DATA MODELLING FOR ENVIRONMENTAL DATA BASES: THE METDATA SYSTEM

Ubirajara Moura Freitas

Gilberto Camara

Ricardo Cartaxo Modesto de Souza

CPTEC
(Centre for Weather Prediction and Climate Studies)
INPE
(National Institute for Space Research)
BRAZIL

ABSTRACT

This paper describes the conceptual data model for METDATA, INPE/CPTEC's environmental data base. This model caters for the diversity of data sources and formats and is being used to develop an object-oriented environmental data base, integrated with a meteorological data visualisation and manipulation system.

1. INTRODUCTION

This paper describes a conceptual data model for a meteorological data base, which serves a basis for a spatial data base management system (METDATA), to be used in Brazil's Centre for Weather Prediction and Climate Studies.

The ideas here presented are intended to serve as the basis for a complete meteorological archival and retrieval system, which caters for all types of data used on a NWP centre. METDATA is able to integrate, in a single data base, spatial information from diverse sources, including fields, satellite and radar imagery, meteorological observations, and geophysical and geochemical data [7].

The METDATA data model is being developed with a view to preserve the concepts used in METVIEW/WS, the meteorological visualization and manipulation system, developed in cooperation by INPE/CPTEC and the European Centre for Medium Range Weather Forecasts [2]. This allows METDATA to share the METVIEW/WS classes and methods for handling spatial data.

At INPE/CPTEC, the combination of METVIEW/WS and METDATA will provide facilities for data retrieval, visualisation and manipulation in a single environment. In a general sense, the information used by METVIEW/WS is simply a collection of data selected and copied by METDATA onto the user's workspace.

The model presented here has been tested in the implementation of a prototype of the METVIEW/WS system.

We believe that the model is generally applicable to the design of complex meteorological systems handling data in different formats. Based on earlier work by the same group ([1], [3] and [5]), the model presented enables the effective application of object-oriented programming concept and to derive an entity-relationship model [4].

2. METDATA REQUIREMENTS

The METDATA system is designed to provide archival, retrieval and selection of meteorological data for INPE/CPTEC. The concept "meteorological data" denotes NWP fields, observations, satellite imagery and graphical products.

2.1. OPERATIONAL ENVIRONMENT

METDATA works on UNIX workstations and supercomputers, with X Window System-based graphical user interfaces, and use the C++ object-oriented language[9], and is built on top of a relational DBMS. The workstations, servers and supercomputers are linked on a local-area network, with Ethernet and FDDI protocols.

A distributed processing architecture, based on the client-server model, is used. RPC ("remote procedure call") and NFS ("network file system") services are used for that purpose.

Use of standard data formats GRIB and BUFR will simplify data organization and retrieval, and present a common access pattern for all applications.

2.2. FUNCTIONAL REQUIREMENTS

The functional requirements for METDATA are:

- facilities for handling large spatial data bases, with heterogeneous information (hundreds of gigabytes);
- data base queries on the properties of objects;
- efficiency on search operations to guarantee interactivity;
- distributed and hierarchical data archival (local and remote), includingsystem;
- dedicated functions for spectral to grid point conversion and sigma to pressure level interpolation.

3. DESIGN CONSIDERATIONS

Meteorological data is completely different from typical commercial data. From a general point of view, it consists of large quantities of unstructured, independent atomic components, with time range variations from one_day model runs to climate simulations that can cover many years. There is no inherent limitation on the way such data is grouped or retrieved, unlike normal commercial uses. Typical meteorological applications simply require fast access to large quantities of 2D and 3D arrays.

The design option chosen was to use relational data base technology to build METDATA. This organization enables easy construction of retrieval operations. The need for developing indexing access methods is minimized, since these operations will be implemented in the data base.

The access methods for spatial data cannot be fully implemented on a traditional relational system. Therefore, in METDATA, relations on the data base store descriptive information about each element. The spatial information is kept in separate files.

4. BASIC CONCEPTS FOR THE MODEL

The challenge for the developers of DBMSs with spatial capabilities lies not so much on special purpose data structures, but in defining abstractions that offer generic spatial data management functions and that can be tailored to the different user requirements [7].

With that challenge in view, the METDATA model aims:

- to provide a representation for meteorological data;
- to enable observations, cartographical data, images and fields to be combined in a single system;
- to identify the data base entities and their relationships;

- to serve as a basis for the definition of clasmethods for object-oriented progra mming.

The data model comprises a series of concepts, with different levels of abstractions, described below.

