

RESEARCH PAPER

Virtual Research Environment for Regional Climatic Processes Analysis: Ontological Approach to Spatial Data Systematization

Andrey Bart¹, Alexander Fazliev², Evgeny Gordov^{1,2,3}, Igor Okladnikov^{2,3}, Alexey Privezentsev² and Alexander Titov^{2,3}

¹ Tomsk State University, Tomsk, RU

² Institute of Atmospheric Optics, Siberian Branch of Russian Academy of Sciences, Tomsk, RU

³ Institute of Monitoring of Climatic and Ecological Systems Siberian Branch of Russian Academy of Sciences, Tomsk, RU

Corresponding author: Alexander Fazliev (faz@iao.ru)

This paper describes a Virtual Research Environment (VRE) based on a web GIS platform 'Climate+', which provides an access to analytic instruments processing 19 collections of meteorological and climate data of several international organizations. This environment provides systematization of spatial data and related climate information and allows a user getting analysis results using geoinformation technologies. The ontology approach to this systematization is described, making it possible to match semantics of meteorological and climate parameters presented in different collections and used in solving various applied problems.

Keywords: Geoinformation; Ontology; Climate Information; Spatial and Climate Data; Virtual Research Environment

Introduction

On-going climate changes, especially their extreme manifestations, such as heat waves, cold periods, heavy rains or snowfalls, storms, floods or droughts, have an increasing impact on economic, political and social processes (IPCC 2012; IPCC 2013; Sillmann et al. 2014). Reliable assessments of their trends and impacts on these processes are critical for the development of adequate local and/or regional strategies for adapting and mitigating negative effects of climate change, for example, for sustainable agriculture and forestry, or planned infrastructure. But such assessments are still missing for various parts of the world. This circumstance is an essential driver for the development of climatic characteristics' monitoring and climate modeling to assess possible future trends. Local and remote observations, as well as numerical modeling of climatic processes, resulted in an unprecedented growth of data archives. For example, the Copernicus programme, the European Union's flagship programme on monitoring the Earth's environment using satellite and *in-situ* observations, anticipates a massive increase in satellite data volume. It is estimated that solely the Sentinel missions, Copernicus' space component, will produce 4 TB of processed data each day (Copernicus 2016). The volume of climate simulation experiments, which formed the basis of the 5th Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), reaches 2 PB (WCRP 2017). The European Centre for Medium-Range Weather Forecasts (ECMWF) archive of meteorological data currently holds more than 90 PB of data and continues to grow by additional 3 PB every month.

The advent of the big data era forced the experts to think about the development of tools for transformation of raw data into information and knowledge (Rowley 2007). At present, much more efforts are spent on data managing and pre-processing than those devoted to actual data evaluation. Such an increase in data archive volume makes using the traditional approach to climate information analysis doubtful, and requires new approaches based on distributed networks and usage of modern information technologies. The fact that emergence of large data archives would create problems for the analysis of current and future

environmental processes was apparent by the end of the last century. As a result, several large programs have been launched in the world to solve them. For example, in the USA, efforts in this direction are united by the general concept of Cyberinfrastructure, which, according to National Science Foundation (NSF), is an information-computational infrastructure that supports distributed science (e-Science), including data, people, computers, and exploits Internet technology (Web 2.0, Grid, etc).

Currently, the main efforts and resources in the world are focused on creating a sustainable distributed cyberinfrastructure for open, permanent, reliable and secure access to high quality Earth observation data and corresponding metadata. Data Observation Network for Earth is one of such projects (<https://www.dataone.org/>). It is the basis for combining multiple federal network repositories to provide search, retrieval and uniform access to the data. Within the framework of international cooperation Earth System Grid Federation (ESGF) portals have been created to integrate geographically distributed scientific datasets. As a part of this collaboration, Earth System Grid (ESG) is developed to facilitate the analysis of global climate change. It provides access to climate projection data. In particular, the historical climatic data and modeling results for climate scenarios, drawn up while preparing the recent IPCC report, were disseminated through ESGF. Currently, about 2 PB of data are stored in ESGF nodes distributed around the world. The next step required is to convert generic climate analysis tools into services (Schnase et al. 2014). For example, the Climate Analytics-As-a-Service (CAaaS) approach combines high-performance computing and data analytics with scalable data management, cloud virtualization, adaptive analytics, and Application Programming Interfaces (APIs) related to subject areas to improve access to large climate data collections. Several years ago World Meteorological Organization (WMO) initiated a large program for development of the climate service system based on thematic web services (<http://gfcs-climate.org/>), intended for Supporting Climate Change Adaptation (WMO 2016).

According to Candela, Castelli & Pagano (2013), all these efforts can be described as a development for climate domain of a Virtual Research Environment (VRE) which is a system with following major features: (i) it is a web-based working environment; (ii) it is tailored to serve the needs of a targeted community; (iii) it is expected to provide a community with the whole array of products needed to accomplish the community's goal(s); and (iv) it promotes sharing of research results.

One of the leading to VRE development directions that emerged as a response of the challenge of big data and their intensive use, was the creation of thematic computing systems accessed via web portals (Gordov & Fazliev 2004). The main idea of this approach is to develop a system that comprises data, tools for processing and visualization of results, computing resources to be used by researchers to perform selected data analysis, and to provide Internet access to the system using a unified user-friendly interface. An example for atmosphere domain is the web portal ATMOS (Gordov, Lykosov & Fazliev 2006). The fact that climate science deals with georeferenced data requires usage of a combination of web and GIS techniques. As a response, the web GIS platform 'Climate' (<http://climate.scert.ru/>) was developed (Gordov et al. 2016; Okladnikov et al. 2015). It is a free, cross-platform, composite application with an open-source computational backend that complies with common data format conventions for climate data and provides functionality of common desktop software in a window of an Internet browser. Forming thematic VRE web GIS 'Climate' is providing interdisciplinary distributed research groups of non-experts in information technologies (climatologists, ecologists, biologists, and decision makers) with easy-accessible reliable online tools for rapid analysis and visualization of multidimensional heterogeneous climatological datasets obtained from various sources.

However, the reliable analysis of climate changes, and nature and society's responses them require skills in dealing with big datasets, abilities to interact with powerful computing resources and complex numerical models, knowledge of modern methods of statistical analysis, and usage of high-level programming languages. The skills mentioned are not typical for specialists in the field of economic, political and social sciences, and unfortunately, it is completely uncharacteristic for decision-makers. It should be added that each thematic domain has its own terminology, sets of data with their standards and models (Lutz et al. 2009). Therefore, an integration of all those in the Internet-accessible research environment to provide specialists and decision-makers with reliable tools for studying economic, political and social consequences of climate change should be done. Moreover, to develop scientifically based foundation to climate change adaptation/mitigation practical measures, one must deal with three types of domains, which are climatology, applied domains and related decision support systems (DSS).

