIM.
- Metview 3.4 Export Version is run.
- GRADS, NCAR Graphics and RIP graphical software packages are also available for postprocessing.

4. IBM pSeries P630 (Test Machine): TEMMUZ
2 CPUs (each 1.45 Ghz)
2 GB total memory size
4x36.4 GB hard disk capacity
AIX Operating System
INTERNET (ECACCESS) gateway.

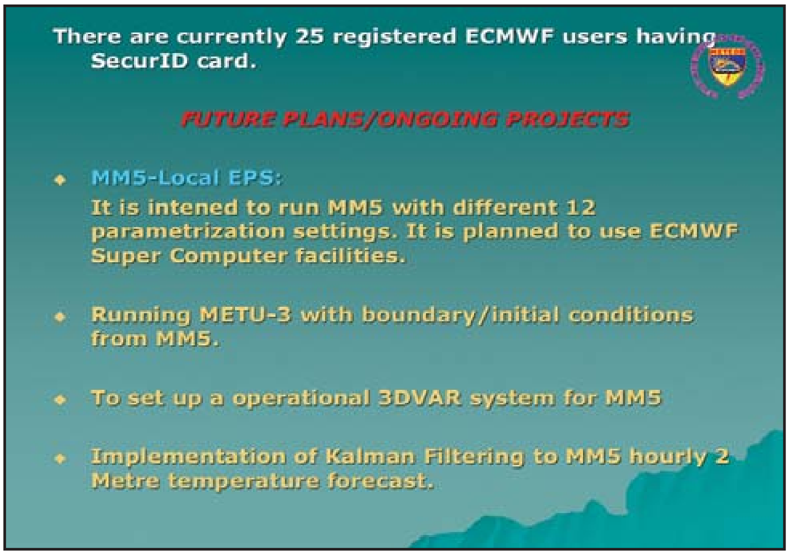


5. Intel P4 based workstations (10) run under SuSE Linux 8.2 and Windows XP under VMWare.
3.0 Ghz CPU
72 GB SCSI hard disk capacity
2 GB RAM

- Metview 3.4 Export version is run on desktops.
- NCAR Graphics and RIP are also available on these machines.

6. SGI ORIGIN 2200 Server, R12000 MIPS: SONBAHAR
(300 Mhz x 2 CPU, 1GB memory, 60 GB HDD)
IRIX Operating System

7. SGI ONYX2 Workstation, R10000 MIPS: ILKBAHAR
(180 Mhz x 2CPU, 256 MB memory, 43 GB HDD)
IRIX Operating System

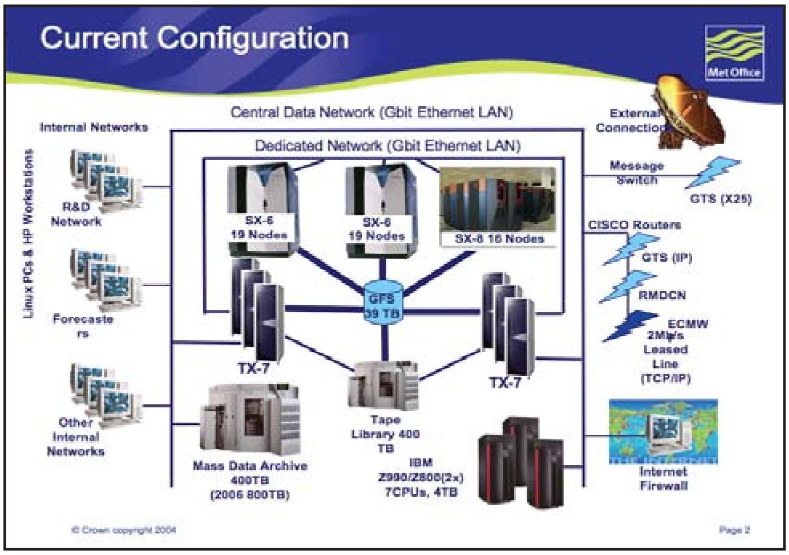


There are currently 25 registered ECMWF users having SecurID card.

FUTURE PLANS/ONGOING PROJECTS

- MM5-Local EPS:
  It is intened to run MM5 with different 12 parametrization settings. It is planned to use ECMWF Super Computer facilities.
- Running METU-3 with boundary/initial conditions from MM5.
- To set up a operational 3DVAR system for MM5
- Implementation of Kalman Filtering to MM5 hourly 2 Metre temperature forecast.

