Systematization of climate data in the virtual research environment on the basis of ontology approach

K. A. Alipova, A. A. Bart, A. Z. Fazliev, E. P. Gordov, I. G. Okladnikov, et al.

K. A. Alipova, A. A. Bart, A. Z. Fazliev, E. P. Gordov, I. G. Okladnikov, A. I. Privezentsev, A. G. Titov, "Systematization of climate data in the virtual research environment on the basis of ontology approach," Proc. SPIE 10466, 23rd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics, 1046675 (30 November 2017); doi: 10.1117/12.2289761

Event: XXIII International Symposium, Atmospheric and Ocean Optics, Atmospheric Physics, 2017, Irkutsk, Russian Federation

Systematization of climate data in the Virtual Research Environment on the basis of ontology approach

K.A. Alipova[1], A.A. Bart[1], A.Z. Fazliev[2], E.P. Gordov[1,2,3], I.G. Okladnikov[2,3],
A.I. Privezentsev[2], A.G. Titov[2,3]

1. Tomsk State University, Lenina av., Tomsk 634010, Russia
2. Institute of Atmospheric Optics SB RAS, Sq. Academician Zuev, 1, Tomsk 634055, Russia
3. Institute of Monitoring of Climatic and Ecological Systems SB RAS, Academichesky av., 10/3, Tomsk 634055, Russia

e-mail: bart@math.tsu.ru, gordov@scert.ru, faz@iao.ru, remake@iao.ru

ABSTRACT

The first version of a primitive OWL-ontology of collections 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 support systems intended for quick search for climate and meteorological data required for solution of a certain class of applied problems.

Keywords: ontology description of domain, systematization of domain data, climatic and meteorological data.

1. INTRODUCTION

Nowadays number of large meteorological centers developed original meteorological or climatic models and used those to calculate meteorological and climatic characteristics. Relevant calculation results can differ for different models and scenarios, both in the degree of detail of the description of processes in the used model and sets of input data. Also sets of calculated characteristics might be different. To store computation results various formats are used, such as grib[1], netCDF[2], HDF5[3], and others.

At the Institute of Monitoring of Climate and Ecological Systems SB RAS (IMCES SB RAS), sets of meteorological data together with sets of metadata that characterize physical quantities are presented to researchers in the Virtual Research Environment (VRE)[4] via human interface. To provide access to the data collections for external user's applications for solution of applied tasks (agent interface), a new system is created—the virtual information platform Climate+[5].

When using climate data from different collections of numerous data producers, the problem arises of ambiguous identification of physical quantities. User of the Climate+ platform should formulate the application in terms of physical quantities recommended by the World Meteorological Organization (WMO). Harmonization of physical quantities from the collections with WMO classification is ensured by personnel supporting the Climate+ platform. The formal data collections description which presented in the form of the applied Climate+ OWL-ontology is intended for selection of data collections within the expert system, which can be used as input to solve an applied task.

The ontological approach chosen to solve the problem is formulated as follows. An ontology description is created for an applied task. This description also includes the mathematical formulation of the task, i.e., a mathematical model. The

23rd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics, edited by
Gennadii G. Matvienko, Oleg A. Romanovskii, Proc. of SPIE Vol. 10466, 1046675
· doi: 10.1117/12.2289761

variables of these equations, which are meteorological parameters, are interpreted using the WMO classification and described in the OWL DL language. Basic meteorological variables, such as sea level pressure, surface pressure, air temperature and humidity, wind vector components, etc., are included into the set of parameters allowing one to compare the calculated values with the results of measurements at meteorological stations. At the same time usage of meteorological and climatic models to solve applied tasks in different subject areas leads to expansion of climatic and meteorological data sets due to appearance of new physical quantities.

2. METADATA DATABASE

2.1 Metadata database architecture

To describe geospatial datasets and their processing routines thus providing effective VRE functioning a dedicated metadata database (MDDB) is required. Currently, there is a lack of such database in the area of Earth sciences. The first attempt of comprehensive description of climatic geospatial datasets and processing routines in a single database is characterized by the following features.

The database (Fig. 1) contains spatial and temporal characteristics of available geospatial datasets, their locations, and run options of software components for data analysis. For internal data representation, the netCDF format is used in the platform. Here the following terminology is used. "Dataset" is a set of data which is a) given on the same temporal and spatial grid, b) covers the same time range and c) obtained under the same simulation or observation conditions (if applicable). It is represented by a collection of netCDF files containing the same set of meteorological parameters. It is necessary to distinguish the term "parameter" ("meteorological parameter") and "variable". Meteorological parameter is the name of some meteorological characteristic: temperature, pressure, humidity, etc. Variable is a unique name of a multidimensional array in a netCDF. Names of meteorological parameters are standardized. In contrast, the names of variables in different datasets could be different, and usually depend on preferences of an institution which produced them. Along with data, netCDF files contain horizontal, vertical and time domain grids. “Data collection” is a collection of datasets created by the organization within the specific project, but specified on different spatial and/or temporal grids, or for different scenarios. The collection may consist of one dataset. As in the framework of the same organization and the same project datasets with different spatial and temporal resolution can be created, the term "data collection" is introduced. Data collection is a collection of datasets created in the same organization within the same project, and given on various spatial and/or time grids, or for various scenarios. The collection may consist of one dataset.