General approach to this type of transdisciplinary problem should rely upon a VRE with three layers of data processing services, which are Data, Information and Knowledge (De Roure, Jennings & Shadbolt 2001). Here Data is understood as uninterpreted bits and bytes of raw data, Information is data equipped with meaning and Knowledge means information applied to achieve a goal, solve a problem or enact a decision.

We refer to desired result of such integration as a web-GIS platform 'Climate+'. The 'Climate+' platform should integrate three sets of thematic domains, which are climatic and meteorological processes, applied impact problems using calculated characteristics of climatic and meteorological processes and decision support systems using applied problem solutions. The important prerequisite required for such integration is an ontology description, since each domain type has its own knowledge representation and integration of those might be controlled on the base of their ontology description. Role and necessity of ontologies usage in geophysical sciences have been demonstrated in the papers of Athanasis et al. (2009), Bogdanović, Stanimirović & Stoimenov (2015), Brodaric, Fox & McGuinness (2009), Husain et al. (2011) and Lutz et al. (2009).

In this paper we describe first steps in development of the 'Climate+' platform, namely applications and ontologies that refer to three layers (Data, Metadata (Information) and Ontology (Knowledge)) of data processing services of the VRE under development. Since the 'Climate' platform forms the backbone of the 'Climate+', firstly we present it in more detail and describe its architecture, major components, and workflow. Then we describe the Ontology layer and developed 'Climate+' platform Applied Ontologies. Special attention will be paid to semantic heterogeneity, solving and formalization of thematic domains related with applied tasks, in particular, a solution of a reduction problem. In the conclusion next steps of the platform development are discussed.

Platform Climate

The web GIS platform 'Climate' developed at the IMCES SB RAS (Okladnikov et al. 2015; Gordov et al. 2016) is aimed at processing and analysis of geospatial gridded datasets in Earth system science, and visualization of results. It's based on a dedicated software framework consisting of three key components: a server-side computational backend; a server-side middleware represented by geoportal; and specialized JavaScript library containing typical widgets of web mapping client GUI which is based on AJAX technology. Geospatial datasets are processed by a set of validated software modules running by the backend. Results are represented by overlapped raster and vector cartographical layers accompanied by corresponding binary data. The platform 'Climate' functionality includes basic and complex statistical analysis of data, whilst online geo-information system (GIS) instruments give a user an ability to combine and map georeferenced results over a chosen cartographical basis. It provides specialists and users without programming skills with reliable and practical online instruments for integrated research of climate and ecosystems changes through a unified web interface.

Some examples of successful applications of the web GIS platform 'Climate' in studies of ongoing Siberia climate change and its impact can be found in the papers of Riazanova et al. (2016), Ryzanova & Voropay (2017) and Shulgina, Gordov & Genina (2011).

An architecture

The 'Climate' platform architecture is shown on **Figure 1**. It represents a typical client-server structure, where in general case the server might be a set of geographically distributed standalone nodes providing common (federated) interface (API), and client applications (basically, Web-GIS client).

The server part of the architecture includes a high-performance computing system with a data storage attached. It is presented by two tiers:

- resources tier, including data and metadata;
- server applications (middleware) tier.

The client part of the architecture is based on modern graphical web browser. It is presented by a single 'Client applications' tier, respectively.

The 'Resources' tier of the platform 'Climate' employs two basic layers, which are Data and Metadata layers. The data layer contains datasets, located on the data storage system either in the form of collections of netCDF files or PostGIS databases. Metadata layer is presented by the Metadata database (MDDDB), which describes geospatial datasets and their processing routines, and provides effective system functioning (Okladnikov, Gordov & Titov 2016). The database contains structured spatial and temporal characteristics of available geospatial datasets, their locations, and configurations of software components for data analysis. Its structure has the 'Dataset' and 'Dataset collection' levels. A dataset is defined as a set of data which is a) given on a single temporal and spatial grid; b) covers the same time range; and c) obtained under the same scenario of simulation. According to the chosen data storage model (Okladnikov, Gordov & Titov 2016), spatial datasets are mostly represented by collections of netCDF files grouped by spatio-temporal features and