Roddy Sharp – Met Office, Exeter



Current Configuration

Met Office

Internal Networks

Linux PCs & HP Workstations

R&D Network

Forecasters

Other Internal Networks

Central Data Network (Gbit Ethernet LAN)

Dedicated Network (Gbit Ethernet LAN)

SX-6 19 Nodes

SX-6 19 Nodes

SX-8 16 Nodes

GFS 39 TB

TX-7

TX-7

Mass Data Archive 400TB (2006 800TB)

Tape Library 400 TB

IBM Z990/Z800(2x) 7CPUs, 4TB

External Connection

Message Switch

GTS (X25)

CISCO Routers

GTS (IP)

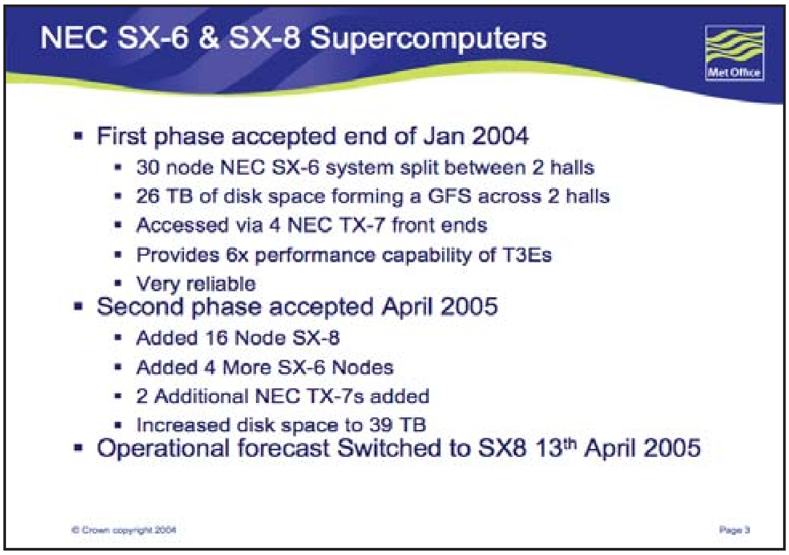
RMDCN

ECMW 2Mb/s Leased Line (TCP/IP)

Internet Firewall

© Crown copyright 2004

Page 2



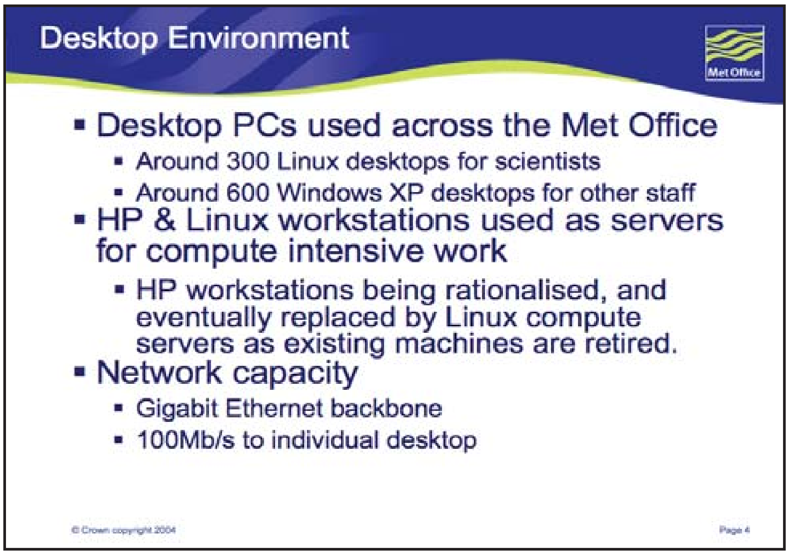
NEC SX-6 & SX-8 Supercomputers

Met Office

- First phase accepted end of Jan 2004
  - 30 node NEC SX-6 system split between 2 halls
  - 26 TB of disk space forming a GFS across 2 halls
  - Accessed via 4 NEC TX-7 front ends
  - Provides 6x performance capability of T3Es
  - Very reliable
- Second phase accepted April 2005
  - Added 16 Node SX-8
  - Added 4 More SX-6 Nodes
  - 2 Additional NEC TX-7s added
  - Increased disk space to 39 TB
- Operational forecast Switched to SX8 13th April 2005

