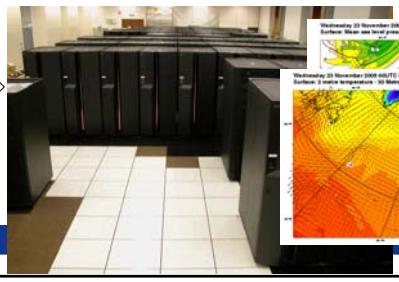
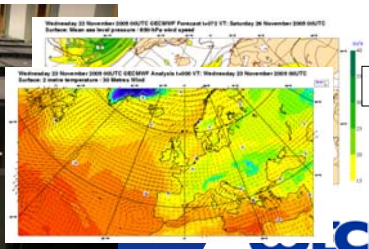
SIMDAT Project
Meteorology Activity
Elements for building the WIS

Guillaume Aubert – Baudouin Raoult, ECMWF

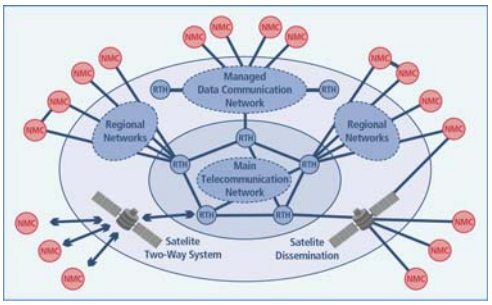
Real-time data flow: From Observation to Numerical Weather Prediction to Decision Making





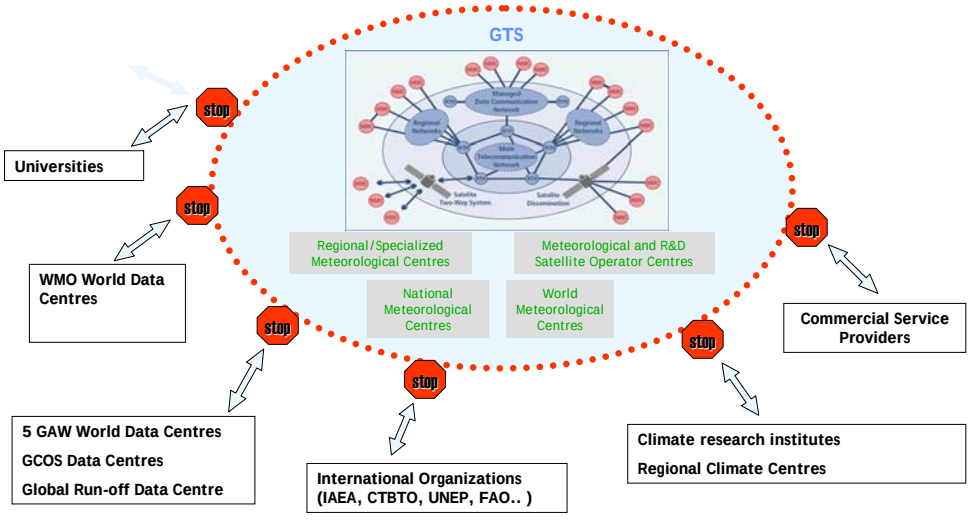
Current Situation: GTS

- **Global Telecommunications System (GTS)**
  - Observations
  - Forecasts
  - Warnings

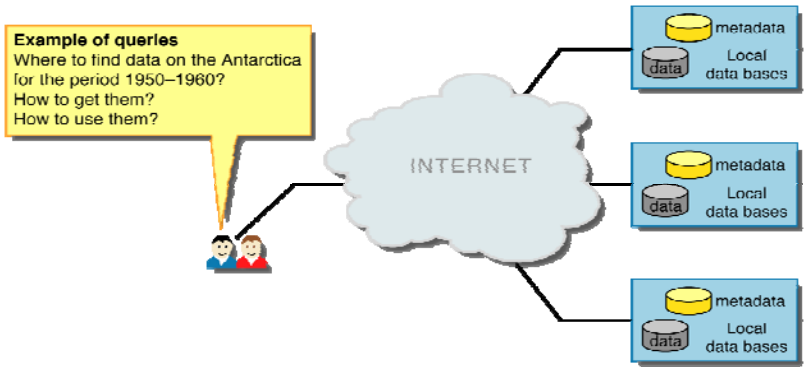


- **Private Network**
  - Node to Node network
  - NOT an Internet

Current Situation: GTS



Current Situation (cont.)



