End–to–end data management prototype

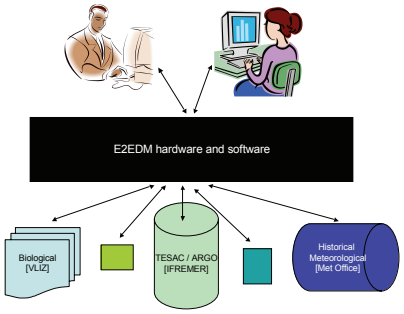
JCOMM Expert Team on Data Management Practices

Overview

The Joint Commission (WMO/IOC) for Oceanography and Marine Meteorology (JCOMM) co-ordinates a number of Expert Teams. At the first meeting of the Expert Team on Data Management Practices (ETDMP) in Gdynia, Poland, in July 2004 there were three prototype projects set up, one of which relates to end-to-end data management (E2EDM).

The end-to-end data management prototype is a proof of concept for a distributed database of marine data, accessed through a central web portal. The Chair of the ETDMP is Dr Nickolay Mikhailov of the National Oceanographic Data Centre of Russia (RIHMI-WDC), and he is also the project leader for the E2EDM prototype.

The Met Office participated throughout the project, from involvement with the conception, to the provision of historical marine meteorological data. This prototype system was demonstrated at JCOMM-II in Halifax, Canada, 19–27 September 2005.



Method

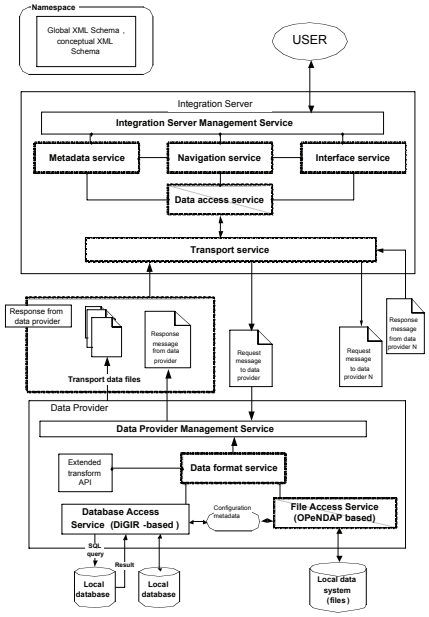
The Met Office contributed by extracting five years of historical marine meteorological data into flat files, and setting up a standalone PC connected to the internet (with firewall) with our data loaded. The Russian NODC provided the integration software to enable its web portal software to extract data from the data sets.

The data are in IMMT (International Maritime Meteorological Tape) format, which is the WMO recommended format for the exchange of delayed-mode marine meteorological data. The area covered by the prototype is the North Atlantic, and the data were requested in 25 separate data sets. These were separated into years (1997–2001) and 10 degree latitude bands (20°–70° North). The data files ranged in size from 6MB to 60MB, depending on the number of observations contained within them (50,000–500,000).

Prototype covers the North Atlantic (20°–70° N and 80° W–40° E)

Data provider	Data type	Elements
Met Office, UK	Historical Marine Meteorological	Air temperature Sea-surface temperature Mean sea-level pressure Wind speed Wind direction Wind-wave height Wind-wave period
USA NODC Russian NODC VLIZ, Belgium	Historical Ocean Cruise	Temperature Salinity Oxygen Major nutrients
MEDS, Canada	Delayed-mode GTSPP	Temperature Salinity
Russian NODC	Real-time Marine Meteorological (SHIP)	Air temperature Sea-surface temperature Mean sea-level pressure Wind speed Wind direction Wind-wave height Wind-wave period
Russian NODC	Real-time Ocean (BATHY and TESAC)	Temperature Salinity
IFREMER, France	Real-time Ocean (TESAC / ARGO)	Temperature Salinity
USA NODC	Monthly climatic fields (average and deviation at standard levels)	Temperature Salinity
Russian NODC	GTS Analysis / Forecast	Sea-surface temperature Wave data
	Ocean SST / colour imagery satellite	