Tables in MDDB are divided into "technical" and "interface". Technical tables contain data intended for computing software components. Interface tables hold string multilingual content for the graphical user interface. Some interface tables contain text records in different languages. Such tables have the suffix "_tr". The list of languages and corresponding codes are stored in a table **lang**. There are two major parts of MDDB providing description of climate datasets and description of data processing software components (computing modules).

2.2 Tables describing climate datasets

This part of MDDB contains characteristics of climate datasets, meteorological parameters and locations of data files on DSSs.

Each dataset is defined by a set of four characteristics: the name of the collection, the horizontal resolution, the time grid step and the name of the scenario (if applicable). Spatial resolutions of the horizontal grid are stored in a table **res**. Time grid steps are in a technical table **tstep**, and their names in different languages - in a table **tstep_tr**. Conventional names of scenarios (simulation conditions) are stored in a table **scenario**, which plays technical and interface roles. Descriptions of data collections in different languages are stored in an interface tables **collection_tr**, and their conventional names are stored in a technical table **collection**. Together with a table **timespan** containing time ranges covered with one data file; a table **filetype**, containing types of data files; a table **dsroot**, containing paths to the datasets; and a table **kind**, containing types of datasets, all these tables fully describe characteristics of climate datasets.
Each climate dataset includes one or several data arrays. Each such data array contains some meteorological parameter values given on a regular or irregular spatial and temporal grids, and is uniquely defined by a dataset, variable (meteorological parameter) and vertical level.