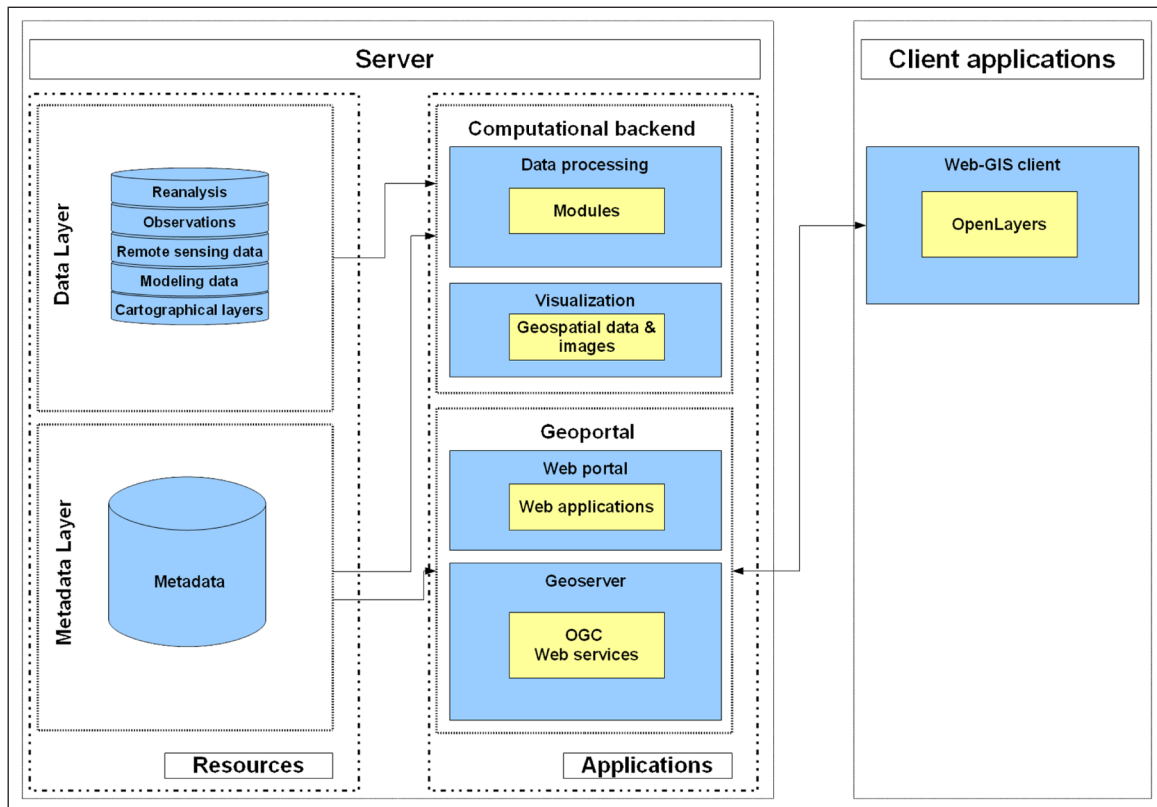


Figure 1: Platform 'Climate' general architecture outline.

placed in the hierarchy of directories on data storage systems. Each netCDF file stores one or more variables containing values of meteorological parameters on a given spatio-temporal domain. The files in a dataset are usually named according to the same pattern to provide their automatic search. Along with data variables, netCDF files contain horizontal, vertical and time domain grids. Dataset collection is defined as a collection of datasets created by an organization in the framework of the same research project but specified on different spatial and/or temporal grids, or for different scenarios. The collection may consist of one dataset.

There are two major parts of MDDB. The first part contains description of all available for analysis datasets: spatio-temporal domain, lists of meteorological parameters and locations of data files on storage system. It is used to search data files and to provide metadata on requests from the backend. The second part contains description of processing routines represented by various pipelined call sequences of dedicated computational modules and their configuration options. Some data analysis routines might be applied to specific meteorological parameters only. Thus the connections between computing modules and data arrays are set in the MDDB. Some tables in the MDDB contain multi-lingual descriptions of datasets and processing routines in a human-readable form, and are used for filling elements of the graphical user interface. This provides rapid actualization of available datasets, parameters and processing routines lists in the graphical user interface just after their integration into the MDDB.

To maintain correctness and synchronization of the MDDB, as well as to provide means for its filling and editing, a dedicated administrative tool (web console) was developed and integrated into the system. Its graphical user interface is shown on **Figure 2**. The administrative web console allows one to create, display, edit and delete information in MDDB with database integrity support. It comprises a set of scenarios written on PHP, JavaScript, and HTML on the base of JavaScript libraries jQuery and jqGrid, which provides user with an intuitive interface. The interface also support API required to process Web client requests on the server side.

The 'Server applications' middleware tier consists of two basic software components: computational backend and geoportal.

Computational backend

The computational backend contains data processing and visualization software components. The data processing is a key software component containing computational modules based on GNU Data Language (GDL, <http://gnudatalanguage.sourceforge.net/>) and Python and providing integral geospatial data

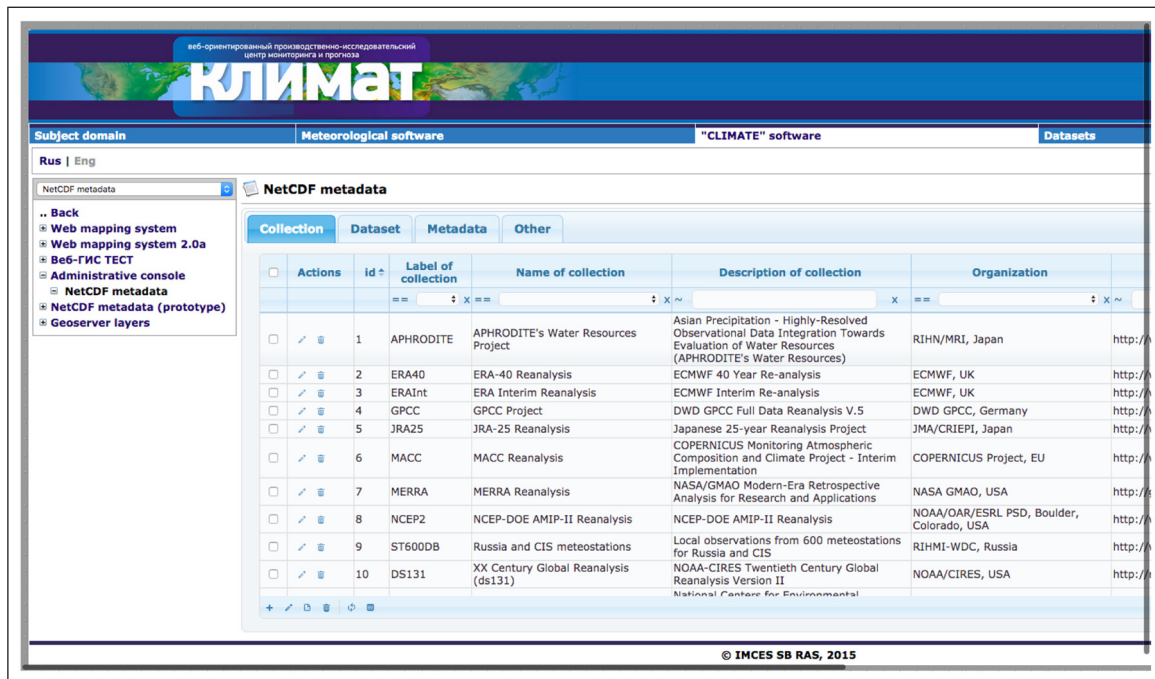


Figure 2: Administrative web console GUI.

statistical processing as well as API to work with netCDF, Hierarchical Data Form (HDF), ESRI Shapefile data files and PostGIS databases. Depending on the result type required visualization component of the backend generates files in the following formats: GeoTIFF, ESRI Shapefile, Encapsulated PostScript, CSV, XML, netCDF, float GeoTIFF.

Geoportal

Spatial Data Infrastructure (SDI) geoportal contains two basic components: web portal and Geoserver (<http://geoserver.org>). Geoserver provides cartographical web services such as Web Mapping Service (WMS), Web Feature Service (WFS) and Web Processing Service (WPS). In general, the Web processing service provides standard HTTP interface for remote configuring and launching data processing software modules and presenting results in generic formats. The services can be used by either standard GIS environments or web applications.