4.1. DATA UNIT

The central concept of the model is that of a data unit. Each data unit corresponds to an invidualizable entity, and the meteorological systems and the data base will be able to manipulate them. Examples of data units are: a TEMP observation, a 2D global temperature field and a spectral band of a METEOSAT image.

Each data unit corresponds to an element of the data base, to be controlled individually. The creation of a data unit within the data base only occurs when it is possible to identify it from others.

Data units have an unique identification, spatial atributes, temporal atributes and properties.

4.2. DATA GROUPS

It is convenient to group the data units in meaningful chunks with identical properties. A natural solution is to aggregate all relevant data units for a given time instance and a specific numerical model. A data group constructed in that way corresponds to the concept of state of the atmosphere.

The state of the atmosphere corresponds to the output generated by a numerical model, and will represent that model's estimation of the atmosphere in a given moment. It is a convenient archival unit. For example, model experiments will request the retrieval of initial conditions, which would be grouped and retrieved together.

4.3. DATA SET

The idea of a data set simplifies the operation and use of the data base. A data set combines data groups that are likely to be retrieved and stored together. Some examples of data sets are:

the archived 8-day forecast data set, consisting of 4 instances per forecast day (plus the 4 initialization analysis), taking 6 variables at 19 model levels, plus 20 diagnostics fields. This data set mighdata groups, each containing one time ins tance of the forecast. A total of 4824 fields is contained in this data set. Typical size of this data set: 800 Mbytes;

a visualization data set, containing 5 variables at levels, with all model time steps (72 instances) for 3 days. Typical size of this data set: 1.2 GBytes. This data set is to be used for visualization experiment purposes.

- The typical sizes reflect the object hierarchy. Fields have between 10-150 Kbytes, data groups have typically 1-80 Mbytes and data sets have roughly 200-1000 Mbytes.
- The actual location of the data sets will consider operational considerations, such as processing efficiency. A possible organization is to have operational forecast resident at the super computer; the visualisation data set, at the graphics superworkstation.

4.4. PARAMETER

In METDATA, the notion of a parameter distinguishes amongst the different kinds of meteorological data. Each parameter represents meteorological data of a different nature, viz.: Temperature, Wind, Satellite Images, Synoptic Charts.

The idea of a parameter helps the user to distinguish the nature of data base components. Since METDATA is a dedicated system, the data selection and retrieval operations should be made close to the user's knowledge domain.

4.5. GEOMETRIC MODELS

The possible number of parameters is enormous. To simplify the development of methods for retrieval, visualization and analysis, parameters are specialised into geometric models. Each geometric model is mapped to a different graphical representation.

The concept of geometric models simplifies the construction of METDATA. The diversity of meteorological data is reduced to a small number of geometric models, namely:

- Digital Model (Fields): spatial distribution of fields).
- Observational model: locational entities, which indicate the existence of a ponctual phenomenon in space. These data may represent measurements made in meteorological data collecting stations, historical climatological information, digital terrain samples and geophysical and geochemical measurements.
- Image model: Images generated by satellites (e.g., GOES, METEOSAT, TIROS and LANDSAT), by radar or by operations in other images.
- Thematic model: geographical regions defined by one or more polygons, where each region corresponds to a different theme. Example: Earth coastlines.

4.6. DATA ARCHIVE

METDATA's data archive is the main facility for data storage. Communication between the data archive and the user is essentially of data retrieval, according to selection criteria defined by him. All requested data is transferred to a on-line data bases.

The idea of a having a data copied from the data archive to on- line data bases has a practical motivation. The access to a large data base can be a time-consuming operation, since network bandwidths offer a limitation to remote data manipulation and visualisation. Additionally, the user may request a given set of data for large periods of time, and it is likely that data access times can be minimized if a local copy is available.

Finally, data consistency is a major concern in large data bases. Since most requests for data are of a "read-only" nature, control of multiple data versions is greatly simplified if such data is copied to on-line data bases.

4.7. ON-LINE DATA BASES

The primary objective of METDATA is to provide data access for the various applications of a meteorological centre. For reasons of performance and organization, the on-line may be organized in one or more on-lin
e data bases.

The on-line data bases will contain data acessible to the user (normally in a read-only manner) and there are no restrictions to the number of data bases the user may attach during a session.

4.8. WORKSPACE

The particular on-line data base under the control of the METDATA and METVIEW/WS user is called a workspace. During a session the system must be attached to a workspace, where he has a write access.