© Crown copyright 2004

Page 3



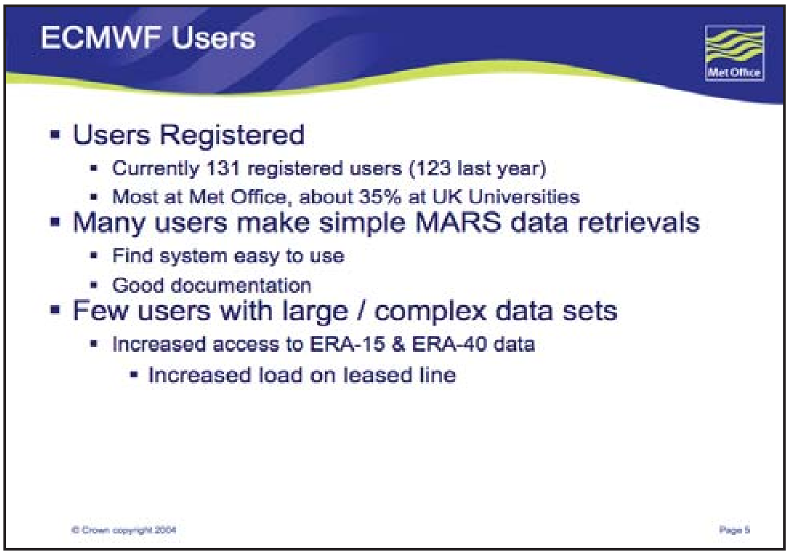
Desktop Environment

Met Office

- Desktop PCs used across the Met Office
  - Around 300 Linux desktops for scientists
  - Around 600 Windows XP desktops for other staff
- HP & Linux workstations used as servers for compute intensive work
  - HP workstations being rationalised, and eventually replaced by Linux compute servers as existing machines are retired.
- Network capacity
  - Gigabit Ethernet backbone
  - 100Mb/s to individual desktop

© Crown copyright 2004

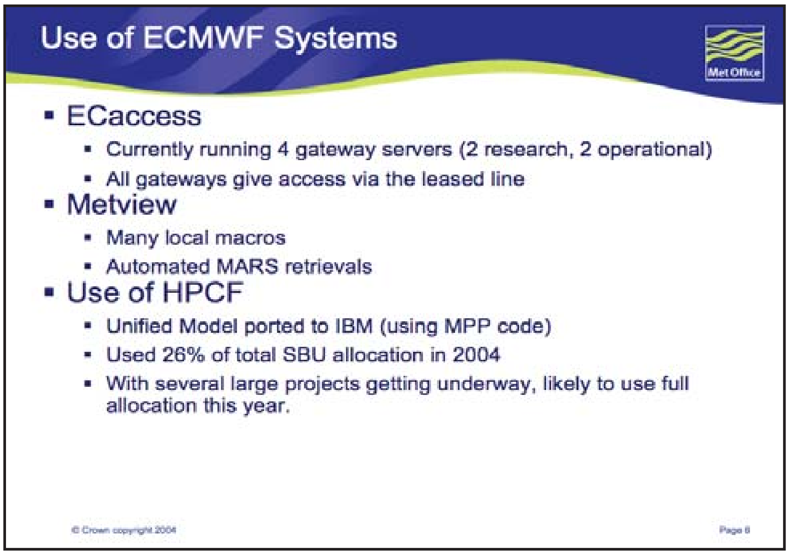
Page 4



ECMWF Users

- Users Registered
  - Currently 131 registered users (123 last year)
  - Most at Met Office, about 35% at UK Universities
- Many users make simple MARS data retrievals
  - Find system easy to use
  - Good documentation
- Few users with large / complex data sets
  - Increased access to ERA-15 & ERA-40 data
    - Increased load on leased line

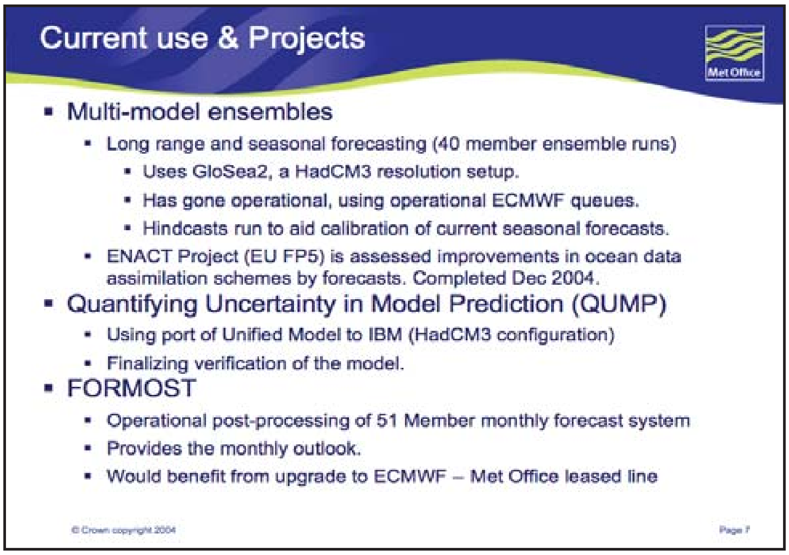
© Crown copyright 2004 Page 5



Use of ECMWF Systems

- ECaccess
  - Currently running 4 gateway servers (2 research, 2 operational)
  - All gateways give access via the leased line
- Metview
  - Many local macros
  - Automated MARS retrievals
- Use of HPCF
  - Unified Model ported to IBM (using MPP code)
  - Used 26% of total SBU allocation in 2004
  - With several large projects getting underway, likely to use full allocation this year.