The web portal serves as a connection point between different SDI elements (geospatial data, metadata, services and client applications). Its main feature is providing unified API for client web applications which comply with the conventional Boundless/OpenGeo architecture (Becirspahic & Karabegovic 2015). The web portal provides server-side part of the Web-GIS client application which complies with general INSPIRE (INfrastructure for SPatial InfoRmation in Europe, <https://inspire.ec.europa.eu>) requirements to geospatial data visualization and implements computational processing services launching to support solving tasks in climate monitoring.

Platform Climate+

The 'Climate' platform described above is aimed at area of basic and applied climatology. Since adaptation to climate change and mitigation of its negative consequences nowadays is urgently required, the thematic VRE should deal with applied tasks of different domains in which climate change impact should be considered. To meet these challenges and to make a step from climate science to climate services we develop the 'Climate+' platform whose architecture is shown on **Figure 3**. The 'Climate' platform is working with resources related only to the climatic data and metadata layers. Subsequent usage and re-usage of results of data processing by procedures of the Computational backend is a concern of a user only. However, while dealing with applied regional problems caused by climate change, a special role is assigned to the decision-support system (DSS), which relies upon solutions of computational problems (Applied Tasks) describing changes of states of spatio-temporal objects (rivers, lakes, roads, etc.) at different time intervals. Resulting information resources (results of calculations) should be saved since they would be requested in different decision-making tasks. In the 'Climate+' platform the results of the calculations performed in the Applied Tasks box are added to the data and metadata layers and can be used later.

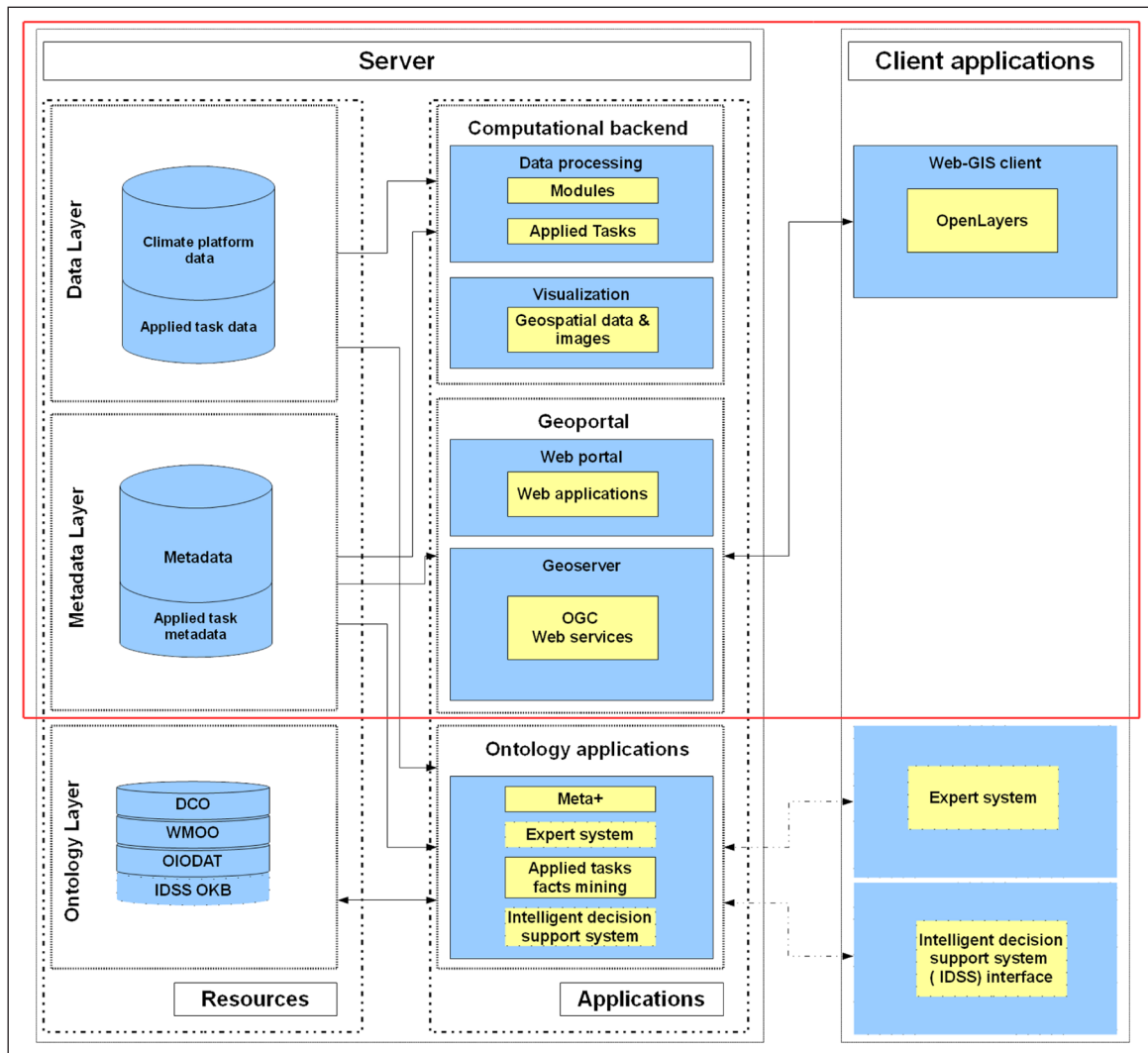


Figure 3: Platform 'Climate+' general architecture outline.

The variety of applied computational problems is significant, and the input data required for their solving is related to data from different subject domains. Therefore, the description of intensions of input data for computational applications and data collections is semantically heterogeneous, that could lead to inadequate decisions related to the interpretation, integration, exchange of data, as well as in retrieving and usage of relevant information. Overcoming semantic heterogeneity in spatial data infrastructures was described in (Lutz et al. 2009). To solve the problem of semantic heterogeneity in the 'Climate+' platform, an ontology layer characterizing the properties of the data collections (Reanalysis, Observations, Modeling Data) is created. This ontology is used to select input data for Applied Tasks applications.

Formalized 156 climatic and meteorological characteristics currently presented in the collections are named in accordance with the WMO taxonomy (WMO 2016). They form a common shared vocabulary of the domain and are presented in the 'Climate+' platform in the form of OWL-ontology described in the next section. At present the contradictions in the definitions of physical quantities in the platform are solved by an expert. The 'Meta+' application is used to automatically create individuals of the ontology of the climate and meteorological data collections (Data collection ontology, DCO). The Expert System application allows one to match the applied problems input data intensions with the corresponding collections of climatic data. The set of applications Applied Tasks Facts Mining creates ontology individuals for each specific application. Examples of such individuals for the ontology of climate data collections properties and the ontology of the input data of the problem of freezing and thawing of a river are given in the next section.

