

# Software for storage and processing coded messages for the international exchange of meteorological information

V N Popov, I A Botygin, A S Kolochev

Tomsk Polytechnic University, 30, Lenina ave., Tomsk, 634050, Russia

E-mail: [vnp@tpu.ru](mailto:vnp@tpu.ru)

**Abstract.** The approach allows representing data of international codes for exchange of meteorological information using metadescription as the formalism associated with certain categories of resources. Development of metadata components was based on an analysis of the data of surface meteorological observations, atmosphere vertical sounding, atmosphere wind sounding, weather radar observing, observations from satellites and others. A common set of metadata components was formed including classes, divisions and groups for a generalized description of the meteorological data. The structure and content of the main components of a generalized metadescription are presented in detail by the example of representation of meteorological observations from land and sea stations. The functional structure of a distributed computing system is described. It allows organizing the storage of large volumes of meteorological data for their further processing in the solution of problems of the analysis and forecasting of climatic processes.

## 1. Introduction

In recent years, coded messages are widely used for the international exchange of meteorological information. These include data of observations and the processed data. Coded messages are also used for the international exchange of data required for the specific application of meteorology in various fields of human activity, and for the exchange of information related to meteorology. The codes are composed of a set of code forms and binary codes consisting of characters (letters or groups of letters) indicating the meteorological or, in certain cases, other geophysical elements. In accordance with the specifications in the messages, these characters are replaced by numbers indicating the amount or the state of the described elements. In some cases, the specification characters allow carrying out their replacement numbers. In other cases, one must use the code numbers, the specifications of which are given in the code tables. In addition, a certain amount of symbolic words and symbolic figure groups have been developed for use as code names, code words, prefixes or distinctive character groups.

In this regard, the development or adaptation of metadata specification, which would help to describe the format and content of data in order to achieve interoperability with their functional processing, is an urgent task. This paper, based on the analysis of international codes of meteorological information (surface meteorological observations, vertical atmospheric sounding, wind sounding of the atmosphere, weather radar observations, observations from satellites and others), describes the software which allows forming a unified data descriptors using a hierarchical scheme of meta descriptions. This scheme includes components such as classes, sections and groups.



## **2. The structure of the basic classes of data of meteorological observations**

The most prevalent codes, containing meteorological information, are codes FM 12-VII SINOP and FM 13-VII SHIP. These codes contain the data of meteorological observations from land and sea stations. They include four main sections. Each section consists of several groups. For example, the first section includes the following groups: a code identifier, the date of the month, an observation period, a station index, a station type pointer, and others. In the second section, the following information is present: meteorological visibility, the total number of cloud layers, the wind direction for a period of observation, the average wind speed for the time of observation, air temperature, a dew point, air pressure at the station, the amount of precipitation over a certain period of time, and others. The third section consists of: the maximum temperature of the day, the minimum air temperature for the night, duration of sunshine per day, instrumental measurements of the height of the cloud base, the number and type of clouds, more information about the weather in the period and between the time of observation, and others. The fourth section describes such points as: the minimum air temperature, the state of the soil in the absence of snow cover, the minimum temperature of the soil surface for the night, the amount of precipitation over a certain period of time, the amount of precipitation per day, and others [1–4].

Other international codes use similar format of structuring, but they have their own sections and groups.

Thus, even a brief description of the structure and composition of the code forms proves the need to develop a special software for storage and processing of these forms.

## **3. Metadescription of the code structure**

In order to manage the processes of creating, storing, updating and processing the huge volume of spatial data and services, metadata should be formed. The main purpose of the formation of metadescription is to increase the availability of a variety of meteorological data together with related information and to improve cooperation and coordination of efforts in data collection and processing. The existing practice of application and using metadescriptions as the formalism, associated with certain categories of resources, is very broad and diverse, as well as the number of used metadescription formats. In the classification of formats, a key feature is usually the description of the subject area. There are metadescription formats for representation of persons and organizations, archives and electronic resources, bibliographic resources, and others.