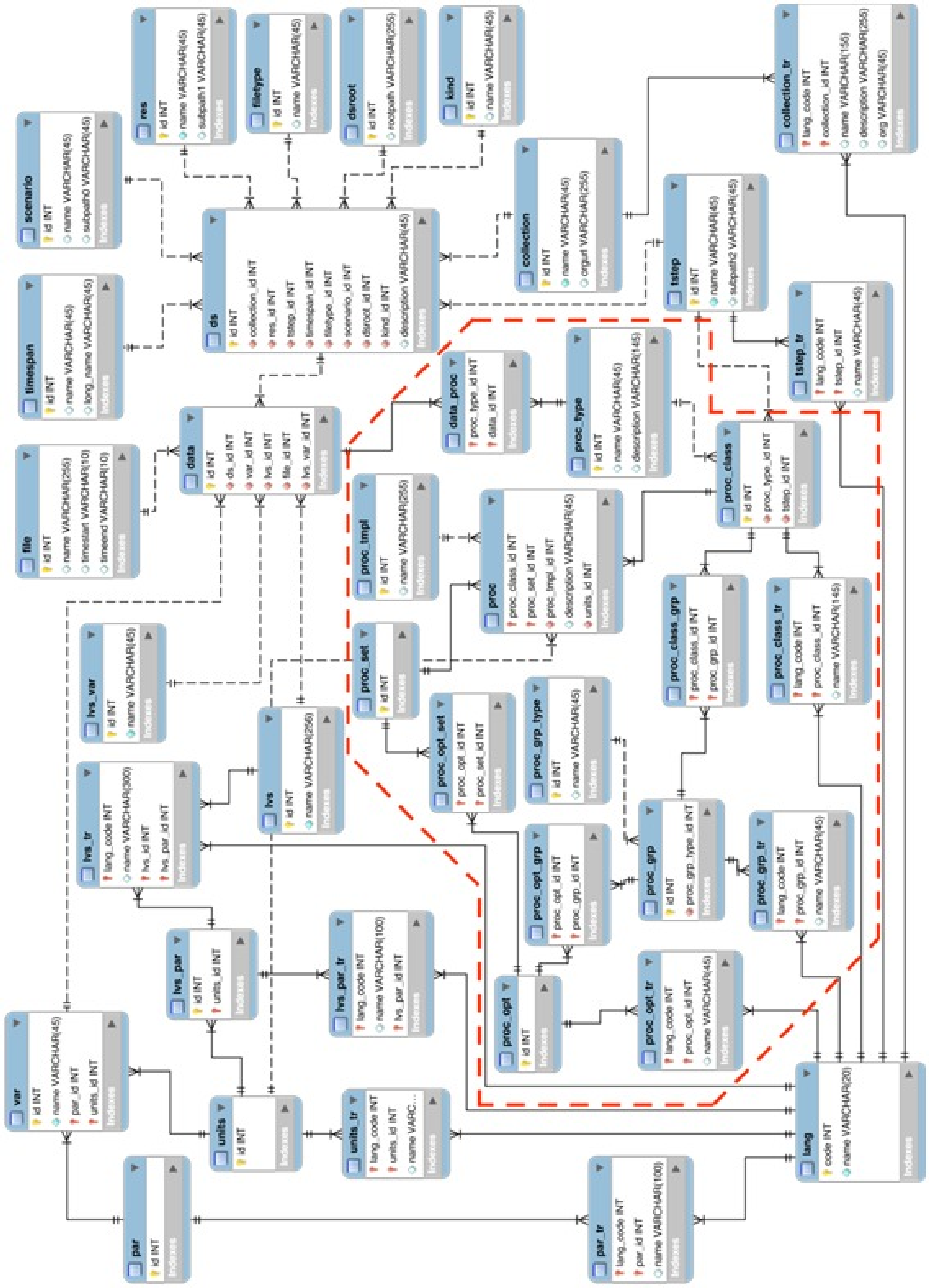

Figure 1. ER-diagram of metadata database. Red polygon interior – tables describing climate data computing components, exterior – tables describing climate data datasets

Names of meteorological parameters in different languages are stored in a table **par_tr**. Names of units in different languages are stored in a table **units_tr**. A table **var** contains names of variables and establish a connection between the names of meteorological parameters and variables in the netCDF files. Names of vertical levels in different languages are contained in an interface table **lvs_tr**, and conventional names of them as they are defined in the netCDF files are in a technical table **lvs**. All data arrays are described in a table **data** and linked with a table **file**. The table **file** contains name templates of data files which, in conjunction with partial paths from **dsroot** table containing the root directory of datasets, and **scenario**, **res** and **tstep** tables, allows one to construct a full path to the data files.

2.3 Tables describing data processing software components

This part of MDDB describes data processing components, their parameters and locations of template files that are required for preparation of task-files for computational software backend.

For data processing using computational backend one needs to prepare a special task-file in XML format. This task-file contains a description of call sequences of data processing software components. Such sequence (or workflow) is hereinafter referred to as "processor". Processors names (e.g., "Calculation of average") in different languages are contained in an interface table **proc_class_tr**. Each processor has some type specified in a table **proc_type**. Compliance of processors and types is contained in a table **proc_class**. Processors related to certain types may only process certain data arrays (e.g., the processor "Cold Night" may accept as an input only air temperature at 2 meters). Compliance of processor types and data arrays that they can handle is given in a table **data_proc**. The operation of some processors may depend on a user input. List of parameters of processors in different languages is stored in an interface table **proc_opt_tr**. The parameters are combined in groups. Each group has a type corresponding to the element of a graphical user interface used for displaying parameters names from this group, wherein the user may select or activate one or more parameters. Parameter groups indices are stored in a table **proc_grp**. Types of groups of parameters are stored in a table **proc_grp_type**. Pairs "parameter - group" (which parameter to which group belongs) is given in a table **proc_opt_grp**. Names of groups of parameters are stored in different languages in an interface table **proc_grp_tr**. Compliance of processors and groups of parameters (which processor which group of parameters has) is set in a table **proc_class_grp**. From each group of parameters corresponding to a certain processor, the user may select 0 or more parameters. Combinations of selected parameters are set in a table **proc_opt_set**, with each combination corresponding to one unique task-file template. Full paths to a variety of task-file templates for each combination "processor – selected parameters" are stored in a table **proc_tmpl**. On the basis of the selected by the user combination of parameters task-file template is selected, filled with necessary additional information about data being processed and spatio-temporal domain and passed to the computing backend.

3. "CLIMATE+" PLATFORM STRUCTURE

The Climate+ virtual information platform is aimed at presenting data using GIS technologies (human interface). Its further development is aimed at enabling researchers to use sets of climate or meteorological data selected by them or parts of the sets as input data in their applied tasks via agent interface. Most of the collections contain data that do not relate to all spatiotemporal objects on the Earth; different data collections often contain different sets of physical quantities. To find spatiotemporal objects and their meteorological and climatic parameters, it is required to create an expert system for selecting the necessary objects and their characteristics. The basis of this expert system is the knowledge base on the spatial objects of data collections and their parameters.