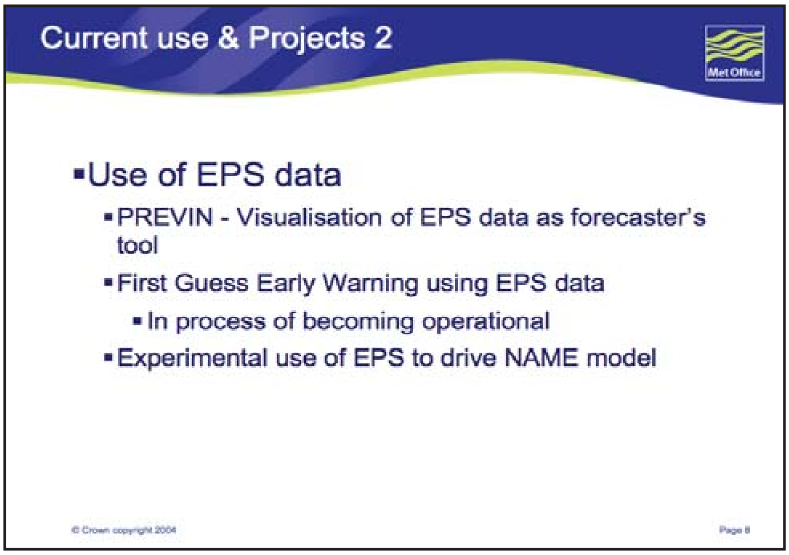
© Crown copyright 2004 Page 6



Current use & Projects

- Multi-model ensembles
  - Long range and seasonal forecasting (40 member ensemble runs)
    - Uses GloSea2, a HadCM3 resolution setup.
    - Has gone operational, using operational ECMWF queues.
    - Hindcasts run to aid calibration of current seasonal forecasts.
  - ENACT Project (EU FP5) is assessed improvements in ocean data assimilation schemes by forecasts. Completed Dec 2004.
- Quantifying Uncertainty in Model Prediction (QUMP)
  - Using port of Unified Model to IBM (HadCM3 configuration)
  - Finalizing verification of the model.
- FORMOST
  - Operational post-processing of 51 Member monthly forecast system
  - Provides the monthly outlook.
  - Would benefit from upgrade to ECMWF – Met Office leased line

© Crown copyright 2004 Page 7



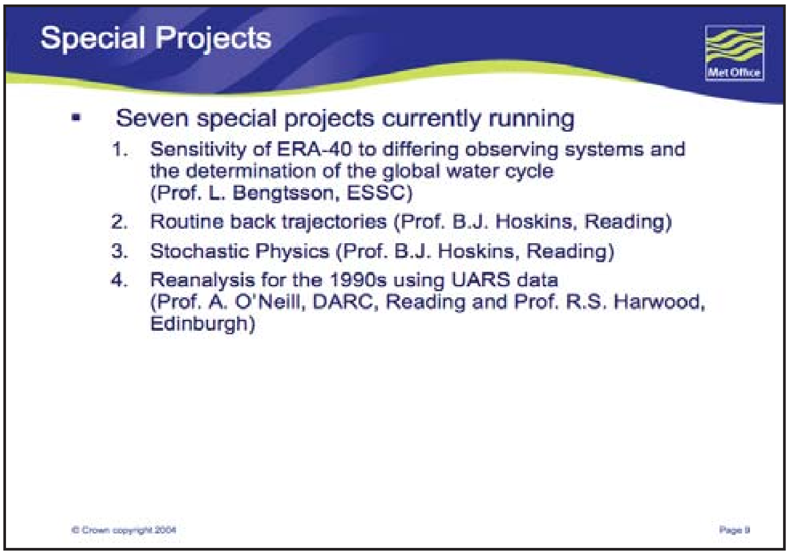
Current use & Projects 2

Met Office

- Use of EPS data
  - PREVIN - Visualisation of EPS data as forecaster's tool
  - First Guess Early Warning using EPS data
    - In process of becoming operational
  - Experimental use of EPS to drive NAME model