The most advanced metadescriptions are in the standards of presentation and modeling of spatial metadata. Note such standards and profiles of spatial metadata as the international standard “ISO 19115: 2003 “Geographic information. Metadata”, the American standard “FGDCSTD–001–1998”, the Russian standard “GOST R 52573–2006 “Geographic information. Metadata”, the European CEN standard (prEN 12657 Geographic Information – Metadata), the Australian and New Zealand standard “ANZLIC”, the UK standard – UK GEMINI, and others.

International codes of representation of meteorological data include more than 40 classes of data presentation: surface meteorological observations (FM 12-IX SYNOP, FM 13-IX SHIP), vertical atmospheric sounding (FM 35-V TEMP, FM 36-V TEMP SHIP), wind sounding of the atmosphere (FM 32-V PILOT и FM 33-V PILOT SHIP), weather radar observations (FM 20-V RADOB), observations from satellites of the Earth (FM 86-VI SATEM, FM 87-VI SARAD, FM 88-VI SATOB), and others.

A specialized model of data representation and functional services was designed on the basis of a common metadata schema and on the basis of the above-mentioned international codes and the analysis of problems of interaction of distributed system components during their processing [5].

Metadata elements support the following functions to work with meteorological information:

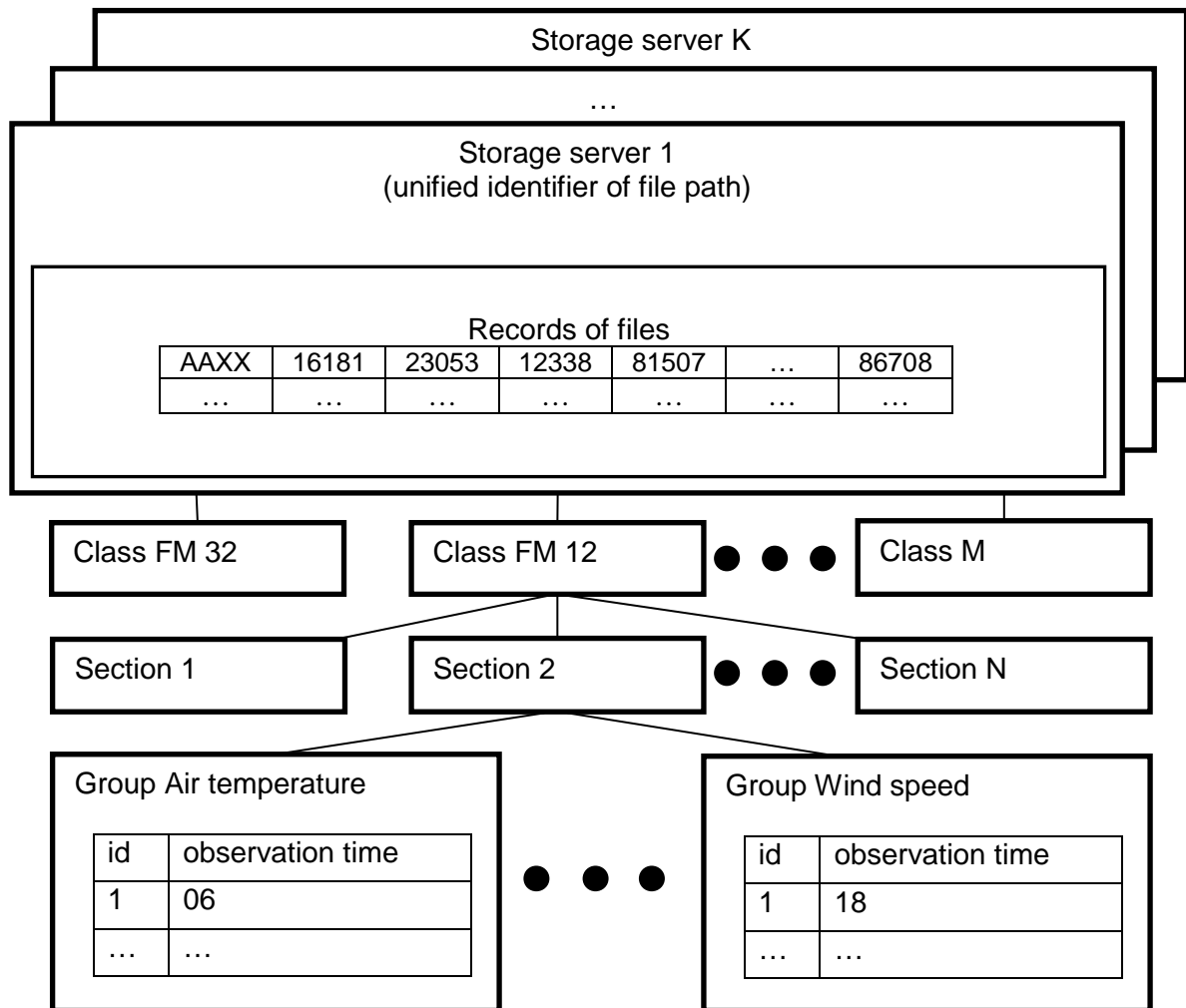
- searching for information needed to determine the sets of data which are available on a specific geographical area;
- definition of the purpose and suitability of the information (conformity assessment of the specific needs of the data set);