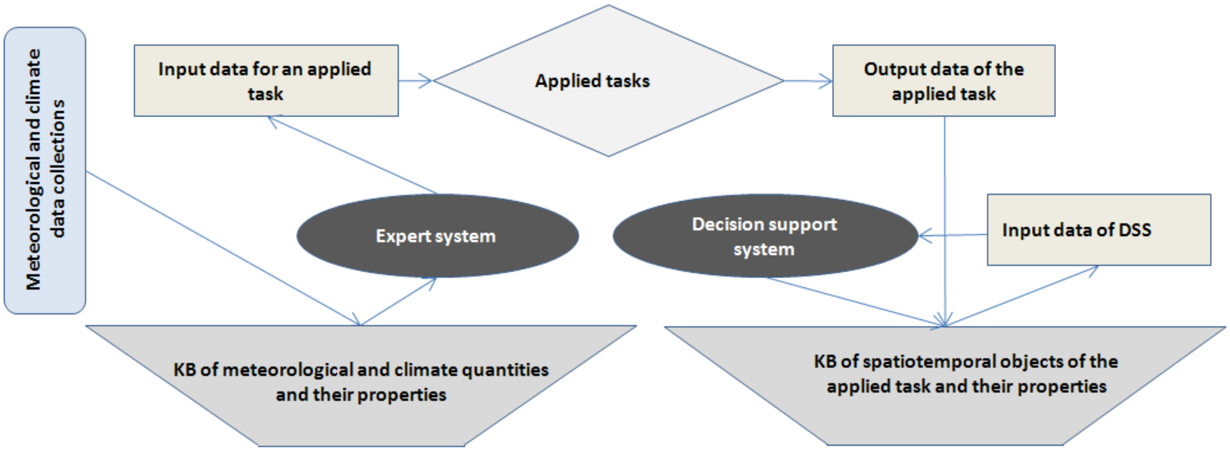

Fig. 2. Simplified block-diagram of "Climate+" platform modification. Data presentation services are omitted.

Figure 2 shows a simplified block-diagram which is a basis for 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.

In this work, we discuss questions of creation of a knowledge base (KB) for the expert system. The main problem which has been solved is the reduction problem[6] that is construction of typical individuals of an OWL-ontology that characterize properties of spatiotemporal objects from the collections. The next step is development of the conceptual part of the ontology (T- and R-box) which provides selection of input meteorological or climatic data with such details as required by the solved applied task. The selection (relevant collections) should be represented as a set of physical quantities from WMO classification.

4. DESCRIPTION OF METEOROLOGICAL PARAMETERS OF THE COLLECTIONS AND DATA OF APPLIED PROBLEMS AND ITS MATCHING WITH THE WMO CLASSIFICATION

When solving applied problems which use meteorological and climate data as an input, a "spatiotemporal object" is the key concept. Such objects can be represented by an atmospheric layer over a bounded territory, the upper soil layer related to the given territory, or, in more specific cases, forests, arable lands, or long roads. Physical and chemical processes occurring with spatiotemporal objects or their parts are described within mathematical models with different degrees of detail. Each model is presented as a set of computational tasks and to solve those one needs in relevant input data (values of physical parameters at given time points). When solving a complex problem, which consists of several subtasks, the matching of input and output data of all the subtasks is a separate technical problem.

Applied problems of the type considered (Fig. 2) are distinguished by a wide choice of different data sets granted by data providers. The main problem in the choice of input data is the correct matching of data required for an applied task with the intensions of data collections available. A typical solution to this problem is the use of expert definitions of physical parameters and their relationships. The WMO classification[7], which is currently used for representation of meteorological and climate resources in the GRIB format[1], is applied in this work as such implicit definitions. Examples of resources based on the WMO classification are WMO code register created for aviation in order to support the exchange of meteorological data in AvXML format[8] and GRIB Discipline Collection[9], which use RDF[10] and SKOS recommendations in the domain description.

The formalization of the relationships between properties of spatiotemporal objects in collections of meteorological data, input and output data of applied problems, and the classifier of types of physical parameters is performed in the OWL DL language[11].

The WMO classification is used in the primitive OWL-ontology CLIMATE+ of meteorological and climate information resources of the IMCES SB RAS and is the linguistic basis for the formation of the names of its classes, properties, and individuals. The individuals characterizing the collections of these resources are described in terms of the names of the physical parameters specified by their producers and are associated with the collections INMCM4.0 (predictive output data of the climate model of the general circulation of the atmosphere and the ocean[12]), ERA-Interim (global climate reanalysis data[13]), and PlaSim (predictive output data of the climatic model of medium complexity[14]). For most of these individuals, equivalence relationships are defined with individuals of the types of physical parameters according to the WMO classification.

5. CLIMATE+ ONTOLOGY

The Climate+ OWL-ontology of climate information resources consists of three independent Genesereth ontologies[15], and a Guarino ontology[16], which combines these independent ontologies.

The independent ontologies are:

- Multiple Data ontology of collections of numerical arrays of the Climate geoportal (http://climate.scert.ru/), which is one of Russian leading meteorological and climate resources;
- ontology of types of WMO physical parameters;
- ontology of applied tasks.

The names of the Climate+ classes are given in the 1st and 3rd rows of Table 1; subclasses of some classes, in the 2nd row. Properties of the Climate+ ontology are presented in Tabl. 2 and 3.