Formalizing the properties of data and metadata and their representation in the form of ontologies in the 'Climate+' platform is useful not only for achieving semantic homogeneity, but also for building a knowledge base in the subject domains for decision support tasks. This comes from the fact that in the decision-making

system, knowledge representations are used from a significant number of subject areas in which modeling is performed with different degrees of granulation. Without an explicit formal representation of terms, statements and concepts, within the framework of one specification language (in our work of the OWL 2 QL language), the interpretation of the adopted decision may be incorrect.

Ontology layer of Climate+ platform

The ontology layer is a part of the knowledge layer and is used to solve the following tasks:

- Semantic search for collections of meteorological and climatic resources, properties of applied problems solutions and decisions made available in the 'Climate+' platform;
- Detection of contradictions between definitions of physical quantities in data collections and their matching in collections and input data of applied problems;
- Formalization of the results of applied problems solutions;
- Building ontological knowledge bases for IDSS.

The ontological layer contains three groups of OWL-ontologies. The first group includes the ontology of the collections of climatic and meteorological data (DCO) (Alipova et al. 2017). To construct it we created the ontology of climatic and meteorological characteristics (WMOO) (Bart, Privezentsev & Fazliev 2017). Key individuals of DC ontology describe the data collection properties and are intended for selection by user input data for its applied tasks. The second group is applied ontologies of input and output applied tasks data (OIODAT) (Bart et al. 2017). Solutions of applied tasks can be used in intellectual decision support system (IDSS). The third group of ontologies is an ontological knowledge base (IDSS OKB) (Kaklauskas 2015). The ontologies of the first two groups are described in more detail below. The third group of ontologies is at process of development now and is not discussed in detail in this paper.

The ontologies are associated with the applications shown in **Figure 3**. The 'Meta+' application supports automatic creation of individuals of the climate and meteorology data collections ontology. The 'Meta+' application is written in Python 3 with the use of the Owlready library (<https://pypi.python.org/pypi/Owlready>). The 'Climate+' platform is aimed at presenting data using GIS technologies (human interface). Its further development is aimed at enabling researchers to select and use sets of climate or meteorological data 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 application '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.

The third group of applications, 'Applied Tasks Facts Mining', is designed to extract facts from numerical solutions of specific applied problems and to present them in the form of individuals of the corresponding OWL ontologies. Below the individuals characterizing such solutions are given for the problems of describing the freezing of a river as an example. Demonstration case study task 'The decision support task' determines the timeframe for closing the navigation on the Ob river and opening the ice crossings on the same river.

Climate+ applied ontologies

Key elements of OWL-ontologies are classes, properties, and individuals. The individuals are defined during the solution of a reduction problem, as well as most properties. Some of the ontology classes are used for specification of the ontology properties domains and ranges.

Three applied ontologies are described below in some details:

- meteorological and climate information ontology of the 'Climate+' platform's web portal (iao:);
- WMO taxonomy of physical quantities (wmo:);
- ontology for description of web portal applied tasks (tsu:).

The classes and properties of these applied OWL-ontologies are listed in the **Tables 1–3**.

Table 1 presents the main classes of the ontologies (Alipova et al. 2017; Bart et al. 2017; Bart, Privezentsev & Fazliev 2017). The namespace prefix is placed before the class name; the class name abbreviation used in the text below is given after the class name in parentheses.

Table 1: Classes of applied ontologies.

Class	Class	SubClassOf
iao:Collection (C)	iao:Data_array (DA)	iao:Physical_data
iao:Organization (O)	iao:Longitudes_list (Lol)	iao:Physical_data
iao:Data_set (DS)	iao:Latitudes_list (Lal)	iao:Physical_data
iao:Scenario (S)	iao:Height_levels_list (HLL)	iao:Physical_data
iao:Spatial_resolution (SR)	iao:Times_list (Tl)	iao:Physical_data
iao:Time_step (Ts)	iao:Spatiotemporal_system (SS)	
iao:Physical_quantity (PQ)	iao:Unit (U)	
iao:Physical_data (PD)		
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:InputData	tsu:Data
tsu:Data	tsu:OutputData	tsu:Data
tsu:MeteorologicalParameter	tsu:SubjectParameter	
tsu: CriticalTimeInterval	tsu:SpatioTemporalObject	tsu:RiverStretch

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

Domain	ObjectProperty	Range	
iao:Collection	iao:has_organization (do)	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 (dpq)	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:has_times_list	iao:Times_list	o13
tsu:Task	tsu:hasInputData	tsu:InputData	o14
tsu:Task	tsu:hasOutputData	tsu:OutputData	o15
tsu:SpatioTemporalObject	tsu:hasMeteorologicalParameter	tsu:MeteorologicalParameter	o16
tsu:SpatioTemporalObject	tsu:hasSubjectParameter	tsu:SubjectParameter	o17
tsu:Data	tsu:hasSpatioTemporalObject	tsu:SpatioTemporalObject	o18
tsu:RiverStretch	tsu:hasCriticalTimeInterval-Shipping	tsu: CriticalTimeInterval	o19
tsu:RiverStretch	tsu:has CriticalTimeInterval-Crossing	tsu: CriticalTimeInterval	o20

Table 3: Datatype property of the ontology of climate information resources.