- implementation of access to information resources which are required for a selected set of data and services;
- usage of resources, i.e. processing and using data sets and services.

Metadata which describe meteorological resources are divided into classes. The base class describes the metadata themselves. Other classes include information about a certain aspect of the described set of meteorological data:

1. the characteristics of a set of metadata are semantic description of stored data;
2. identification contains a unique identifier of the data set;
3. information about updates is the date and time of updating of the meteorological data;
4. the organization and maintenance of meteorological data in the set describes the functional structure of the code;
5. the applicable set of symbols is used to refer to non-numerical information;
6. custom metadata extensions are additional meta-description tags;
7. the processing application scheme is program code sets, which perform standard statistical transformations on the data;
8. spatio-temporal length specifies the time interval of observations.

Figures 1 shows a hierarchical structure illustrating the placement of several groups of data representation in a metadescription.



**Figure 1.** The hierarchical structure of metadescription components.

The class consists of one or more sections which are connected relationships of generalization. The sections can be repeated within the same class to solve user problems. Each section is a collection of elements (groups) which characterize one or another aspect of meteorological data. Metadata elements can be mandatory, optional or conditional.

In the hierarchical structure of the metadescription components at the lowest level, there are the metadata which describe the specific measured parameters of meteorological observations (wind speed, air temperature, etc.). The main characteristic of this metadata is an identification number or index in records of the physical file storage. The metadata of the top-level metadescriptions define the unified identifier of the storage server and the path to the physical file of meteorological observation data storage. All metadata of intermediate levels provide forming parts of a unified identifier path to the storage server (the class of data representation, the section, the group). The specific filename with the stored data is generated based on the table of the template names for classes of data representation [6].

The unified path to the file with specific data is a set of special tags (keywords). Thus, access to specific data of different classes is performed uniformly and rapidly. An exposed set of keywords, in fact, is a hash value (address) of required data.

#### **4. The functional structure of the software for storing and processing coded reports**

The functional structure of the software is based on the concept of an open horizontally scalable distributed information-computing framework for distributed computing and distributed data management [7, 8]. Note that this approach is not contrary to the modern concepts in the development of Computer Science and is formulated by experts as a "personal grid." Nodes of this distributed computing system are common desktop computers available practically in all researcher workplaces, which have special network applications – agents. Agents are additional networking software modules operating in an infrastructure of distributed information and a computing framework at its nodes, communicating with each other and ensuring the proper implementation of parallel computing in a distributed system.

