Ontological Description of Meteorological and Climate Data Collections

© A.A. Bart © V.V. Churuksaeva

Tomsk State University,
Tomsk, Russia

bart@math.tsu.ru

© A.Z. Fazliev © A.I. Privezentsev © E.P. Gordov © I.G. Okladnikov © A.G.Titov

Institute of Atmospheric Optics SB RAS,
Tomsk, Russia

faz@iao.ru remake@iao.ru gordov@scert.ru oig@scert.ru titov@scert.ru

Abstract. The first version of the primitive OWL-ontology of collections of climate and meteorological data of Institute of Monitoring of Climatic and Ecological Systems SB RAS is presented. The ontology is a component of expert and decision-making support systems intended for quick search for climate and meteorological data suitable for solution of a certain class of applied problems.

Keywords: ontology description of object domains, systematization of domain data, climate and meteorological data.

1 Introduction

Today every large meteorological center uses original meteorological models for calculation of climate and meteorological parameters, which can differ both in the level of detail and set of calculated values of physical parameters. During the reanalysis of a meteorological situation, key meteorological parameters corresponding to measurements at weather stations are usually taken into account.

The results of climatic numerical simulation, weather forecast, or reanalysis of meteorological fields are collections of meteorological parameters that characterize the state of the atmosphere. They are represented by data arrays in common formats, e.g., grib [7], netCDF [12], HDF5 [8], etc.

At Institute of Monitoring of Climate and Ecological Systems SB RAS (IMCES SB RAS), the data processing environment [3] has been developed for representing collections of meteorological data; the environment is provided by sets of metadata that characterize physical parameters entering into the above collections. The practice showed restriction of the use of only localized applications in this environment. Inclusion of external applications resulted in creation of a new system – virtual information platform “Climate+” [17], where data are represented in the netCDF format.

When using climate data from different collections of numerous data manufacturers, the problem arises of ambiguous identification of physical parameters from these collections. The sense of physical parameters in the collections agrees with physical parameters advised by World Meteorological Organization (WMO). They are described in the taxonomy of the WMO ontology Codes Registry [19], as well as in the taxonomy of the ontology of the GRIB Discipline Collection [16] intended for the use in the Climate Information Platform for Copernicus (CLIPC).

The ontology description of data collections in the form of a primitive (simplified) formal OWL-ontology is intended for the selection of data collections within an expert system, which can be used during solution of an applied task of an object domain.

The ontology approach selected for the solution of the problem stated consists in the following. An ontology description is constructed for an applied problem. In addition to the physical statement, the description should include the mathematical statement of the task, i.e., a mathematical model with equations. Variables, which conform the WMO classification, and limitations are described in the form of an OWL-ontology. On the one hand, the set of parameters includes common meteorological parameters, such as sea level pressure, surface pressure, air temperature and humidity, wind speed and direction, and so on. This allows comparison of the computed values with the weather station measurement results. On the other hand, both meteorological and climate models supplemented by an applied task compose a component of a more complex model, where the results of prognostic calculations by climate/meteorological parameters are used for the solution of applied problems in different fields of human activity. This, in turn, enriches collections of climate and meteorological data with values of new physical parameters.

Proceedings of the XIX International Conference “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL’2017), Moscow, Russia, October 10-13, 2017

2 Virtual data processing environment

Approaches used in the creation of the prototype of a subject virtual data processing environment (VDPS) for the analysis, estimation, and forecast of the impacts of global climate changes on the natural environment and climate of a region were mainly developed during the design of the “Climate” web GDS [4,5]. This sub-ject GDS has been designed with the use of up-to-date information and communication technologies, is based on the conceptions of spatial data infrastructure (SDI) [2, 10], and grounds a software infrastructure for the complex use of geophysical data and information sup-port of integrated multidisciplinary scientific researches in the modern quantitative meteorology. We have se-lected it as a subject component of VDPS for Earth sci-ences. A web geoportal [1, 9] is a single access point to subject spatial data, processing procedures and results [1, 9]. The portal allows a user to search for geoinfor-mation resources in metadata catalogues, to form sam-ples of spatial data according to their characteristics (access functionality), and to manage tools and applica-tions for data processing and mapping.