The Multiple Data ontology describes the intension of the base of metadata (see Section 2.1). It's main classes are: data collection (iao:Collection), data set (iao:Data_set), data array (iao:Data_array), spatiotemporal system (iao:Spatiotemporal_system), organization (iao:Organization), model scenario (iao:Scenario), spatial resolution (iao:Spatial_resolution), time step (iao:Time_step), physical quantity (iao:Physical_quantity), measurement unit (iao:Unit), physical data (iao:Physical_data), list of longitudes (iao:Longitudes_list), list of latitudes (iao:Latitudes_list), list of altitude levels (iao:Height_levels_list), list of time values (iao:Times_list).

A spatiotemporal object physically exists in three dimensions in a certain time interval. In the Multiple Data ontology, this object and the related physical properties are specified in the form of a spatiotemporal grid and are called a spatiotemporal system. A collection can include several spatiotemporal systems characterizing a spatiotemporal object. An individual that characterizes a spatiotemporal system belongs to the iao:Spatiotemporal_system class and is determined by individuals describing longitudes, latitudes, altitude levels, time intervals and physical parameters. These individuals belong to the iao:Longitudes_list, iao:Latitudes_list, iao:Altitude_levels_list, and iao:Times_list classes, which are subclasses of the iao:Physical_data class (physical data). Individuals that characterize physical quantities (iao:Physical_quantity class) in certain measurement units (iao:Unit) and are described by the number of values of the physical quantity (iao:has_number_of_values) in the array, uniform step of its values (iao:has_step_value), minimal (iao:has_minimum_value) and maximal values (iao:has_maximum_value) or by their numerical (iao:has_value) values belong to extensions of the iao:Physical_data class. The list of time labels (iao:Times_list) is additionally described by the initial (iao:has_initial_time) and final (iao:has_final_time) time labels.

The data array (iao:Data_array) is an ordered set of numerical values of a physical quantity (iao:Physical_quantity), which characterizes a spatiotemporal system (iao:has_spatiotemporal_system) at each 4D point (longitude, latitude, altitude, and time) of the spatiotemporal system. In terms of the OWL language, the data array (iao:Data_array) is a subclass of the physical data class (iao:Physical_data) and, thus, it is a numerical array of values of a physical quantity (iao:Physical_quantity) in certain measurement units (iao:Unit) and is described by the number of values of the physical quantity (iao:has_number_of_values) in the array, minimal (iao:has_minimum_value) and maximal values (iao:has_maximum_value) of the physical quantity. The set of data arrays (iao:Data_array) belongs (iao:has_data_array) to one data set (iao:Data_set), which differs from other data sets by a model scenario (iao:Scenario), spatial resolution (iao:Spatial_resolution), time step (iao:Time_step), and membership in one collection (iao:Collection).

Several types of physical quantities are distinguished at the upper level of the taxonomy of WMO physical quantities: hydrological, land surface (wmo:Land_surface_products), atmospheric (wmo:Meteorological_products), and space quantities. Numerous proper categories are distinguished in each of these areas. Figure 3 shows individuals of the ontology of the WMO physical quantities, which are representatives of the classes of physical quantities of the Soil (wmo:Soil_category) and Temperature (wmo:Temperature_category) categories. Ontology individuals wmo:Soil_temperature and wmo:Soil_moisture are representatives of the wmo:Soil_category class, and the individual wmo:Skin_temperature is a representative of the wmo:Temperature_category class.

Ontology descriptions of physical problems are based on the representation of a scientific task (tsu:Task) in the form of input (tsu:hasInputData) and output data (tsu:hasOutputData). The scientific task (tsu: Task) can have references to mathematical (tsu:hasMathematicalStatement) and physical (tsu:hasPhysicalStatement) models.

At the bottom of Fig. 3, an individual of the applied task is shown. This problem deals with a soil layer of the depth L with constant moisture. The soil starts freezing gradually, and the depth of the frozen layer changes with time. Thus, the soil layer under study is divided into frozen and thawed zones, with a boundary in between. Under the assumption that the heat is transferred in the soil only due to the heat conduction, the task is reduced to a two-phase one-dimensional

Stefan’s problem with a second-kind boundary condition at the lower boundary. The input data for this task are the temperatures on the soil surface and at the depth L, soil moisture data, the coefficient of thermal conductivity, specific heat, and densities of thawed and frozen soil. The output data are a 1D temperature field for the soil layer under study and the depth of the phase boundary[17].

The ontology of this applied task includes individuals of input and output data belonging to the extensions of the tsu:InputData and tsu:OutputData classes that are subclasses of the data class tsu:Data. These individuals are associated (via specific properties Datatype Properties) with individuals that characterize meteorological variables (tsu:hasMeteorologicalVariable), subject variables (tsu:hasSubjectVariable), and a spatiotemporal object (tsu:hasSpatioTemporalObject), which is described by these variables. An individual describing a spatiotempotal object (tsu:SpatioTemporalObject class) physically exists in three dimensions throughout a certain time interval.