- **At present, WMO Programmes do not offer appropriate response to such queries**

WMO Information System - Vision

- **Real-time collection and dissemination:**
  - Real-time "push" through dedicated telecommunication for operation-critical data
- **Timely delivery of data and products:**
  - Delayed mode "push" through dedicated telecommunication means and public data networks, especially the Internet
- **Data discovery and retrieval service:**
  - "Pull" through the Internet (HTTP, FTP,…)
- **Unified procedures**
  - More efficient data exchange
- **Coordinated and standardized metadata**
  - Data interoperability between programmes
  - Improved data management
  - ISO 191xxx series

Structure of WIS

- **Functional centres interconnected by data communication networks**
- **National Centres (NC)**
  - **Links national data providers and users to regional and global data exchange nodes**
- **Data Collection and Production Centres (DCPC)**
  - **Provides for regional and mission-oriented exchange of data and products**
  - **Supports information push and pull**
- **Global Information System Centres (GISC)**
  - **Provides for 24/7 reliable global exchange services**
  - **Collects and provides metadata**
  - **Supports data and information discovery and retrieval**

SIMDAT Project

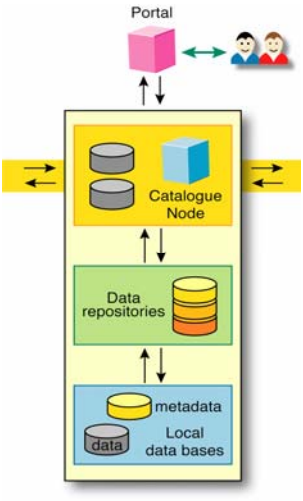
- **EU funded project**
  - **4 Activities: Aero, Pharma, Automotive and Meteo**
- **Meteorology activity of SIMDAT:**
  - **5 Partners: DWD, Meteo-France, UKMO, EUMETSAT and ECMWF**
  - **European effort for developing the WIS**
- **Coordination and main developments: ECMWF**
- **Distribution of the SIMDAT software**
  - **Contract with the EU, already specifies that the software will be made available to the entire WMO community**
  - **Wider distribution, looking at an Open Source licence**
- **An evaluation copy can be requested from ECMWF**

Meteorology Activity: Project Aims

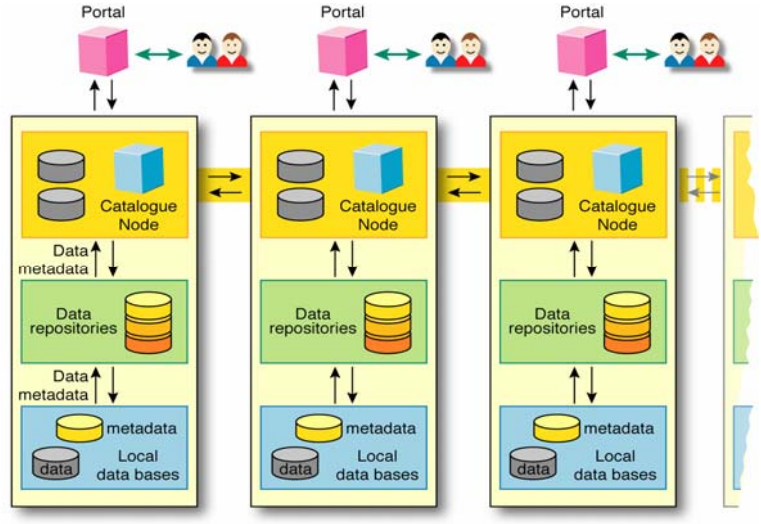
- **To build an integrated and scalable framework for the collection and sharing of distributed data**
  - **Targeting meteorology, hydrology, climate and other environmental data**
- **To provide a unified view of all available data**
- **To provide a transparent access to distributed resources**
  - **Discovery service, Cataloguing service, Subscription service,…**
- **To implement a non-intrusive system**
  - **Provide access to existing local databases**
  - **Provide a global access control policy managed by the partners and integrated into their existing security infrastructure**

Prototype: Architecture

- **Based on three components *(plus one)*:**
  - **Portal**
  - **Catalogue nodes**
  - **Data repositories**
  - ***(Local database)***