The GDS Web Client [6, 13] is the main tool of the user’s desktop. It ensures the fulfillment of OGC re-quirements for web services: spatial data visualization (Web Map Service—WMS), data representation in vec-tor (Web Feature Service—WFS) and bitmap formats (Web Coverage Service—WCS), and their geospatial processing. It provides for the access to collections of climate data and tools for their analysis and visualiza-tion of the results via typical GDS graphical web browser. The Web Client satisfies the general require-ments of INSPIRE standards and allows selection of data set, processing type, geographic region for the analysis of processes, and representation of the pro-cessing results of spatial data sets in the form of WMS/ WFS map layers in bitmap (PNG, JPG, Geo-TIFF), vector (KML, GML, Shape), and binary formats (NetCDF).

Today, the VDPS prototype combines data collec-tions (reanalyses and climate simulation results and weather station measurements) within the unified geo-portal, supports the statistical analysis of archive and required data, and provides access to the WRF and «Planet Simulator» models. In particular, a user can run a VDPS-integrated model, preprocess the results, pro-cess them numerically and analyze, and gain the results in graphical representation. The prototype provides for specialists that participate in a multidisciplinary re-search process prompt tools for integral study of climate and ecological systems on the global and regional scales. With these tools, a user that does not know pro-gramming is able of processing and graphically repre-senting multidimensional observation and simulation data in the unified interface with the use of the web browser.

3 VDPS prototype capabilities

Support of the following data sets is built in the pro-totype: NCEP/NCAR reanalysis, ed. II, JMA/ CRIEPI JRA-25 reanalysis, ECMWF ERA-40 reanalysis, ECMWF ERA Interim, MRI/JMA APHRODITE’s Wa-ter Resources Project data, DWD Global Precipitation Climatology Centre data, GMAO Modern Era-Retrospective analysis for Research and Applications (MERRA), reanalysis of the joint Project «Monitoring atmospheric composition and climate (MACC)», NO-AA-CIRES Twentieth Century Global Reanalysis, ver. II, NCEP Climate Forecast System Reanalysis (CFSR), simulation results obtained with the use of global and regional climate and meteorological models. Observa-tion data from weather stations from the territory of the former USSR for the 20th century included in the Post-GIS database are also accessible.

Data processing

1. Statistical characteristics of meteorological parame-ters: sample mean, variance, excess, median, mini-mum and maximum, and asymmetry.
2. Derived climate parameters: vegetation period dura-tion, sum of effective temperature, Selyaninov hy-drothermal coefficient.
3. Periodic variations: standard deviation, norms, aber-rations, amplitudes of diurnal and annual variations.
4. Non-periodic variations: duration and repeatability of atmospheric phenomena with meteorological pa-rameters below or above the limits specified at dif-ferent time points.

Then a user can either analyze the results or contin-ue adding new layers on the map. To study the results, the user is provided for a possibility of selecting a geo-graphical region, scaling, getting values from all layers at a point, additionally processing earlier results (e.g., comparison between data from different layers). In ad-dition to the direct analysis of geophysical data, a used can carry out joint researches with other user, share the results, and use proper data collections in the pro-cessing. In general, this hardware-software complex provides for distributed access, processing and visuali-zation of large collections of geospatial data with the use of cloud technologies.

The data processing environment “Climate” devel-oped at IMCES SB RAS limits possibilities of users by local software applications. A current task is to extend the environment by external user applications. For this, the corresponding problems should be specified in gen-eral. Below we describe one of possible classes of prob-lems connected with decision-making.

4 General definition of the problem

The “Climate+” virtual information platform in-cludes collections of meteorological and climate data. It is intended for the data representation with the use of GIS technologies. Its further development is oriented to providing researchers possibilities of using selected data sets or their parts as input data. Most collections include data related to some (not all) spatiotemporal objects of the Earth; different collections often include different sets of physical parameters. To search for required spa-tiotemporal objects and their meteorological and clime characteristics, it was necessary to create a corresponding expert system on the basis of a knowledge base on spatial objects of the data collections and their parameters.