Domain	DatatypeProperty	Range	
iao:Physical_data	iao:has_number_of_values (dn)	int	d01
iao:Physical_data	iao:has_minimum_value (dmiv)	float	d02
iao:Physical_data	iao:has_maximum_value (dmav)	float	d03
iao:Physical_data	iao:has_value (dv)	float	d04
iao:Physical_data	iao:has_step_value	str	d05
iao:Times_list	iao:has_initial_time (dit)	str	d06
iao:Times_list	iao:has_final_time (dft)	str	d07
tsu:Task	tsu:hasMathematicalStatement	anyURI	d08
tsu:Task	tsu:hasPhysicalStatement	anyURI	d09

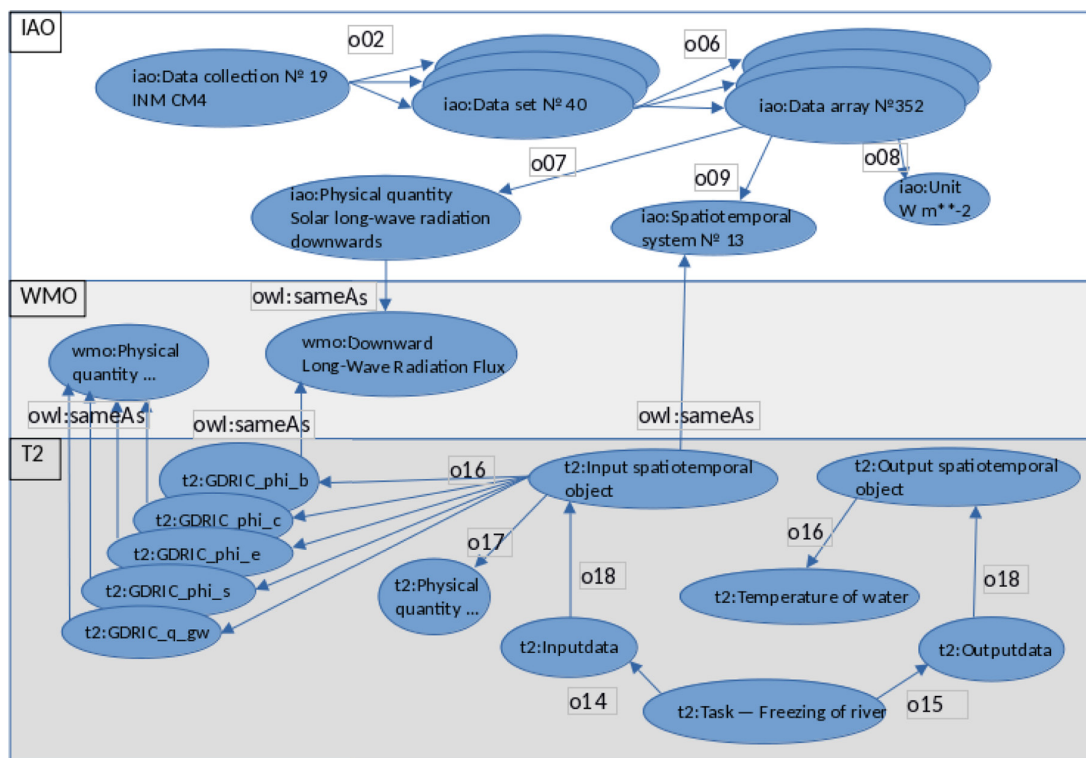


Figure 4: Simplified representation of individuals, describing PLASIM and INM CM4 data collections, taxonomy of WMO physical quantities, and individuals of T1 (soil freezing and thawing) and T2 (freeze-up on a river) applied tasks.

Properties of the applied ontologies are presented in **Tables 2** and **3**. The first three columns contain the domain, object property and range, respectively; the fourth column is the property abbreviation, used in the text and in **Figure 4** (Bart et al. 2017).

Ontology of climate and meteorological information collections. Classes and properties

Collections of climate and meteorological data form the basis for the work of ‘Climate+’ platform services. Facts included in the ontology represent the properties of data climate collections of the ‘Climate’ platform over 19 numerical data collections, which include 40 data sets and 793 numerical data arrays. The climate information ontology includes the description of 170 spatiotemporal objects, characterized by 156 physical quantities. The ontology of climate data collections created is a formal OWL 2 QL description of properties of these data. Its components are specified in the namespace iao:.

The ontology model of numerical arrays of the 'Climate' web portal uses the following information model for storage and presentation. Numerical data are represented in data arrays, stored in netCDF files. The data arrays are grouped into data sets. All the data arrays in a set should satisfy the following conditions: (a) obtained at the same spatial or temporal grid; (b) they should be collected under the same simulation or observation conditions, (c) netCDF files that include the same physical quantities. The data sets are grouped into data collections. Data collection is an ensemble of data sets obtained by an organization within a project but represented on different spatial or temporal grids or for different model scenarios. A collection can consist of one or several data set.

The ontology model for description of numerical arrays is based on the presentation of a real spatiotemporal 3D object, that exists on a certain geographical territory and during a certain time interval. Numerical values of the physical properties of this object are estimated in the model specified in a form of a spatiotemporal grid; the model is called the spatiotemporal system. The spatiotemporal system (iao:SS) is a 4D grid that is defined by lists of numerical values of: longitudes (iao:Lol), latitudes (iao:Lal), heights (iao:HLL), and time (iao:Tl). These lists are subclasses of the class (iao:PD)—physical data; therefore, they contain numerical values of a physical quantity (iao:PQ) in certain units (iao:U) and can be described by: number of values of the array of this physical quantity (iao:dn), proportional step of the physical quantity values (iao:dpq), physical quantity minimum (iao:dmiv), physical quantity maximum (iao:dmax), or numerical values of the physical quantity (iao:dv). The list of time labels (iao:Tl) is additionally described by the initial (iao:dit) and final time labels (iao:dft).

A data array (iao:DA) is an ordered list of numerical values of a physical quantity (iao:PQ); it is a property of a spatiotemporal system (iao:osts) at each 4D point (longitude, latitude, height, and time) of the spatiotemporal system (iao:STS). In the context of OWL language, the data array (iao:DA) is a subclass of the class (iao:PQ) of physical data; therefore, it is the numerical array of values of one physical quantity (iao:PQ) in certain measurement units (iao:U) and is described by the number of array components (iao:dnc) and the physical quantity maximum (iao:dmax) and minimum (iao:dmiv). The set of data arrays (iao:DA) belongs (iao:dda) to a data set (iao:DS), which differs from other data sets in the model scenario (iao:S), spatial resolution (iao:SR), time step (iao:Ts), and membership in a collection (iao:C). A data collection (iao:C) consists of a set (iao:dds) of data sets (iao:DS) and belongs (iao:do) to one organization (iao:O).

Short version of WMO ontology of physical quantities

Main classes of the WMO ontology generalized WMO taxonomy (STFC 2016) are listed in **Table 1**. The structure of this table is described below. The top level of the taxonomy of WMO physical quantities distinguishes several groups of physical quantities: hydrological, land surface (wmo:LSP), air (wmo:MP), and space quantities; each of these groups includes subgroups. **Figure 4** shows individuals of the WMO ontology (Bart, Privezentsev & Fazliev 2017). They represent the classes of physical quantities of the soil category (wmo:SC), long-wave radiation (wmo:LWRC), and temperature category (wmo:TC). The ontology individuals wmo:ST and wmo:SM are instances of the wmo:SCy class, the individual wmo:WMO_DLWRFlux is an instance of the wmo:LWRC class, and the individual wmo:ST, of the wmo:TC class.