© Crown copyright 2004

Page 8



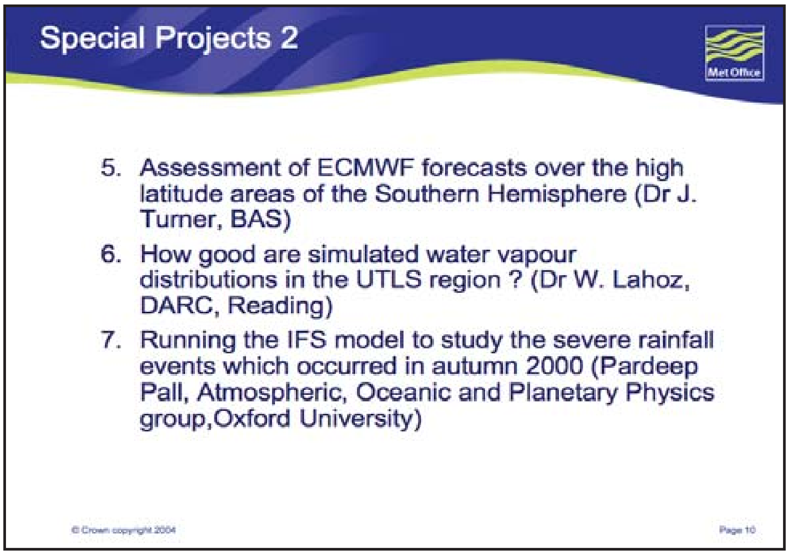
Special Projects

Met Office

- Seven special projects currently running
  1. Sensitivity of ERA-40 to differing observing systems and the determination of the global water cycle (Prof. L. Bengtsson, ESSC)
  2. Routine back trajectories (Prof. B.J. Hoskins, Reading)
  3. Stochastic Physics (Prof. B.J. Hoskins, Reading)
  4. Reanalysis for the 1990s using UARS data (Prof. A. O'Neill, DARC, Reading and Prof. R.S. Harwood, Edinburgh)

© Crown copyright 2004

Page 9



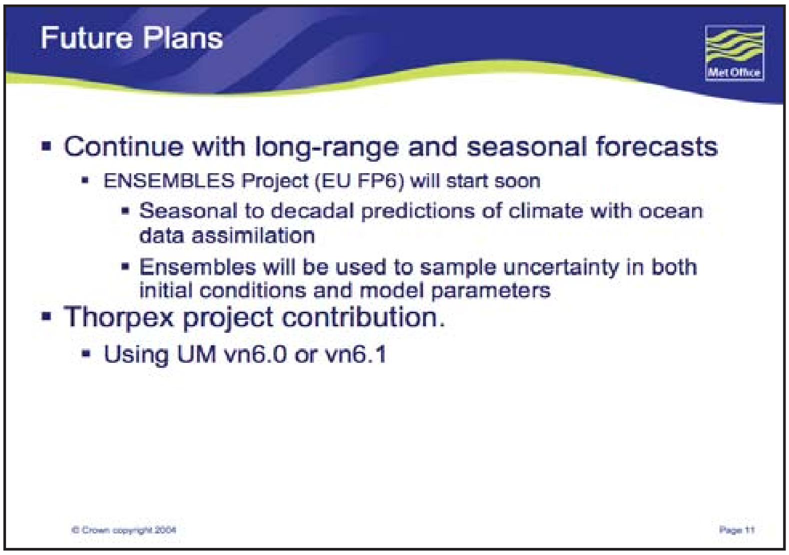
Special Projects 2

Met Office

5. Assessment of ECMWF forecasts over the high latitude areas of the Southern Hemisphere (Dr J. Turner, BAS)
6. How good are simulated water vapour distributions in the UTLS region ? (Dr W. Lahoz, DARC, Reading)
7. Running the IFS model to study the severe rainfall events which occurred in autumn 2000 (Pardeep Pall, Atmospheric, Oceanic and Planetary Physics group,Oxford University)

© Crown copyright 2004

Page 10



Future Plans

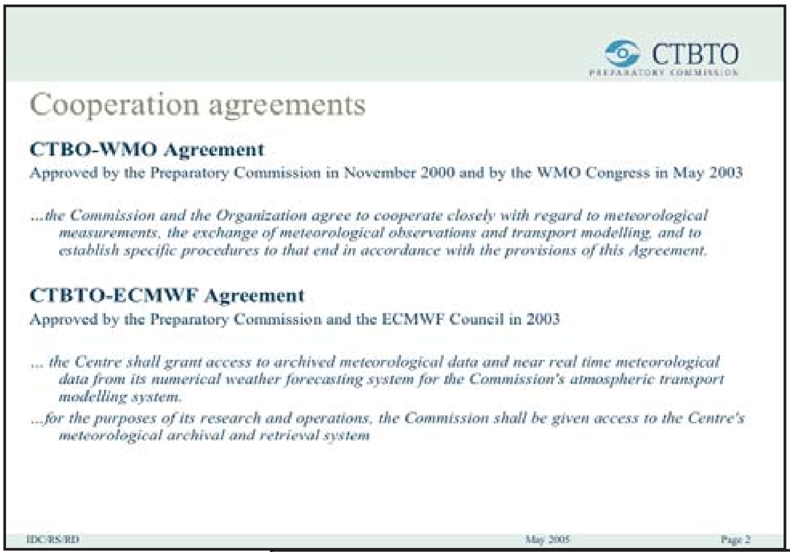
- Continue with long-range and seasonal forecasts
  - ENSEMBLES Project (EU FP6) will start soon
    - Seasonal to decadal predictions of climate with ocean data assimilation
    - Ensembles will be used to sample uncertainty in both initial conditions and model parameters
- Thorpex project contribution.
  - Using UM vn6.0 or vn6.1

© Crown copyright 2004

Page 11

M Pithon asked what the four NEC TX -7 front ends were used for. R. Sharp replied that they were mainly used as file servers and for interactive job submission, although they can also be used to run anything which is not suitable for the SX supercomputers.