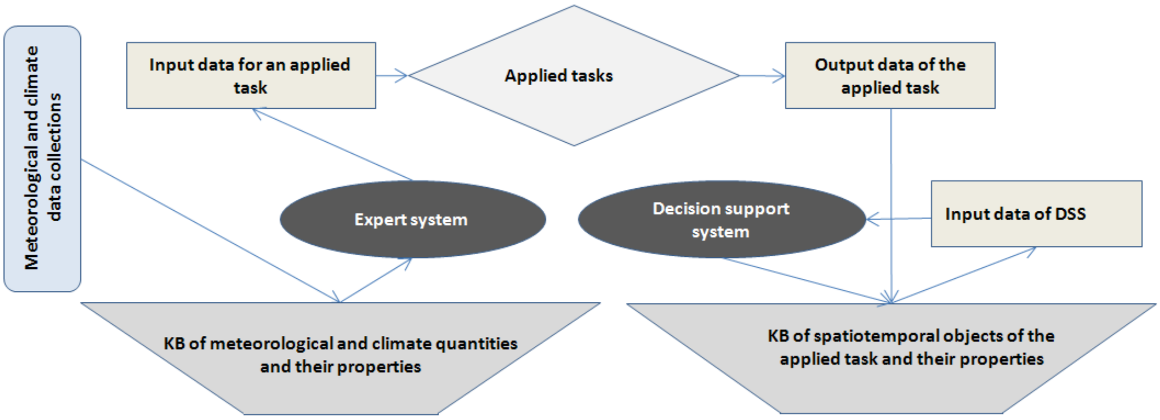

Figure 1 Simplified block-diagram of “Climate +” platform modification

Figure 1 shows a simplified block-diagram which is a basis of the “Climate +” platform modification. There are three groups of subsystems: meteorological and climate data collections; subsystem for work with knowledge bases (expert system for selecting input data for applied tasks and decision-making support system), and applied tasks with their input and output data. The data representation services are omitted.

In this work, we discuss questions of creation of a knowledge base for the expert system. The main problem which has been solved is substantiation of the reduction problem solution [20] or, in other words, construction of typical individuals of an OWL-ontology that characterize properties of spatiotemporal objects from the collections. The development of the conceptual part of the ontology (T- and R-box) is connected in our solution with classification of meteorological and climate parameters and is briefly described below.

5 Taxonomy of meteorological parameters

The OWL DL language [14] is used for the ontology description of the domain that generalizes, in particular, related spatiotemporal objects. These objects can be an air layer over a bounded territory, upper soil layer on this territory, or, in more specific cases, forests, fields, or long roads. There are physical and chemical processes connected with the objects; they are described by numerical models and used in calculations. Input values of the physical parameters are required for the calculations. The processes under study can relate to different temporal and spatial scales and be described on different levels of detail. Let us note that coupling of several mathematical models requires knowledge of sets of input and output parameters and their spatiotemporal characteristics.

The taxonomy of physical parameters allows forming sets of properties of spatiotemporal objects of a domain for solution of specific applied tasks. This taxonomy is used in the OWL-ontology for T-box construction.

When developing the decision-making support system on the basis of both meteorological and climate data, the parameters should be matched. Therefore, the WMO classification in version [11] is included in the ontology. This matching allows describing applied tasks of the domain in common terms.

There are climatic and meteorological resources [16, 19] that use the WMO classification of names of meteoparameters for the GRIB format for data storage [7]. First of all, WMO Codes Registry created for the aviation with the aim of supporting data exchange in the AvXML format; it is based on RDF and SKOS recommendation.