Functionality of this system includes: searching, collecting, cataloging of meteorological data; placement of meteorological data in the repository and providing access to them; providing access to distributed meteorological data via standard protocols; visualization and editing elements of meteorological data in different formats.

The framework is represented in the form of the decentralized architecture of resources with decentralized computing and data management [9, 10]. The nodes of the storage and computing cluster receive software data processing modules and meteorological data input buffers from the control center. A special component of the computing center determines the order of order data placement and the process of their distribution on the physical nodes of the framework. Data flows between the software modules of the placement and computing is implemented on a common field of external disk storage. Access to the nodes of clusters is based on a socket interface. The problem of the horizontal scaling data-processing capacity of such distributed system (the storage and computing clusters) can be solved simply by connecting additional nodes and deploying of software agents on these nodes, the number of which can be increased at any time.

The interaction between researchers and the environment is carried out through the web-portal. A control web-portal is a site that provides services to the user to work with analytic- computing functional of the framework. The web-portal performs a graphical user interface function, which consists of a set of control components. These components provide researchers with a simple and convenient access to the framework services (communication services, infrastructure of computing service management, infrastructure of network storage). In addition, the web-portal is used to perform the system administration functions of the framework. Running functional applications in the web-portal is based on the Java Web Start technology.

#### **5. Conclusion**

The functional structure of the software for the organization of multi-level distributed repository of coded messages for the basic forms of the international exchange of meteorological information is presented. To represent data of codes for international exchange, the scheme of metadescription is developed, including such categories of resources as the class, the section, the group. The structure and the content of the metadata are based on the analysis of data representation of surface meteorological observations, vertical atmospheric sounding, wind sounding of the atmosphere, weather radar observations, observations from satellites, and others.

The functional-logical scheme of the distributed computing system based on personal computers is offered for transformation of the metadescription on the physical infrastructure of the storage and computing cluster. The unified path to the file with specific meteorological data is a set of special tags. The access to specific data of various forms of international codes is performed uniformly. An exposed set of keywords is a hash value (uniform resource identifier) of stored meteorological data.

The studies show the effectiveness of using metadescriptions for the design of distributed storage of large amounts of diverse data in order to further their parallel processing in high performance cluster systems for the problems of analysis and forecasting of climatic processes.

## References

- [1] 2015 *Manual on Codes – International Codes* Vol I.1 (WMO) p 466
- [2] 2015 *Manual on Codes – International Codes* Vol I.2 (WMO) p 924
- [3] 2015 *Manual on Codes – International Codes* Vol I.3 (WMO) p 134
- [4] 2015 *Manual on Codes – International Codes* Vol II (WMO) p 354
- [5] Botygin I A, Popov V N, Tartakovsky V A and Sherstnev V S 2015 Architecture of scalability file system for meteorological observation data storing *Proc. SPIE 21st Int. Symp. Atmospheric and Ocean Optics: Atmospheric Physics* **9680** 1–4
- [6] Popov V N, Botygin I A and Koshelev N V 2016 Uniform Model of Representation Heterogeneous Data Hydrometeorological Observations *Key Engineering Materials* **685** 925–929
- [7] Pogrebnoy V K and Pogrebnoy A V 2004 Efficient placement of stations of topologically distributed multiprocessing computing systems *In. Proc. 8th Korean-Russia International Symposium on Science and Technology (KORUS-2004)* Vol 1 (Tomsk: Tomsk Polytechnic University) pp 137–141
- [8] Pogrebnoy A V and Pogrebnoy An V 2016 Allocation of Clusters of Objects on Topological Graphs of High Cardinality *Key Engineering Materials* **685** 952–956
- [9] Mesheryakov R, Moiseev A, Demin A, Dorofeev V and Sorokin V 2016 Using Parallel Computing in Queueing Network Simulation *Key Engineering Materials* **685** 943–947
- [10] Moiseev A, Demin A, Dorofeev V and Sorokin V 2016 Discrete-event approach to simulation of queueing networks *Key Engineering Materials* **685** 939–942