Gerhard Wotawa – Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization



CTBTO PREPARATORY COMMISSION

Cooperation agreements

CTBO-WMO Agreement

Approved by the Preparatory Commission in November 2000 and by the WMO Congress in May 2003

...the Commission and the Organization agree to cooperate closely with regard to meteorological measurements, the exchange of meteorological observations and transport modelling, and to establish specific procedures to that end in accordance with the provisions of this Agreement.

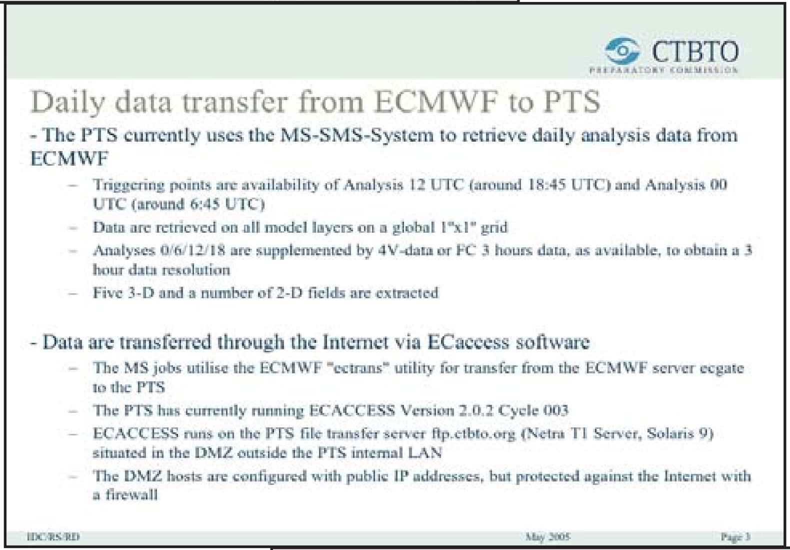
CTBTO-ECMWF Agreement

Approved by the Preparatory Commission and the ECMWF Council in 2003

... the Centre shall grant access to archived meteorological data and near real time meteorological data from its numerical weather forecasting system for the Commission's atmospheric transport modelling system.

...for the purposes of its research and operations, the Commission shall be given access to the Centre's meteorological archival and retrieval system

IDC/RS/RD May 2005 Page 2

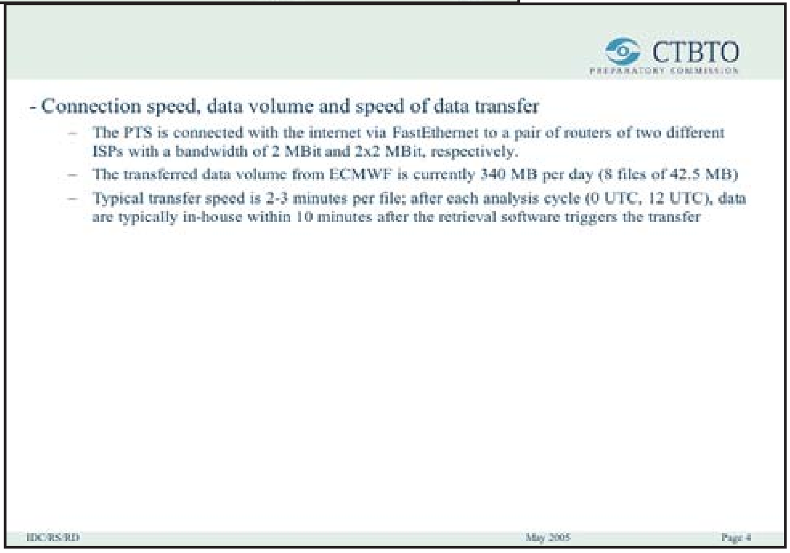


CTBTO PREPARATORY COMMISSION

Daily data transfer from ECMWF to PTS

- The PTS currently uses the MS-SMS-System to retrieve daily analysis data from ECMWF
  - Triggering points are availability of Analysis 12 UTC (around 18:45 UTC) and Analysis 00 UTC (around 6:45 UTC)
  - Data are retrieved on all model layers on a global 1°x1° grid
  - Analyses 0/6/12/18 are supplemented by 4V-data or FC 3 hours data, as available, to obtain a 3 hour data resolution
  - Five 3-D and a number of 2-D fields are extracted
- Data are transferred through the Internet via ECaccess software
  - The MS jobs utilise the ECMWF "ectrans" utility for transfer from the ECMWF server ecgate to the PTS
  - The PTS has currently running ECACCESS Version 2.0.2 Cycle 003
  - ECACCESS runs on the PTS file transfer server ftp.ctbto.org (Netra T1 Server, Solaris 9) situated in the DMZ outside the PTS internal LAN
  - The DMZ hosts are configured with public IP addresses, but protected against the Internet with a firewall