In our OWL-ontology of climate information resources, we created classes and individuals that correspond to names of meteorological parameters, e.g., the Meteorological_Products class and subclasses, according to [11]. In the primitive OWL-ontology of climate information resources described below, classes and individuals are created that correspond to names of meteorological parameters according to [11]. Individuals that unambiguously characterize physical parameters by their name [11] have been created in each subclass Thermodynamic_Stability_category, Atmospheric_Chemical_Constituents_category, Electrodynamics_category, Mass_category, Long-wave_radiation_category, Temperature_category, Short-wave_radiation_category, Aerosols_category, Moisture_category, Radiology_Imagery_category, Momentum_category, Trace_Gases_category, Cloud_category, and Physical_Atmospheric_category.

For the INMCM4 collection, which corresponds to output data of the INMCM4 climate model of general atmospheric and ocean circulation [18], classes and subclasses were created corresponding to model variables. These classes agree to the corresponding WMO classes.

6 Primitive ontology of “Climate+” platform data

The OWL DL developed and formalized ontology of climate information resources describes the current state of collections of data arrays of the data processing environment as one of the main Russian information resources on climate data. Numerical data are represented by data arrays that are stored in netCDF files. The data arrays are grouped in data sets. All data arrays in a set should: (a) be received at one temporal or spatial grid; (b) cover the same time interval; (c) be received under the same simulation or observation conditions (if possible); (d) be represented by a set of netCDF files, which include the same physical parameters. The data sets are grouped in data collections. A data collection is an ensemble of data sets received by an organization within a project, but represented on different spatial or temporal grids or for different model scenarios. In particular, a collection can consists of the only data set.

The basic classes in the OWL-ontology are: Collection, Spatiotemporal_object, Organization, Data_set, Data_array, Scenario, Spatial_resolution, Physical_quantity, Physical_quantity_values, Unit, Longitudes_array, Time_step, Latitudes_array, Height_levels_array, and Times_array. The spatiotemporal system is a four-dimensional object determined by arrays of numerical values of longitudes (Longitudes_array), latitudes (Latitudes_array), height levels (Height_levels_array), and time labels (Times_array), which are subclasses of the class of the list of values of a physical parameter and, therefore, numerical arrays of one physical parameter (Physical_quantity) in certain measurement units (Unit). They can be described by: the number of members of the array of a physical parameter (has_number_of_values), its minimal value (has_minimum_value) and maximal value (has_maximum_value), or by numerical values of the parameter (has_value). A data array (Data_array) is an ordered list of numerical values of a physical parameter (Physical_quantity), as a property of the spatiotemporal system (has_spatiotemporal_system), at each 4D point (longitude, latitude, height level, and time) of the spatiotemporal system (Spatiotemporal_system). In the OWL-ontology, a data array (Data_array) is a subclass of the class Physical_quantity_values and, hence, is a numerical array of values of one physical parameter (Physical_quantity) in certain measurement units (Unit); it is described by the number of members (has_number_of_values), maximal values (has_minimum_value) and minimal values (has_maximum_value) of the physical parameter. A data array (Data_array) belongs (has_data_array) to a data set (Data_set), which differs from other data sets by the model scenario (Scenario), spatial resolution (Spatial_resolution), time step (Time_step), and belonging to one collection (Collection). A data collection (Collection) consists of (has_data_set) data sets (Data_set) and belongs (has_organization) to one organization (Organization). The OWL properties of the climate data ontology are represented in Tables 1 and 2.

Table 1 Object properties of the ontology of climate information resources

Domain	Object Property	Range	id
Collection	has_organization	Organization	o01
Collection	has_data_set	Data_set	o02
Data_set	has_scenario	Scenario	o03
Data_set	has_spatial_resolution	Spatial_resolution	o04
Data_set	has_time_step	Time_step	o05
Data_set	has_data_array	Data_array	o06
Physical_quantity_values	has_physical_quantity	Physical_quantity	o07
Physical_quantity_values	has_unit	Unit	o08
Data_array	has_spatiotemporal_object	Spatiotemporal_object	o09
Spatiotemporal_object	has_longitudes_array	Longitudes_array	o10
Spatiotemporal_object	has_latitudes_array	Latitudes_array	o11
Spatiotemporal_object	has_height_levels_array	Height_levels_array	o12
Spatiotemporal_object	hat_times_array	Times_array	o13

Definitions of object properties are given in first three rows of Table 1; their unique identifying properties, in the fourth row; the range of definition (the first row) and range of values (the third row) are specified for each property. Definition of the data array properties are given in the first three rows of Table 2; unique identifying properties are given in the fourth row. The range of definition (the first row) and range of values (the third row) are specified for each property from the second row.