Ontology individuals

Figure 4 shows examples of individuals of the OWL-ontology of climate information characterized INM CM4 data collection, WMO taxonomy of physical quantities, and freeze-up on a river applied tasks (T2). The last problem is formulated in papers (Shen 2016; Shen and Chiang 1984; Bart et al. 2018). Data representing the solution of task T2 is located in the data layer (Applied Tasks Data).

Individuals of the OWL-ontology are shown as ovals, and literal values, as dashed rectangles; arrows show properties with unique identifiers from the above tables located in small empty rectangles above. Similarly to a RDF-graph, each arrow connection represents a 'subject–predicate–object' triad.

Figure 4 shows schematically the individuals describing the properties of the data collections and the tasks could be solved. The simplification in **Figure 4** is to consider only one of the four types of structures of individuals of the 'freeze-up on a river' task. The individual shown in **Figure 4** characterizes the temperature of the water surface in the river during the autumn period. The IAO namespace contains the properties of the collections and data sets of the 'Climate+' platform. Each data array is characterized by its space-time resolution as well as by the name of the physical quantity and its dimensionality. The namespace T2 contains individuals describing the properties of the arrays that are input and output data for the freezing model of a river (Bart et al. 2018). To solve the applied task the model should be provided with these input data. Sometimes

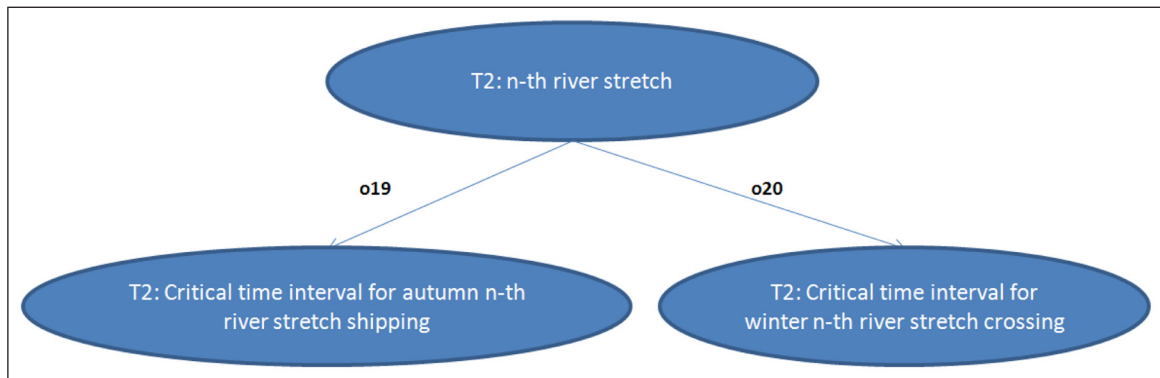


Figure 5: An example of an individual characterizing a river stretch used when deciding on the end of navigation in various stretches of the Ob River and the time of the opening of the stretch for transport crossings on ice.

in practice the names of variables in climatic data collections and in applied models can have different names for the same variables. The task of comparing these names can be simplified and partially automated by usage of the WMO taxonomy. Using saved in the knowledge base rules for comparing the names of climatic data from the collections with those accepted by WMO and using the WMO terms in the model description one can determine availability in the collections of the data required for calculations. The characteristics determining location and/or time of the beginning of the process of ice formation on a river are of interest for the semantic evaluation of the results.

IDSS ontological knowledge base

Critical for decision-making individuals such as time intervals for the formation of ice on the river, ice thicknesses permitting the movement of different weight vehicles over ice, the stage of ice melting on which traffic on the ice of the river is prohibited, and the time interval during which not all river stretches are accessible for shipping are exported to the ontological knowledge base (OKB) used by the 'IDSS' application. The description of this IDSS OKB, 'IDSS' application and interfaces for working with 'Expert System' and 'IDSS' applications is omitted here. **Figure 5** shows the structure of the individual characterizing the time intervals that are critical for decision-making in the autumn season.

Conclusion

The VRE distributed architecture presented allows smooth adding new nodes, computing and data storage systems as well as provides solid computational infrastructure for regional climate change studies based on modern Web and GIS technologies. Usage of the metadata database improves system functional capabilities in terms of extending geospatial dataset archives and statistical processing routines as well as providing computational resources as web services.

A general ontology of climate information resources has been developed and formalized in OWL DL; it reflects the current state of collections of numerical arrays of the 'Climate' web portal, WMO taxonomy of physical quantities, and taxonomy of physical problems of the decision support system. The software has been developed for the formation of the A-box (a set of OWL-ontology individuals) of the climate information ontology of the 'Climate' web portals. Using the software developed, individuals of the climate information ontology have been formed. The climate information ontology model created is a simple and easily extendible taxonomy of information resources required for further work on the development of the decision support system. The solution of reduction problem for the ontologies under study have been described.

Acknowledgement

The authors thank the Russian Science Foundation for the support of this work under the grant No 16-19-10257.

Competing Interests

The authors have no competing interests to declare.