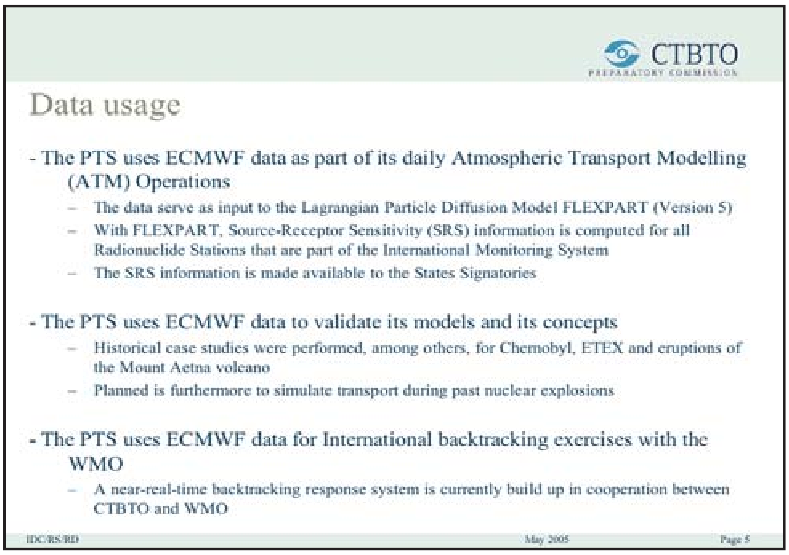
IDC/RS/RD May 2005 Page 3



CTBTO PREPARATORY COMMISSION

- Connection speed, data volume and speed of data transfer
  - The PTS is connected with the internet via FastEthernet to a pair of routers of two different ISPs with a bandwidth of 2 MBit and 2x2 MBit, respectively.
  - The transferred data volume from ECMWF is currently 340 MB per day (8 files of 42.5 MB)
  - Typical transfer speed is 2-3 minutes per file; after each analysis cycle (0 UTC, 12 UTC), data are typically in-house within 10 minutes after the retrieval software triggers the transfer

IDC/RS/RD May 2005 Page 4

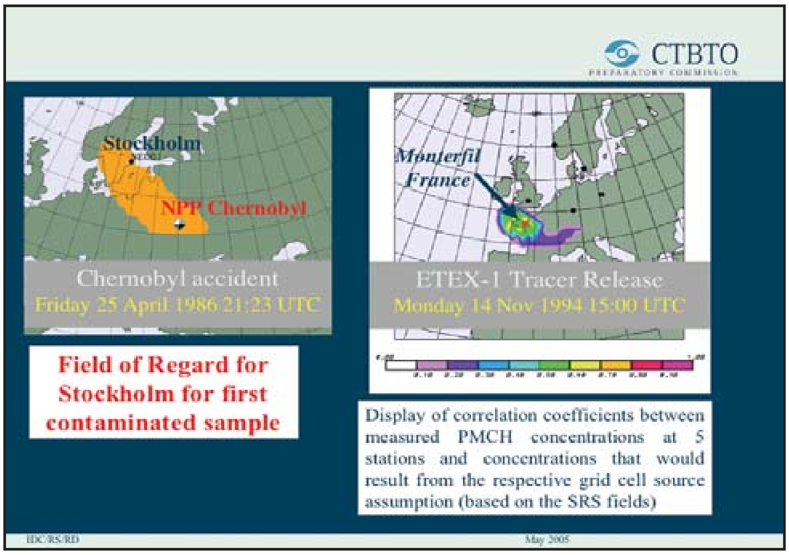


CTBTO PREPARATORY COMMISSION

Data usage

- The PTS uses ECMWF data as part of its daily Atmospheric Transport Modelling (ATM) Operations
  - The data serve as input to the Lagrangian Particle Diffusion Model FLEXPART (Version 5)
  - With FLEXPART, Source-Receptor Sensitivity (SRS) information is computed for all Radionuclide Stations that are part of the International Monitoring System
  - The SRS information is made available to the States Signatories
- The PTS uses ECMWF data to validate its models and its concepts
  - Historical case studies were performed, among others, for Chernobyl, ETEX and eruptions of the Mount Aetna volcano
  - Planned is furthermore to simulate transport during past nuclear explosions
- The PTS uses ECMWF data for International backtracking exercises with the WMO
  - A near-real-time backtracking response system is currently build up in cooperation between CTBTO and WMO