In this work, we describe task T1 of soil freezing and thawing (tsu:Task_T1_–_freezing_and_thawing_of_soil). Task T1 has input (tsu:Inputdata_for_task_T1) and output (tsu:Outputdata_for_task_T1) data. The input data include (tsu:Inputdata_for_task_T1) a spatiotemporal object of task T1 (tsu:Spatiotemporal_object_for_task_T1), subject variables (tsu:Specific_heat_capacity_of_thawed_ground, tsu:Thickness_of_soil_layer, tsu:Density_of_thawed _soil, tsu:Specific_heat_capacity_of_frozen_soil, tsu:Freezing_point_of_soil, tsu:Heat_of_phase_transition, tsu:Coefficient_of_thermal_conductivity_of_thawed_ground, tsu:Density_of_frozen_soil, tsu:Coefficient_of_ thermal_conductivity_of_frozen_soil), and meteorological variables (tsu:Soil_moisture, tsu:Skin_temperature, tsu:Initial_temperarure_ of_soil). The output data of task T1 (tsu:Outputdata_for_task_T1) include meteorological variables (tsu:Temperature_of_the_thawed_soil_layer, tsu:Temperature_of_the_frozen_soil_layer, tsu:Boundary_of_the_phase_transition).

Table 1. Classes of Climate+ ontology

Class	SubClassOf	Class
iao:Collection		iao:Organization
iao:Scenario		iao:Data_set
iao:Spatial_resolution		iao:Unit
iao:Time_step		iao:Physical_data
iao:Physical_quantity		iao:Spatiotemporal_system
iao:Data_array	iao:Physical_data	
iao:Longitudes_list	iao:Physical_data	
iao:Latitudes_list	iao:Physical_data	
iao:Height_levels_list	iao:Physical_data	
iao:Times_list	iao:Physical_data	
wmo:Products		
wmo:Land_surface_products	wmo:Products	
wmo:Soil_category	wmo:Land_surface_products	
wmo:Meteorological_products	wmo:Products	
wmo:Temperature_category	wmo:Meteorological_products	
tsu:Task		
tsu:Data		
tsu:InputData	tsu:Data	
tsu:OutputData	tsu:Data	
tsu:SpatioTemporalObject		
tsu:MeteorologicalVariable		
tsu:SubjectVariable		

In the ontology, combining the three independent ontologies described above, individuals of the independent ontologies are matched. Figure 3 shows equivalence relations (owl:sameAs) between individuals of these ontologies. To test the unifying ontology for consistency, it is suggested to use an output machine. An output machine used in the descriptive logic checks the resulting unifying ontology for conformity and consistency, supplementing it with facts and statements derived from existing models. The following five facts in the form of triads are shown in the unifying ontology in Fig. 3:

- iao:Physical_quantity_Soil_temperature owl:sameAs wmo:Soil_temperature;
- tsu:Initial_temperarure_of_soil owl:sameAs wmo:Soil_temperature;
- tsu:Soil_moisture owl:sameAs wmo:Soil_moisture;
- tsu:Skin_temperature owl:sameAs wmo:Skin_temperature;
- tsu:Spatiotemporal_object_for_task_T1 owl:sameAs iao:Spatiotemporal_system_No_2.

This ontology allows one to systematize and obtain knowledge on the whole volume of data represented by all ontological models participating in the integration. Thus, it is possible to satisfy complex information requests taking into account the integration facts, which could not be possible without their presence. For example, one can ascertain through the integration ontology which data arrays (iao:Data_array) can be used as input to solve the problem of soil freezing and thawing (tsu:Task_T1_–_ freezing_and_thawing_of_soil).

The definition of subject properties is presented in the first three columns of Table 2; the fourth column gives unique property identifiers. For each property from the second column, the definition area (the first column) and the range of values (the third column) are specified.

Table 2.Object properties of the ontology of climate information resources.

Domain	Object Property	Range	
iao:Collection	iao:has_organization	iao:Organization	o01
iao:Collection	iao:has_data_set	iao:Data_set	o02
iao:Data_set	iao:has_scenario	iao:Scenario	o03
iao:Data_set	iao:has_spatial_resolution	iao:Spatial_resolution	o04
iao:Data_set	iao:has_time_step	iao:Time_step	o05
iao:Data_set	iao:has_data_array	iao:Data_array	o06
iao:Physical_data	iao:has_physical_quantity	iao:Physical_quantity	o07
iao:Physical_data	iao:has_unit	iao:Unit	o08
iao:Data_array	iao:has_spatiotemporal_system	iao:Spatiotemporal_system	o09
iao:Spatiotemporal_system	iao:has_longitudes_list	iao:Longitudes_list	o10
iao:Spatiotemporal_system	iao:has_latitudes_list	iao:Latitudes_list	o11
iao:Spatiotemporal_system	iao:has_height_levels_list	iao:Height_levels_list	o12
iao:Spatiotemporal_system	iao:hat_times_list	iao:Times_list	o13
tsu:Task	tsu:hasInputData	tsu:InputData	o14
tsu:Task	tsu:hasOutputData	tsu:OutputData	o15
tsu:Data	tsu:hasMeteorologicalVariable	tsu:MeteorologicalVariable	o16
tsu:Data	tsu:hasSubjectVariable	tsu:SubjectVariable	o17
tsu:Data	tsu:hasSpatioTemporalObject	tsu:SpatioTemporalObject	o18

The definition of data type properties is given in the first three columns of Table 3; and unique property identifiers, in the fourth column. For each property in the second column, the domain (the first column) and the range (the third column) are specified.