Prototype: Connectivity

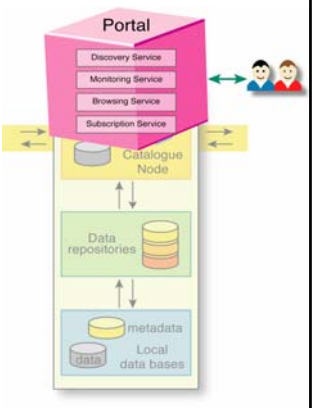


SIMDAT project

ECMWF

Components: Portal

- **Web based user Interface of the system**
- **Offers discovery facilities**
  - **Catalogue Browsing**
  - **Searching (keyword, time and space)**
- **Allows data selection and request submission**
- **Offers per user request management**
  - **Progress monitoring, …**
- **Offers data download facility**



SIMDAT project

ECMWF

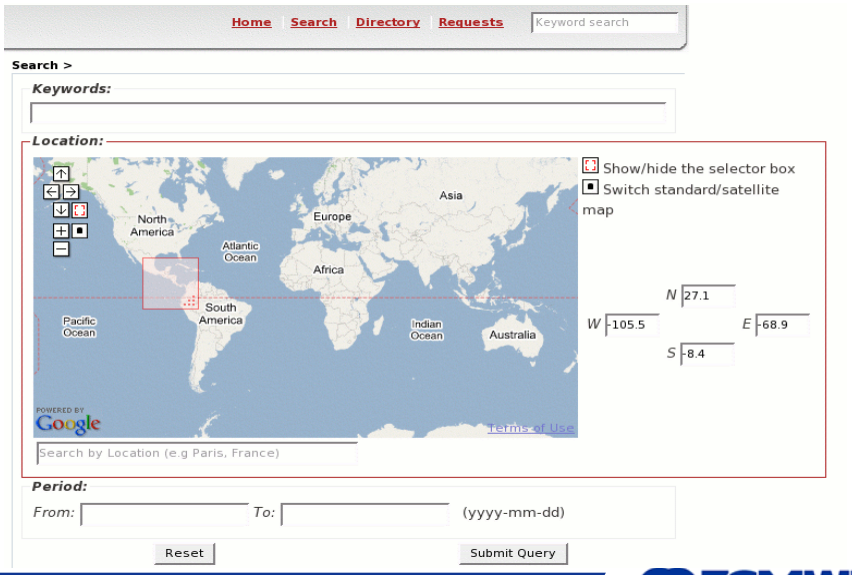
Focus on usability

- **Usability from an end-user point of view**
  - **Portal based on concepts used by major search engines: no learning necessary**
- **Usability from a data publisher point of view**
  - **Effort to minimise prerequisite technological knowledge (Metadata, XML, …)**

Finding data

- **Google like search**
  - **Fast: people are now used get results in a few milliseconds**
  - **Scoring: people are used to find the most relevant hits in the first result page**
- **Yahoo like directory**
  - **Datasets can be categorized**
  - **A dataset can be in several categories at once**

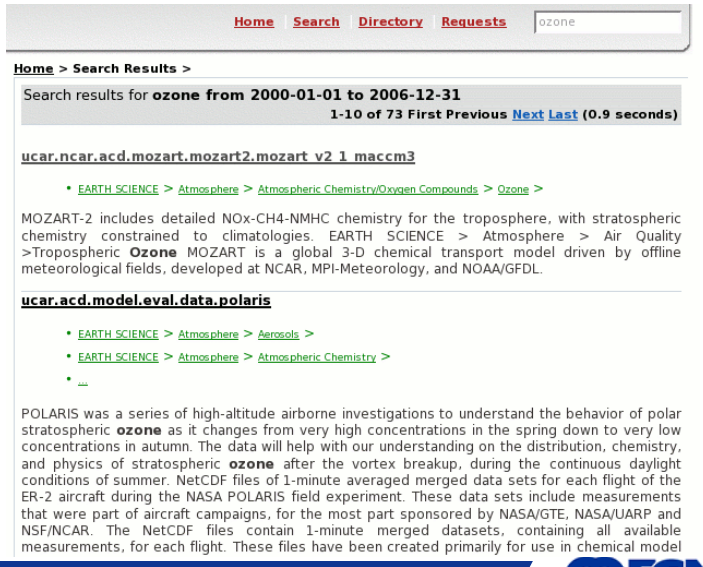
Portal: Searching...