Detailed description from project website



Vision statement

- Pilot should demonstrate real-time access to, and fusion of, data
  - at operational time scale
  - across multiple disciplines
  - preferably non-traditional variables
  - from multiple source formats
  - from multiple providers in different geographic regions
  - of utility to some user group
- The pilot should demonstrate the full range of processes including data discovery, access, and visualisation
- It should use pre-existing components where possible and be achievable with modest incremental effort

High-level functionality

The following functionality is envisaged:

- a user can enter the system, either via a web browser or a dedicated client, and request data of a single or multiple types, from a distributed set of sources, over a single (or possibly multiple) space-time region(s)
- appropriate data to the user's request will be automatically sourced from wherever it resides, and returned to the requesting machine (which may be the user's machine, or an intermediate portal providing value-added services)
- tools will exist (again either on a dedicated client, or on an intermediate portal) to fuse the aggregated data in real time to produce a newly created data product of value to the user.

Conceptual components required

The pilot E2EDM system requires the following components:

- data sources, with data of potential interest to the system, and the technological means for such data to be accessed;
- a master list of such sources, which could be generated as a virtual list by querying one or multiple sources, or reside as an independent entity;
- 'system search' metadata for each source, which describes at a high level, in a machine-readable structured way, at least the following:
  - data class — according to an agreed semantic model yet to be defined (e.g. satellite data, in situ oceanographic data, biological data ...)
  - parameter list (according to agreed semantic model)
  - overall space, time footprint (according to ISO metadata standard)
  - location of, and access protocol for remote requests to connect to the data
- for complex data providers, e.g. sources of data on multiple parameters with discontinuous distributions in time and/or space, more detailed search metadata describing the individual space-time footprints of every parameter (e.g. different biological species distributions);
- one or more 'request brokers' capable of querying first the search metadata, then the relevant data sources, to retrieve data relevant to the user's request. (Such a request broker could either be client software installed on the user's machine, or a dedicated portal to which the user connects via a standard web browser);
- one or more user interfaces permitting the user to formulate an appropriate request;
- one or more applications capable of generation of real-time data products from the data returned as a result of the distributed query;
- relevant software and hardware to connect the various components of the system;
- relevant data and metadata models to ensure that requests can be formulated by the request broker, and responded to, in a consistent manner.

Commentary: This list attempts to identify the components which will be required, but makes no final decision as to whether they may exist as real or distributed entities, or where they should reside. For example, the system search metadata described has an obvious overlap with the conventional thematic metadata directories (GCMD, EDMED, MEDI, others) and could conceptually reside there in 'distributed' form. Alternatively it could reside in a separate 'registry' more directly under control of the 'owners' of the distributed JCOMM system (one could even start with one model, and migrate to another over time). Similarly, the more detailed search metadata could reside in an intermediate registry or cache, or be generated on demand from the data sources in real time, or simply be ignored for the purpose of the pilot project.

Future

As yet (1 Oct 2005) the outcome of the demonstration at JCOMM-II is unknown to those not at the meeting.

Unfortunately, the technology used in the project is not aligned with the current Met Office IT policy, so irrespective of the outcome of the JCOMM-II meeting, the Met Office will be decommissioning its section of the prototype in December 2005. This is being done with the full agreement of the prototype project management team.

For further information, and to try out the system, please visit:

Prototype demo
http://data.meteo.ru:8080/iserv/

Project website
http://data.meteo.ru/e2edm/index.php?section=1

There are many people to thank for their participation in this prototype project, in particular Dr Nickolay Mikhailov, Chair, Russian NODC, and Sergey Belov, Technical Expert, Russian NODC

Also the Project Board:
- Alan Douglas, Project Sponsor, Met Office
- Geoff Smith, Project Executive, Met Office
- Patricia MacKenzie, Project Manager, Met Office
- Elanor Gowland, Senior User, Met Office
- Martin Ouldridge, Senior Supplier, Met Office
- Dave Truckle, Project Assurance, Met Office
- Jeremy Tandy, IT Architect, Met Office

and PC Desktop, Met Office

Met Office Saughton House Broomhouse Drive Edinburgh EH11 3XQ United Kingdom
Tel: +44 (0)131 528 7313 Fax: +44 (0)131 528 7345
Email: elanor.gowland@metoffice.gov.uk