Table 3. Data type properties in the ontology of climate information resources.

Domain	Datatype Property	Range	
iao:Physical_data	iao:has_number_of_values	int	d01
iao:Physical_data	iao:has_minimum_value	float	d02
iao:Physical_data	iao:has_maximum_value	float	d03
iao:Physical_data	iao:has_value	float	d04
iao:Physical_data	iao:has_step_value	str	d05
iao:Times_list	iao:has_time_start	str	d06
iao:Times_list	iao:has_time_end	str	d07
tsu:Task	tsu:hasMathematicalStatement	anyURI	d08
tsu:Task	tsu:hasPhysicalStatement	anyURI	d09

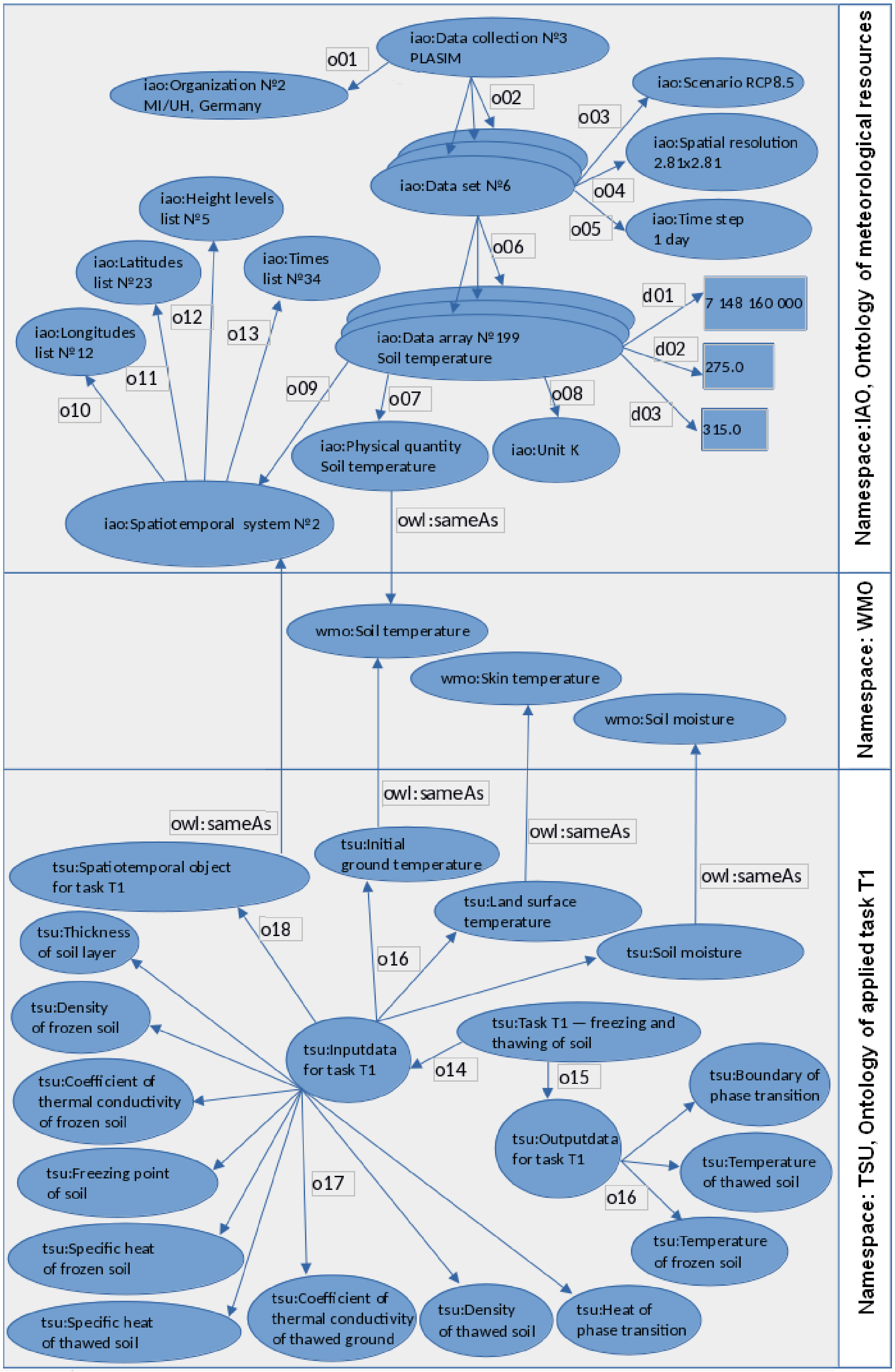

Fig. 3. Simplified individuals that describe a PLASIM data collection; some individuals of the taxonomy of WMO physical parameters, and an individual of T1 physical problem (soil freezing and thawing).