SIMDAT project

ECMWF

Portal: Search results



SIMDAT project

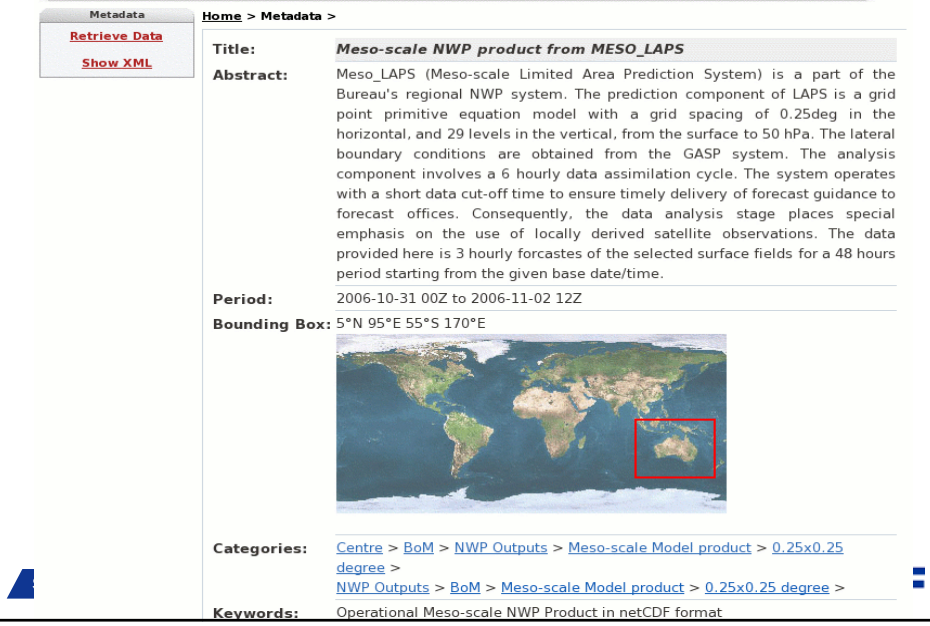
ECMWF

Portal: Browsing by Category

SIMDAT project

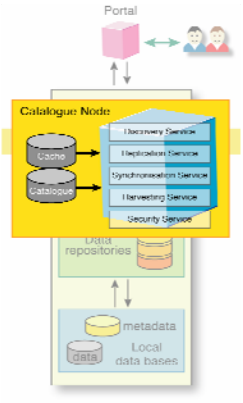
ECMWF

Portal: Legible Metadata



Components: Catalogue Node

- **Provides connectivity between Partners**
  - Can reach any other Catalogue Node
- **Holds the metadata catalogue**
  - Provide discovery services to the Portal
- **Implement peer-to-peer synchronization of metadata with other Catalogue Nodes**
- **Forward data requests**
  - To its Data Repositories
  - To peer Catalogue Nodes
- **Stream retrieved data between Data Repository and End User**

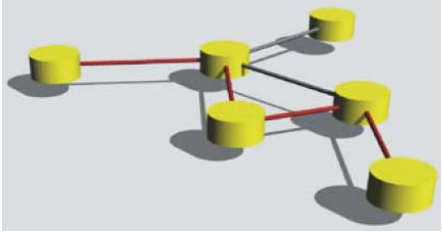


Peer-to-peer synchronization

- **Each Catalogue Node synchronizes its catalogue with its neighbours in the Meshed Network**
- **Based on a journaling mechanism**
  - Create, update, delete records
- **Transactional access to the journal and metadata catalogue**
  - Maintains catalogue consistency
- **Clean separation between the metadata management and the synchronization**
- **Demo ...**

Mesh Network: Peer-to-peer routing

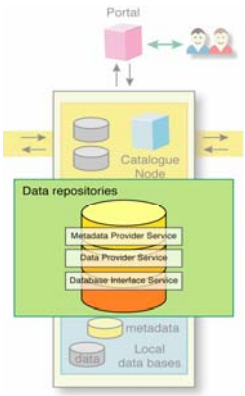
- **Year I Prototype:**
  - **All the nodes had to be fully interconnected to get a complete catalogue (N x N interconnections)**
  - **It is not scalable, not manageable as the number of nodes is growing**
- **Year II Prototype:**
  - **Implements a Mesh Network**
  - **Each node is connected to 2-3 peers at most**
  - **Each node can reach any nodes which is part of the network**