IDC/RS/RD May 2005 Page 5



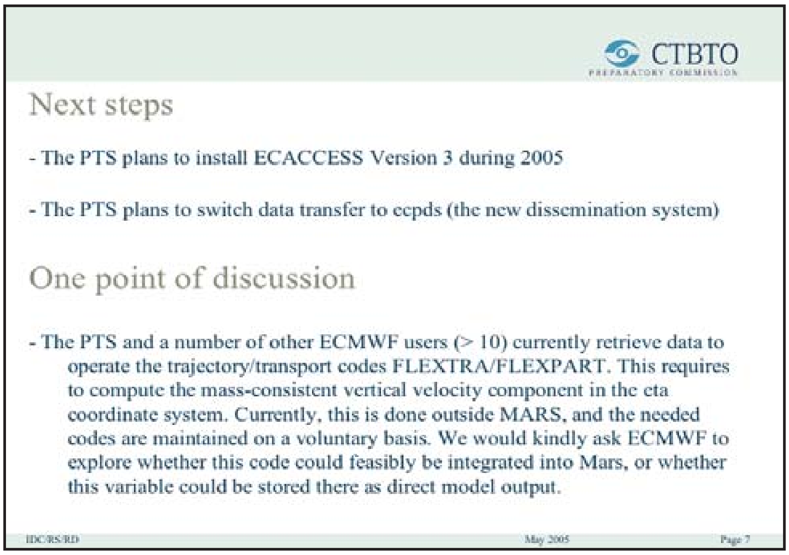
Chernobyl accident
Friday 25 April 1986 21:23 UTC

ETEX-1 Tracer Release
Monday 14 Nov 1994 15:00 UTC

Field of Regard for Stockholm for first contaminated sample

Display of correlation coefficients between measured PMCH concentrations at 5 stations and concentrations that would result from the respective grid cell source assumption (based on the SRS fields)

IDC/RS/RD May 2005



CTBTO PREPARATORY COMMISSION

Next steps

- The PTS plans to install ECACCESS Version 3 during 2005
- The PTS plans to switch data transfer to ecpds (the new dissemination system)

One point of discussion

- The PTS and a number of other ECMWF users (> 10) currently retrieve data to operate the trajectory/transport codes FLEXTRA/FLEXPART. This requires to compute the mass-consistent vertical velocity component in the eta coordinate system. Currently, this is done outside MARS, and the needed codes are maintained on a voluntary basis. We would kindly ask ECMWF to explore whether this code could feasibly be integrated into Mars, or whether this variable could be stored there as direct model output.

IDC/RS/RD May 2005 Page 7

ANNEX 1

ANNEX 1

Seventeenth Meeting of Computing Representatives

ECMWF, Shinfield Park, Reading, U.K., 19–20 May 2005

Participants

Austria	Cornelia Hammerschmid
Belgium	Liliane Frappez
Croatia	Vladimir Malovic
CTBTO	Gerhard Wotawa
Czech Republic	Karel Ostatnicky
Denmark	Niels Olsen
Finland	Kari Niemelä
France	Marion Pithon
Germany	Elisabeth Krenzien
Greece	loannis Alexiou
Hungary	László Tölgyesi
Iceland	Halldor Björnsson
Ireland	Paul Halton
Netherlands	Gert-Jan Marseille
Norway	Rebecca Rudsar
Romania	Catalin Ostroveanu
Serbia & Montenegro	Vladimir Dimitrijevic
Slovenia	Petar Hitij
Spain	Eduardo Monreal
	Julio González Breña
Sweden	Rafael Urrutia
Switzerland	Peter Roth
Turkey	Ahmet Erturk
United Kingdom	Roddy Sharp
ECMWF:	Sylvia Baylis
	Petra Berendsen
	Jens Daabeck
	Paul Dando
	Matteo Dell'Acqua
	Françis Dequenne
	Richard Fisker
	Helene Garçon
	Laurent Gougeon
	John Greenaway
	Fredi Hofstadler
	Petra Kogel
	Dominique Lucas
	Carsten Maass
	Umberto Modigliani
	Pam Prior
	Sylvia Rozemeijer
	Neil Storer
	Isabella Weger

ANNEX 2

ANNEX 2

Programme

Thursday 19 May 2005

09.30	Coffee	
10.00	Welcome	
	ECMWF’s computer status and plans	I. Weger
	Member States and Co-operating States presentations	
12.30	Lunch	
13.30	Visit of Computer Hall (optional)	
14.00	Member States and Co-operating States presentations (continued)	
	HPCF and DHS update	N. Storer
	SIMDAT and DEISA projects	M. Dell’Acqua
	Introduction to ECPDS	L. Gougeon
16.00	Coffee	
16.30	Planned model resolution upgrade in operations	A. Hofstadler
	Graphics update	J. Daabeck
	The ECMWF Linux cluster: one year on	P. Kogel
	ECMWF Disaster recovery plans	F. Dequenne
18.00	Cocktails	
20:00	Informal dinner at restaurant	

Friday, 20 May 2005

09.00	Member States and Co-operating States presentations (continued)	
10.30	Coffee	
11:00	User Registration: update on the interface	P. Dando
	Results of the survey of external users	User Support
12.30	Discussion	
13.00	End of meeting	