Figure 3 presents the structure of one individual of the Climate+ ontology of climate information resources in description of the PLASIM data collection and physical task T1 (soil freezing and thawing), as well as their relationships with the physical quantities of the WMO taxonomy through equivalence relations.

The ontology individuals are shown in ovals, and literal values, in fill rectangles; the arrows represent properties with unique identifiers in small empty rectangles from the Tables 2 and 3. In the RDF graph[10] shown on Fig. 3, every two individuals connected by a directed arc represent a "Subject-Predicate-Object" triad. Triple arrows indicate possible cardinality of a property higher than unit, and three overlapped ovals mean a possible number of individuals greater than one.

6. SUMMARY

The Climate+ ontology of climate information resources of the IMCES SB RAS is developed and formalized using the OWL DL language. It describes the current state of collections of numerical arrays of climate and meteorological data of the "Climate" geoportal (http://climate.scert.ru/), the taxonomy of WMO physical quantities, and the applied ontology of the problem of soil freezing and thawing (Stefan's problem). A-, R- and T-box of the Climate + ontology are created for the work of the expert system that determines necessary input data for applied tasks. The ontology of climate information resources includes a description of 170 spatiotemporal objects characterized by 156 physical quantities. Software has been developed for the formation of ontology individuals of collections of climate information resources. The individuals formed characterize 13 collections, which contain 36 data sets and 793 numerical data arrays of more than 80 Tb in volume.

ACKNOWLEDGMENTS

The authors thank the Russian Science Foundation for the support of this work under the grant No16-19-10257.

REFERENCES

[1] World Meteorological Organization Extranet, “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,” , 2017, http://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB2_062006.pdf

[2] Unidata UCAR, “Network Common Data Form (NetCDF),” , 2017, https://www.unidata.ucar.edu/software/netcdf/

[3] HDF Group - HDF5, “Current Releases: HDF5-1.8.19, HDF5-1.10.1,” ,2017, https://support.hdfgroup.org/HDF5/

[4] 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 Novosibirsk State University. Series: Information Technology, 9(4), 94-102 (2011).

[5] Titov, A.G., Gordov, E.P., Okladnikov, I.G. “Hardware-software platform Climate as a basis for local spatial data infrastructure geoportal,” Vestnik Novosibirsk State University. Series: Information Technology”, 10(4), 104-111 (2012).

[6] Zinov'ev, A.A., Foundations of the logical theory of scientific knowledge (Complex Logic), D. Reidel Publishing Company, Dordrecht-Holland, 1-304 (1973).

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

[8] WMO Codes Registry, “GRIB edition 2,” , 2017, https://codes.wmo.int/

[9] Science and Technology Facilities Council, “The GRIB Discipline Collection,” , 2017, http://vocab-test.ceda.ac.uk/collection/grib/grib-content/Discipline.html

[10] Graham Klyne, Jeremy J. Carroll, “Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation 10 February 2004,” ,2017, http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/

[11] M. Schneider, F.J. Carroll, I. Herman, P.F. Patel-Schneider , “OWL 2 Web Ontology Language. RDF-Based Semantics (Second Edition),” , 2017, http://www.w3.org/TR/2012/ REC-owl2-rdf-based-semantics-20121211/.

[12] Volodin, E.M., Dianskii, N.A., Gusev, A.V., “Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations,” Izvestiya. Atmospheric and Oceanic Physics, 46(4), 414-431 (2010).

[13] ECMWF, “Details of ERA-Interim parameters,” , 2017, https://software.ecmwf.int/wiki/display/CKB/Details+of+ERA-Interim+parameters

[14] Universitat Hamburg, “The Planet Simulator (PlaSim): a climate model of intermediate complexity for Earth, Mars and other planets,” , 2017, https://www.mi.uni-hamburg.de/en/arbeitsgruppen/theoretische-meteorologie/modelle/plasim.html

[15] Genesereth, M.R. and Nilsson, N.J., Logical foundation of Artificial Intelligence, Morgan Kaufman, Los Altos, 1-200, 1987.

[16] Oberle, D., Semantic management of middleware, Springer, Berlin, 1-268 (2006).

[17] Alipova, K.A., Bogoslovskiy, N.N., “Stefan’s problem for heat transfer equation,” Proc. All-Russian Youth Symposium “All sides of a mathematic and mechanic”, P. 92-99 (2016).