Table 2 Data type properties in the ontology of climate information resources

Domain	Datatype Property	Range	id
Physical_quantity_values	has_number_of_values	int	d01
Physical_quantity_values	has_minimum_value	float	d02
Physical_quantity_values	has_maximum_value	float	d03
Physical_quantity_values	has_value	float	d04
Times_array	has_time_start	str	d05
Times_array	has_time_end	str	d06

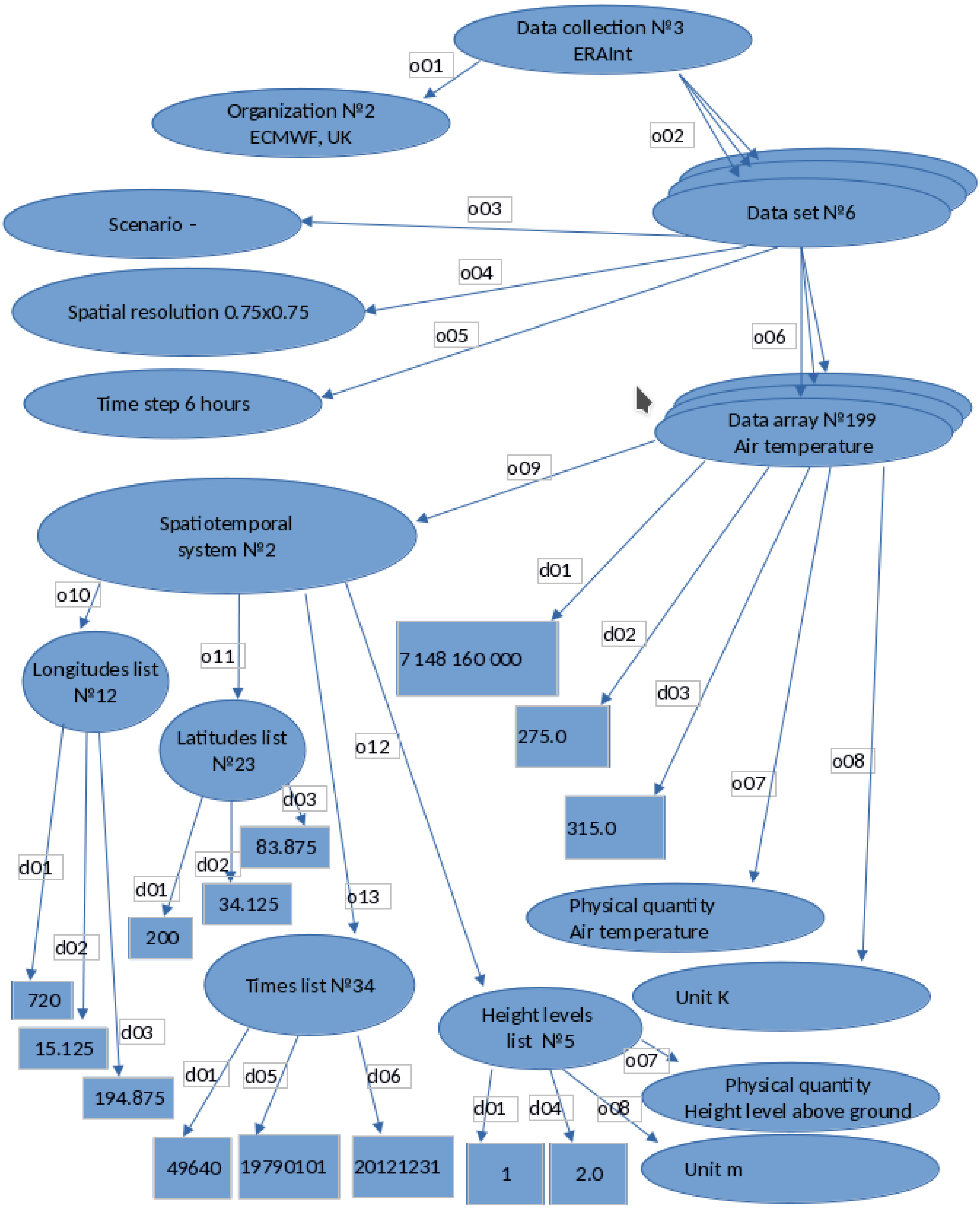

Figure 2 Simplified representation of individual describing ERAInt data collections

Figure 2 exemplifies a simplified individual of the OWL-ontology of climate information resources, used in the description of a ERAInt data collection, within the formal description of RDF resources [15].

Individuals of the OWL-ontology are shown in ovals; literal values are given in rectangles; the arrows show properties with unique identifiers in small rectangles, taken from Tables 1 and 2. Three arrows mean probable property cardinality higher than unity. Three overlapped ovals mean probable number of individuals of the OWL-ontology larger than unity. The individual “Data_collection” is connected by the property “has_data_set” with the individuals “Data_set”, each of which is connected by the property “has_data_array” with individuals ”Data_array”.

The domain analysis of climate numerical data arrays of the “Climate+” platform, stored as NetCDF files, allows the description of a primitive ontology of climate data of this platform in the OWL DL language. The primitive ontology is a simple and easily extended systematization of information resources required for the further work on the development of the decision-making support system.

To construct the climate data ontology of the “Climate+” platform the software has been developed for the formation of the fact-based block (A-box). An A-box has been formed for the climate data ontology using this software. Facts have been retrieved from the analysis of 80 Tb of climate data from the “Climate+” platform over 13 numerical data collections, which include 36 data sets and 793 data arrays. All the climate data collections include description of 170 spatiotemporal systems and 156 physical parameters that characterize properties of these systems.