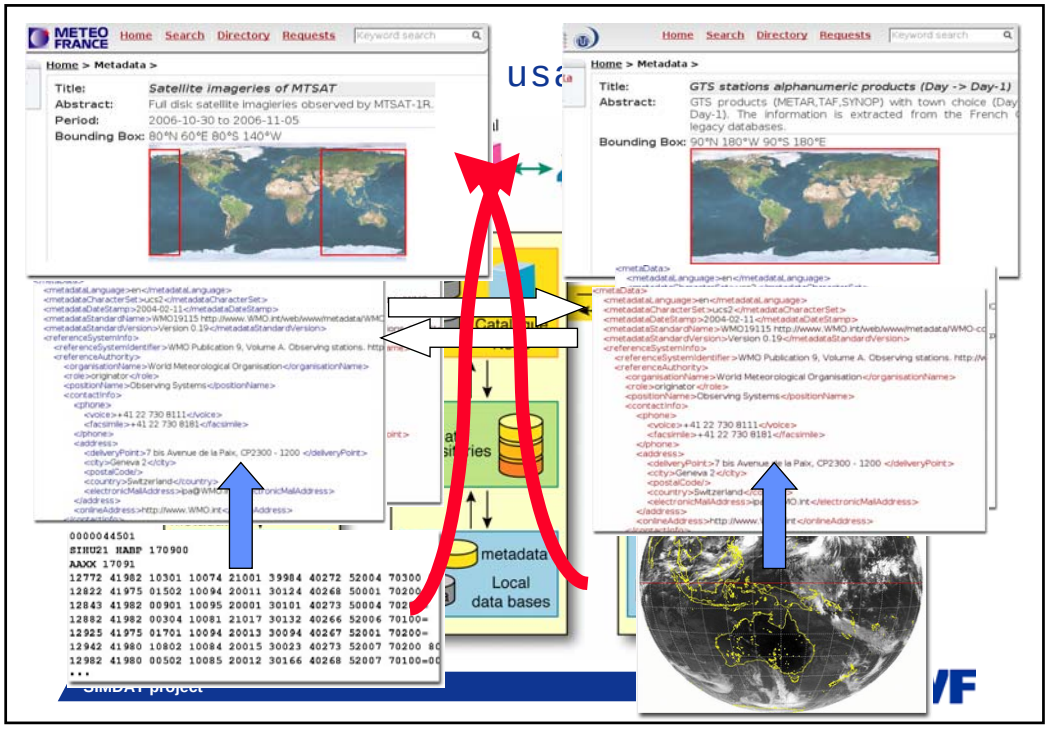
ECMWF

Components: Data Repository

- **Provides a unified interface between the Catalogue Node and local databases**
- **Acts as a metadata provider**
  - **Publish metadata for harvesting by the Catalogue Node**
- **Acts as a data provider**
  - **Accept data requests from the Catalogue Node**
  - **Translate data requests into request for local databases**
  - **Implement asynchronous handling of data requests**
    - **Support for access to off-line data**