There are no restrictions on data that make up a workspace. Data can be generated by maniputation over other data and stored in the workspace. The user can also define a selection criteria, and there will be functions that retrieve data into it from the data archive, from individual files or from other on-line data bases.

4.9. AUXILIARY CONCEPTS

Besides the basic concepts described above, METDATA needs a number of auxiliary notions that are used for its implementation.

The auxiliary concepts projection, level, and spectral band are all associated to a data unit, and the notion of a type to a data set.

Each data unit is associated to a cartographic projection supported by METDATA. This notion is used for the storage of satellite imagery. The isobaric or model level is an attribute of a data unit, used as a filter for selection and visualisation. The spectral band is the logical equivalent of the isobaric level in the case of imagery.

The type of a data set is the selection criteria used for its definition. are examples of different meteorological types.

5. ENTITY-RELATIONSHIP SCHEMA FOR METDATA

The conceptual model of METDATA has been mapped into an entity- relationship schema, where the basic entities are data unit, data group and data set.

The relationship between the data unit and the data group is one of aggregation. Data uniand geographic region are related to one and only one da
ta group tuple.

The relationship between data group and data set is an arbitrary one and depends on opertional considerations.

The relation between data units and geometric models is one of specialisation. Depending on the parameter associated to it, a data unit is specialized in one of the geometric models. The current METDATA implementation deals with the following specialisation entities: Grid, Image, Observation, Sample, Polygonal Map, Point and Chart.

The other entities of METDATA schema are: parameter, projection, level, spectral band and report type. These entities and associated relationships define the possible attributes values of the basic entities.

6. THE MODEL IN PRACTICE

The model has enabled the seamless integration of the METDATA system with METVIEW/WS, since the entities and relationships of METDATA correspond to the classes of METVIEW/WS. There is only one menu-driven interface to the system, that hides the spatial DBMS, the image processing and graphical methods and the data structures. This achieves a seamless integration of the various components of an meteorological information system.

A prototype of METDATA is being developed at INPE/CPTEC, using the PostGRES DBMS [8]. The methods for archival, retrieval and manipulation of data are integrated to a DATA BASE version of METVIEW/WS 0.1. This special version operates over a workspace data base with the presented model. Each concept described above has been implemented as a C++ class and the hierarchy of classes reflects the relationships between the entities.

Special care has been taken to use only the basic facilities that are available in all comercial systems. METDATA does not require large fields or some other special features.

An interactive interface generates a query for selection of data units into the workspace. The desired atributes are specified producing a data selection equivalent to a MARS query [6].

The current version has facilities to operate with satellite images and fields in GRIB format. Observations in BUFR format are planned to be included in the first semester of 1992. The first version of METDATA will be operational in the second semester of 1992, at INPE/CPTEC instalations.

ACKNOWLEDGMENTS

Arry Buss, Elisa Nishimura and Andre Battaiola from INPE/CPTEC and Jens Daabeck from ECMWF have provided useful and important comments on earlier versions of this document.

The authors are indebted to Drs. Marco Casanova and Luiz Tucherman from the IBM Rio Scientific Center for many useful discussions on the general subject of spatial data bases.

REFERENCES

[1] D. S. Alves, Data Models for Geographic Information Systems, PhD. Thesis. University of Sao Paulo, 1990. (in Portuguese).

[2] Camara,G.; Daabeck, J. "METVIEW/WS: The conception of a new system for visualization of meterological and climatological data ", in Proc. 7th. International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hidrolology, New Orleans, LA, Jan. 1991.

[3] G. Camara, J.R. Oliveira, F.Y. Yamaguchi, F.R.D. Velasco, R.C.M. Souza, "Abstract Data Types in Computer Graphics and Image Processing", in Proc. II Brazilian Congress on Software Engineering, Canela, Brazil, October 1988 (in Portuguese).

[4] P. Chen, "The Entity-Relatioship Model-Toward a Unified View of Data", ACM Trans. Database Sys., vol.1, no.1, Mar. 1986, pp. 9-36.

[5] G. Erthal, G. Camara and D.S. Alves, "A General Data Model for Geographic Information Systems", in International Symposium on Photogammetry and Remote Sensing, Kyoto, Japan, 1988.

[6Gibson,
Workshop on Meteorological Operations Systems, 1987. Proceedings.

[7] Guenther,O.; Buchmann, A. "Research Issues on Spatial Databases", ACM SIGMOD Record, 19:4, December 1990.