7 Conclusions

The prototype of subject virtual data processing environment has been developed to provide for researchers, specialists, and people that make decisions an access to different geographically distributed and georeferenced resources and climate data processing services via a typical web browser. It includes a geoportal, systems for distributed storage, processing, and providing of spatial data and results of their processing. In particular, it allows the simultaneous analysis of several subject sets of climate data with the use of up-to-date statistical methods and, thus, revealing the impacts of climate changes on ecological processes and human activity. After finishing the work on the prototype, different interactive web tools are to be developed for the profound analysis of climatic variables and their derivatives provided by the subject geoportal.

The developed software is used for processing spatial datasets, including observation and reanalysis data, for the spatiotemporal analysis of recent and probable climate changes, with the special focus on extreme climate phenomena in northern latitudes.

The primitive OWL-ontology of climate and meteorological collections of IMCES SB RAS is constructed; it can be used for the search and selection of data for classes of applied problems in coupled decision support systems. The matching of physical parameters of applied tasks with IMCES SB RAS collections is carried out in WMO accepted terms.

8 Acknowledgment

The authors thank the Russian Science Foundation for the support of this work (developing of web-services and solution of reduction problems) under the grant No16-19-10257. We also thank Russian Foundation for Basic Research (16-07-01028) for the support of work (conceptualization of domains) partially described in the sections 4, 5, 6 of the article.