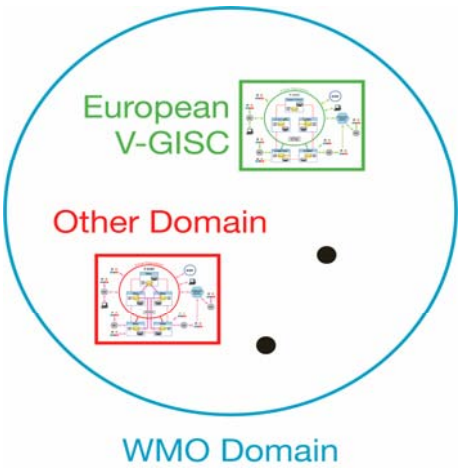
ECMWF



Virtual Organisation

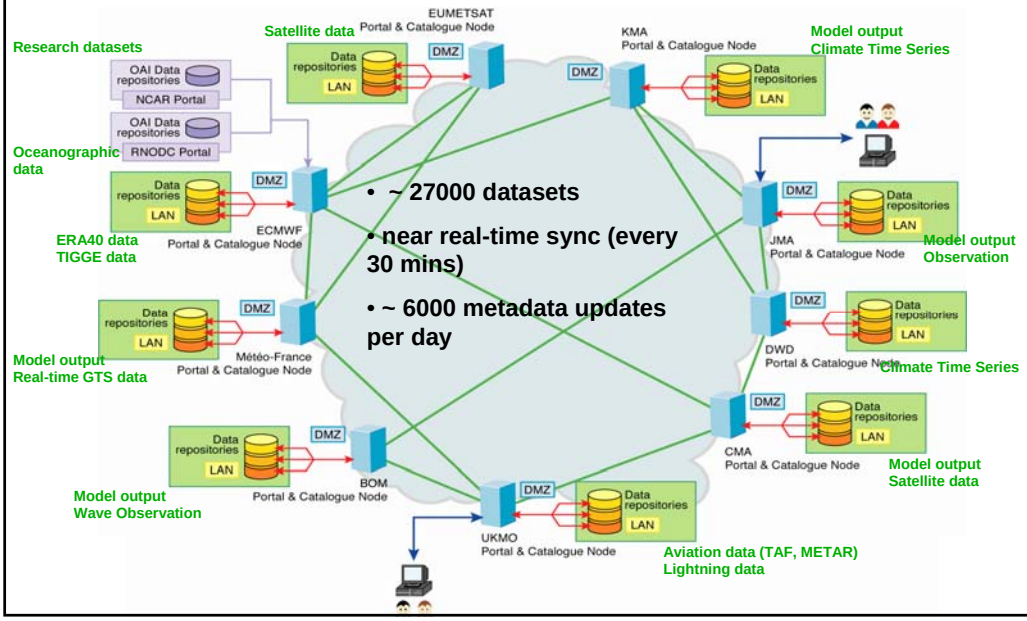
- **We want security**
  - Users needs to be authenticated (who are they?)
- **We want to enforce data policies**
  - Can this user access this data?
  - Users have “roles”, e.g. this user is a “researcher”
  - Data have “policies”, e.g. this data is accessible by “researchers”.
- **These issues have to be solved in a distributed environment**
  - Non centralised solution: user may login at one site (authenticated)…
  - … to get data from another (authorized)
- **Problems**
  - How to we make sure that all sites have the same understanding of roles and policies (e.g. what is a “researcher”)
  - How do we solve technically the fact that authentication and authorisation do not take place at the site.

Virtual Organisations: Domains & Trust



- A domain represents a group of organisations with a common data access policy.
- Organisations within a domain trust each other to authenticate users
- Authorization is performed by the Site hosting the data

Status: 11 sites connected



Data available in demonstrator

BoM	NWP outputs, Observations	GRIB, ASCII
CMA	NWP output, satellite data	GRIB, BUFR
DWD	Climate Time Series, NWP outputs	XML, HTML, GRIB
ECMWF	NWP outputs (ERA15, ERA40, TIGGE), Sample GTS observations (1 day)	GRIB1, GRIB2, BUFR, PDF
EUMETSAT	METEOSAT 8 images	JPEG
JMA	NWP outputs, MTSAT images, Observations	GRIB, BUFR, animated GIFs
KMA	Climate Time Series, NWP outputs	XML, GRIB
Météo France	NWP outputs Real-time GTS data	GRIB, ASCII
NCAR	~ 6000 datasets	NetCDF, other
RNODC	Oceanographic data (BATHY, SHIP, TESAC)	ASCII
UKMO	Aviation data (METAR, TAF)	XML

Support of WIS requirements

Routine collection of observation data	Work in progress.
Automated dissemination of all observed data and products, both real-time and non real-time : Push model	Work in progress.
Ad-hoc requests for data and products : Pull model	Done.
Support of different user profiles and data policies	Work in progress.
Support of data and network security	Done. Based on SSL
Support the integration of diverse datasets	Done. Support for Any data type, from any data repository.
Reliable infrastructure	Done.
Technologically sustainable and appropriate to local expertise	Done. Based on Standards. Run on a PC. Open Source components.
Modular, flexible and scalable	Done.

Plans

- **Security [Q3 2007]**
  - **Tools for user and policy administration [Q2 2007]**
  - **Data access policy [Q3 2007]**
- **Routine Collection of observations [Q2 2008]**
  - **Push mechanism for critical data [Q2 2007]**
  - **Data Replication/Caching [Q2 2008]**
- **Subscription service [Q1 2008]**
- **Monitoring Tools [Q4 2007]**
- **Data delivery of very large datasets [Q3 2008]**

Conclusion

- **Virtual distributed database**
- **Flexible, non intrusive architecture**
- **Peer-to-peer synchronisation mechanisms implemented**
- **Mesh Network communication implemented**
- **Generates interest in meteorology and other environmental communities**
- **Software running continuously, catalogues are constantly updated**
- **Data available from 11 sites worldwide**

Thank you !

SIMDAT project

ECMWF