SIMDAT
Elements for building the WIS

Guillaume Aubert
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
  - Point to point network
  - NOT an Internet
Current Situation: GTS

GTS

- 5 GAW World Data Centres
- GCOS Data Centres
- Global Run-off Data Centre
- WMO World Data Centres
- Universities
- Regional/Specialized Meteorological Centres
- Climate research institutes
- Regional Climate Centres
- Meteorological and R&D Satellite Operator Centres
- International Organizations (IAEA, CTBTO, UNEP, FAO...)
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
The Fourteenth WMO Congress (2003) approved the concept of the WMO Information System (WIS)

- “a single, co-ordinated global infrastructure for the collection and sharing of information in support of all WMO and related international programmes”

The WIS defines three functional components:

- National Centres (NC)
- Data Collection or Production Centres (DCPC)
- Global Information System Centres (GISC)

DWD, Météo France and the UK Met Office have volunteered to collaborate on development of a virtual GISC (VGISC)

ECMWF and EUMETSAT are included in the project as DCPCs.
SIMDAT and the VGISC Project

- SIMDAT is a four years EU funded project under FP6
  - It started in 2004

- SIMDAT was an opportunity to fund to development of the VGISC
  - ECMWF is the project coordinator
  - Most of the developments are done at ECMWF

- The available budget is approximately 1.1 M€.
  - ECMWF: 84 person/months (2 consultants)
  - DWD: 36 person/months
  - UK Met Office: 6 person/months
  - Météo France: 6 person/months
  - EUMETSAT participates as a non-funded partner.
SIMDAT 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
The VGISC architecture

Virtual Organisation

VGISC

MétéoFrance

data
meta data

UK MetOffice

data
meta data

DWD

data
meta data

EUMETSAT

data
meta data

ECMWF

data
meta data

External DCPC

GISC

NC

11th ECMWF Workshop - Nov 2007 - GA
Access to existing data repositories in a non-intrusive fashion
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
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
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
Connectivity to another site via secure connection layer
A scalable system: addition of new sites
Catalogue Synchronisation: each site has a copy of the global catalogue
Data from anywhere can be accessed from everywhere
Datasets are indexed (text, location, time)
Datasets are categorised (browsing)

- Atmosphere (18 categories)
- Biosphere (1 product, 3 categories)
- Climate Indicators (1 category)
- Cryosphere (1 category)
- Human Dimensions (2 categories)
- Hydrosphere (2 categories)
- Land Surface (6 categories)
- Marine meteorology (3 products)
- Oceanography (8 products)
- Oceans (11 categories)
- Paleoclimate (1 category)
- Solid Earth (1 category)
- Spectral/Engineering (3 categories)
- Sun-earth interactions (2 categories)
Datasets are described using Metadata

Title: MEGAN (Model of Emissions of Gases and Aerosols from Nature)

Abstract: MEGAN is a modeling system for estimating the net emission of gases and aerosols from terrestrial ecosystems into the atmosphere. It is driven by landcover, weather, and atmospheric chemical composition. MEGAN is a global model with a base resolution of ~1 km. It can run as a stand-alone model for generating emission inventories but is also being incorporated as an on-line component of chemistry/transport and earth system models. The MEGAN collection is currently being updated to version 2.0 so these data are not currently available. This notice will be removed when the update is complete. You may email cdp@ucar.edu if you have further questions.

Bounding Box: 84°N 180°E 57°S 0°W

Categories: EARTH SCIENCE > Atmosphere > Air Quality > Emissions >
EARTH SCIENCE > Atmosphere > Air Quality > Volatile Organic Compounds >
EARTH SCIENCE > Atmosphere > Air Quality > Nitrogen Oxides >
EARTH SCIENCE > Atmosphere > Atmospheric Chemistry > Trace Gases/Trace Species >
18 more...

On Line Source: https://cdp.ucar.edu/getCatalog.do?ID=ucar.ncar.acd.software.megan

11th ECMWF Workshop - Nov 2007 - GA
Multiple data types support

Stationname: Aachen

<table>
<thead>
<tr>
<th>Date</th>
<th>Value</th>
<th>Quality Level</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-09-17</td>
<td>19.2</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>2006-09-18</td>
<td>17.5</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>2006-09-19</td>
<td>19</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>2006-09-20</td>
<td>21.5</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>2006-09-21</td>
<td>27.2</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>2006-09-22</td>
<td>27.3</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>2006-09-23</td>
<td>25.3</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>2006-09-24</td>
<td>23.8</td>
<td>high</td>
<td>50.7839</td>
<td>6.095</td>
<td>202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Status: 11 sites connected

- ~ 27000 datasets
- near real-time sync (every 30 mins)
- ~ 6000 metadata updates per day
Status and Plans

● Status
  - The prototype has been deployed in a large worldwide test bed (11 sites)
  - Provides discovery capabilities
  - Allows data retrieval
  - Support a wide variety of data types
  - Implements user management and data policies
  - The software is available under an Open Source Licence

● Plans
  - Ingestion of GTS data
  - Offer subscription services
  - Follow the development of the INSPIRE directive
New Release with support for VO

- We want security
  - Users need 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 do 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.
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.
Project successes

- With the EU
  - The achievements of the meteorological activity of SIMDAT have been acknowledged by EU reviewers

- With the WMO community
  - The software was presented during expert team meetings

- Evaluation copies of the software
  - Australia, Brazil, Canada, China, Finland, Japan, Korea, Morocco, Norway, Portugal, Russia, Sweden, Zimbabwe

- Other communities are interested
  - Global Monitoring for Environment and Security (GMES)
  - EUMETNET OPERA project
  - International Polar Year (IPY)
Conclusion

- SIMDAT is a virtual distributed database
  - Fully decentralised and all sites have equal rights
  - Decentralised user and data policies management
  - Integrated catalogue and data retrieval functions (one-stop-shop)
  - Designed for operational use

- Interfacing with any existing data repositories
  - Without any impact on the local infrastructure or disruption of operational activities
  - Support for any data types (GTS bulletins, Model outputs, Satellite images, climate time-series, …)

- Generates interest in meteorology and other environmental communities
SIMDAT General Overview

SIMDAT is Data grids for process and product development using numerical simulation and knowledge discovery. The project is funded by the European Commission under the Information Society Technologies Programme (IST-2004-51438).

SIMDAT focuses on four application areas: product development in automotive, aerospace and pharma industries as well as a service provision in meteorology.

This WIKI provides information about the results and software developed during the lifetime of the the SIMDAT project for the Meteorological application. General information about the SIMDAT project and the other application areas are found in: http://www.scail.fraunhofer.de/simdat.html

SIMDAT Virtual Global Information System Centre

The objective of SIMDAT for the meteorology application is to develop a VGISC, a virtual and consistent view of all meteorological data distributed in the real-time and archived databases of the partners. The system will provide a secure, reliable and efficient mechanism to collect, exchange and share these distributed data, in order to support research and operational activities of the meteorological community. New ECMWF, EUMETSAT, Météo-France and the Met Office will use SIMDAT to select and share a distributed set of meteorological data.

The use of Grid technologies and standards for metadata, interoperability and distributed data, and the need for Grid technologies and distributed data, and a virtual and consistent view of all meteorological data distributed in the real-time and archived databases of the partners.

Infrastructure, the SIMDAT stack is deployed in the prototype. (Including Atmosphere, Ocean and Land model)

http://code.ecmwf.int/trac/vmc

27,000 datasets discoverable

543x116
Slide 30

http://code.ecmwf.int/trac/vmc
Thank you