References

[1] Becirspahic, L., Karabegovic, A.: Web Portals for Visualizing and Searching Spatial Data. Inform. Comm. Techn., Electr. and Microelectr. (MIPRO), 2015, 38-th International Convention on, Opatija, pp. 305-311 (2015). doi: 10.1109/MIPRO.2015.7160284

[2] Frans, J. M., van der Wel: Spatial Data Infrastruc-Ture for Meteorological and Climatic Data. Meteorol. Appls., 12 (1), pp. 7-8 (2005)

[3] Gordov, E. P., Okladnikov, I. G., Titov, A. G.: Application of Web Mapping Technologies for Development of Information-Computational Systems for Georeferenced Data Analysis, Vestnik NGU, Ser. Information Technologies, 9 (4), pp. 94-102 (2011) (in Russian)

[4] Gordov, E.P., Lykosov, V.N., Krupchatnikov, V. N., Okladnikov, I. G., Titov, A. G., Shulgina, T. M.: Computational-information Technologies for Monitoring and Modeling of Climate Change and its Consequences. Novosibirsk: Nauka, 199 p. (2013) (in Russian)

[5] Gordov, E. P., Okladnikov, I. G., Titov, A. G. : Information and Computing Web-system for Interactive Analysis of Georeferenced Climatic Data Sets, Vestnik NGU, Ser. Information Technologies, 14 (1), pp. 13-22 (2016) (in Russian)

[6] Gordov, E., Shiklomanov, A., Okladnikov, I., Prusevich, A., Titov, A.: Development of Distributed Research Center for analysis of regional climatic and environmental changes, IOP Conf. Series: Earth and Environmental Science, 48, 012033 (2016)

[7] Guide to the WMO Table Driven Code Form Used for the Representation and Exchange of Regularly Spaced Data In Binary Form: FM 92 GRIB Edition 2. World Meteorological Organization Extranet. 2003. URL: http://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB2_062006.pdf

[8] HDF Group - HDF5: https://support.hdfgroup.org/HDF5/

[9] Koshkarev, A. V.: Geoportal as a Tool to Control Geospatial Data and Services, Geospatial Data, 2, pp. 6-14 (2008) (in Russian)

[10] Koshkarev, A. V., Ryakhovskii, A. V., Serebryakov, V. A.: Infrastructure of Distributed Environment of Spatial Data Storage, Search and Processing, Open Education, 5, pp. 61-73 (2010) (in Russian)

[11] NCEP/NCO Production Management Branch. NCEP WMO GRIB2 Documentation. National Weather Service Organization NCEP Central Operations. 2005. http://www.nco.ncep.noaa.gov/ pmb/docs/grib2/grib2_doc.shtml

[12] Network Common Data Form (NetCDF). https://www.unidata.ucar.edu/software/netcdf/

[13] Okladnikov, I. G., Gordov, E. P., Titov, A. G.: Development of Climate Data Storage and Processing Model. IOP Conf. Series: Earth and Environmental Science, 48, 012030 (2016)

[14] OWL 2 Web Ontology Language. RDF-Based Semantics (Second Edition), Eds: M. Schneider, F. J. Carroll, I. Herman, P. F. Patel-Schneider. W3C Recommendation 11 December 2012, http://www.w3.org/TR/2012/ REC-owl2-rdf-based-semantics-20121211/

[15] Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation 10 February 2004, Eds: Graham Klyne, Jeremy J. Carroll, http://www.w3.org/TR/2004/ REC-rdf-concepts-20040210/

[16] The GRIB Discipline Collection: [site] (2004). URL: http://vocab-test.ceda.ac.uk/collection/ grib/Discipline

[17] Titov, A. G., Gordov, E. P., Okladnikov, I. G.: Hardware-Software Platform «CLIMATE» as a Basis for Local Spacial Data Infrastructure Geoportal, Vestnik NGU, Ser. Information Technologies, 10 (4), pp. 104-111 (2012) (in Russian)

[18] Volodin, E. M., Dianskii, N. A., Gusev, A. V.: Simulating Present-day Climate with the INMCM 4.0 Coupled Model of the Atmospheric and Oceanic General Circulations, Izvestiya, Atmospheric and Oceanic Physics, 46 (4), pp. 414-431 (2010)

[19] WMO Codes Registry (2013). URL: http://codes.wmo.int/grib2

[20] Zinov'ev, A. A.: Foundations of the Logical Theory of Scientific Knowledge (Complex Logic), D. Reidel Publishing Company, 264 p.