[8] M. Stonebraker and L. Rowe, "The Design of PostGRES", in Proc. ACM-SIGMOD Int. Conf. Management of Data, Washington, DC, May 1986.

[9] B. Stroustroup and M. Ellis, The Annotated C++ Reference Model, Reading, MA: Addison-Wesley, 1990.

APPENDIX A
DESCRIPTION OF METDATA RELATIONS

The following paragraphs define, in simplified way, the relevant relations (tables) of METDATA and its atributes.

1. DATA SET

Aggregates data groups related to a same time reference time and data type. Its attributes are:

- Data set identification;
- Suite type, e. g. operational, research;
- Type of data, e. g. visualisation, climatology, verification, analysis, initialized analysis, first guess, forecast;
- Time reference.

2. DATA GROUP

Agregates data units related to an time instance, geographical region and projection. Its attributes are:

- Data group identification;
- Data set identification;
- Time reference, initial time and final time;
- Geographical region (latitude and longitude of left- down and upper-right corners of the target area);
- Projection identification, key to the projection relation, which define all parameters required to the navigation over the data;
- Bounding box in projection coordinates;
- File path, host machine and directory where the associated data are stored, if data is online;
- Tape Identification, name and tape number where a copy of data are stored.
- Control information like created date, security tags.

3. DATA UNIT

Contains the attributes of all meteorological data stored system. Its attributes are:

- Data unit identification;
- Parameter identification;
- MetClass identification;
- Representation (grid, image, observation, sample, poligonalmap, point, chart);

4. GRID

Data unit specialisation used to store regular grid data, like meteorological fields. Its attributes are:

- Data unit identification;

- Level Type and Level specification;
- Number of grid points (lines and columns);
- Number of components, fields like wind have two components);
- GRIB file name and indices to the GRIB data. The fields itsef are stored in a Standard GRIB file.

5. IMAGE

Data unit specialisation that stores images, like METEOSAT and GOES images. Its attributes are:

- Data unit identification;
- Sensor identification, key to Sensor table which defines the satellite sensor or other devices;
- Channel identification, key to Sensor related table Channel, which define the spectral band and other atributes of the image;
- Number of image points (lines and columns);
- GRIB file name and index.

The images are stored in a extended GRIB file format. The METEOSAT and GOES Images use the space view image format grid proposed to WMO. TIROS/NOAA and other satellite images use a sligth different format, to permit navigation with landmarks. Images, in classical cartographical projections, use a other format. These formats respect the original ideas in GRIB with differences to accept the new requeriments.

6. SAMPLE

Data unit specialisation that stores data in a non regular grid. Data obtained from meteorological stations are suitable to be stored in this representation. This organization correspond to a Metview class and special methods can be used to generate regular grids from non regular ones. Its atributes are:

- Data unit.IP - Level Type and Level specification;
- Number of grid points;
- Number of components, parameters like wind have two components;
- BUFR file name and indices. All grid points are stored in a standard BUFR file.

7. OBSERVATION

Data unit specialization used to store meteorological observations. Each observed parameter is stored as a data unit in the relation which specify the location. The atributes are:

- Data unit identification;
- Level Type and Level specification;
- Location (latitude and longitude);
- Observation value;

8. POLYGONAL MAP

Data unit specialisation use to store opened and closed polygons. This data normally corresponds to map information like coast lines, aerial routes,dams and rivers used as a visual overlay of meteorological information or images. Its attributes are:

- Data unit identification;
- Bounding box of polygon map in projection coordinates;
- Area;

- Vector and topological file name;

9. POINT

Data unit specialisation which stores special symbols, for visualization purposes. Its attributes are:

- Data unit identification;
- Position, in projection coordinates;
- Symbol value;

10. CHART

Data unit specialisation which deals with visual ready to use products, like a sinotic chart or a meteogram. These data units can not be manipulated and are stored to distribution purposes. Its attributes are:

- Data unit identification;
- Chart name and type;
- Meta file standard and file name. PostScript is the normal format used.

11. DESCRIPTIVE RELATIONS

METDATA defines other relations that are referenced by some atributes of the basic relations, establishing relationships among them. These relations are:

- Parameter;
- Metclass, an week relation associated with Paramete.IP - Projection, defines a projection type, projection atributes and earth ellipsoid atributes. These atributes define the convertion equations from projection coordinates to latitude and longitude and vice-versa;
- Sensor;
- Channel;
- Report type, for observations, e. g. TEMP, SYNOP, SHIP