References


- Alipova, KA, Bart, AA, Fazliev, AZ, Gordov, EP, Okladnikov, IG, Privezentsev, AI and Titov, AG.** 2017. Systematization of climate data in the virtual research environment on the basis of ontology approach. In: *Proceedings of SPIE 10466, 23rd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*, 104667. Irkutsk, Russia, 3–7 July 2017. DOI: <https://doi.org/10.1117/12.2289761>
- Athanasios, N, Kalabokidis, K, Vaitis, M and Soulakellis, N.** 2009. Towards a semantics-based approach in the development of geographic portals. *Computers & Geosciences*, 35(2): 301–308. DOI: <https://doi.org/10.1016/j.cageo.2008.01.014>
- Bart, AA, Churuksaeva, VV, Fazliev, AZ, Gordov, EP, Okladnikov, IG, Privezentsev, AI and Titov, AG.** 2018. Ontological Description of Applied Tasks and Related Meteorological and Climate Data Collections. In: *Lecture Notes in Computer Science*. Springer (in press).
- Bart, AA, Churuksaeva, VV, Fazliev, AZ, Privezentsev, AI, Gordov, EP, Okladnikov, IG and Titov, AG.** 2017. Ontological Description of Meteorological and Climate Data Collections. In: *CEUR Workshop Proceedings 2022, Selected Papers of the XIX International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2017)*, 266–272. Moscow, Russia, 9–13 October 2017. Available at: <http://ceur-ws.org/Vol-2022/paper43.pdf> [Last accessed 13 April 2017].
- Bart, AA, Privezentsev, AI and Fazliev, AZ.** 2017. Ontological description of climatic data collections used in decision support system. In: *Proceedings of the Russian Conference “Knowledge-Ontology-Theory” (KONT)*, 47–53. Novosibirsk, Russia, 2–6 October 2017. (In Russian).
- Becirspahic, L and Karabegovic, A.** 2015. Web portals for visualizing and searching spatial data. In: *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 305–311. Opatija, Croatia on 25–29 May 2015. DOI: <https://doi.org/10.1109/MIPRO.2015.7160284>
- Bogdanović, M, Stanimirović, A and Stoimenov, L.** 2015. Methodology for geospatial data source discovery in ontology-driven geo-information integration architectures. *Web Semantics: Science, Services and Agents on the World Wide Web*, 32: 1–15. DOI: <https://doi.org/10.1016/j.websem.2015.01.002>
- Brodaric, B, Fox, P and McGuinness, DL.** 2009. Geoscience knowledge representation in cyber infrastructure. *Computers & Geosciences*, 35(4): 697–699. DOI: <https://doi.org/10.1016/j.cageo.2009.01.001>
- Candela, L, Castelli, D and Pagano, P.** 2013. Virtual Research Environments: An Overview and a Research Agenda. *Data Science Journal*, 12: GRDI75–GRDI81. DOI: <https://doi.org/10.2481/dsj.GRDI-013>
- Copernicus.** 2016. Data Volume, October 2015. Available at: <http://newsletter.copernicus.eu/article/data-volume> [Last accessed 29 November 2017].
- De Roure, D, Jennings, N and Shadbolt, N.** 2001. Research Agenda for the Semantic Grid: *A Future e-Science Infrastructure. Report commissioned for EPSRC/DTI Core e-Science Programme*, December 2001. Available at: http://www.nesc.ac.uk/technical_papers/DavidDeRoure.etal.SemanticGrid.pdf [Last accessed 29 November 2017].
- Gordov, E, Shiklomanov, A, Okladnikov, I, Prusevich, A and Titov, A.** 2016. Development of Distributed Research Center for analysis of regional climatic and environmental changes. *IOP Conf. Series: Earth and Environmental Science*, 48. DOI: <https://doi.org/10.1088/1755-1315/48/1/012033>
- Gordov, EP and Fazliev, AZ.** 2004. Scientific web resources for support of atmospheric investigations. *Computational technologies*, 9(1): 123–136.
- Gordov, EP, Lykosov, VN and Fazliev, AZ.** 2006. Web portal on environmental sciences “ATMOS”. *Advances in Geosciences*, 8: 33–38. DOI: <https://doi.org/10.5194/adgeo-8-33-2006>
- Husain, MF, Al-Khateeb, T, Alam, M and Khan, L.** 2011. Ontology based policy interoperability in geo-spatial domain. *Computer Standards & Interfaces*, 33(3): 214–219. DOI: <https://doi.org/10.1016/j.csi.2010.03.011>
- IPCC.** 2012. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press.
- IPCC.** 2013. *Fifth Assessment Report ‘Climate Change 2013’*. Cambridge, UK: Cambridge University Press.
- Kaklauskas, A.** 2015. *Biometric and Intelligent Decision Making Support*. Berlin, Germany: Springer International Publishing. DOI: <https://doi.org/10.1007/978-3-319-13659-2>
- Lutz, M, Sprado, J, Klien, E, Schubert, C and Christ, I.** 2009. Overcoming semantic heterogeneity in spatial data infrastructures. *Computers & Geosciences*, 35(4): 739–752. DOI: <https://doi.org/10.1016/j.cageo.2007.09.017>

- Okladnikov, IG, Gordov, EP and Titov, AG.** 2016. Development of climate data storage and processing model. *IOP Conference Series: Earth and Environmental Science*, 48. DOI: <https://doi.org/10.1088/1755-1315/48/1/012030>
- Okladnikov, IG, Gordov, EP, Titov, AG and Shulgina, TM.** 2015. Information-computational System for Online Analysis of Georeferenced Climatological Data. In: *Selected Papers of the XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015)*, 1536: 76–80. Obninsk, Russia, 13–16 October 2015. CEUR Workshop Proceedings.
- Riazanova, AA, Voropay, NN, Okladnikov, IG and Gordov, EP.** 2016. Development of computational module of regional aridity for web-GIS 'Climate'. *IOP Conference Series: Earth and Environmental Science*, 48. DOI: <https://doi.org/10.1088/1755-1315/48/1/012032>
- Rowley, J.** 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33: 163–180. DOI: <https://doi.org/10.1177/0165551506070706>
- Ryazanova, AA and Voropay, NN.** 2017. Droughts and Excessive Moisture Events in Southern Siberia in the Late XXth – Early XXIst Centuries. 2017. *IOP Conference Series: Earth and Environmental Science*, 96. DOI: <https://doi.org/10.1088/1755-1315/96/1/012015>
- Schnase, JL, Duffy, DQ, McNerney, MA, Webster, WP and Lee, TJ.** 2014. Climate Analytic as a Service. In: *Proceedings of the Conference on Big Data from Space BiDS'14*, 90–93. Frascati, Italy, 12–14 November 2014. DOI: <https://doi.org/10.2788/1823>
- Shen, HT.** 2016. River Ice Processes. In: Wang, LK, Yang, CT and Wang, MHS (eds.). *Advances in water resource management*, 483–530. Switzerland: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-22924-9_9
- Shen, HT and Chiang, LA.** 1984. Simulation of growth and decay of river ice cover. *Journal of Hydraulic Engineering*, 110(7): 958–971. DOI: [https://doi.org/10.1061/\(ASCE\)0733-9429\(1984\)110:7\(958\)](https://doi.org/10.1061/(ASCE)0733-9429(1984)110:7(958))
- Shulgina, TM, Gordov, EP and Genina, Yu, E.** 2011. Dynamics of climatic characteristics influencing vegetation in Siberia. *Environmental Research Letters*, 6. DOI: <https://doi.org/10.1088/1748-9326/6/4/045210>
- Sillmann, J, Donat, MG, Fyfe, JC and Zwiers, FW.** 2014. Observed and simulated temperature extremes during the recent warming hiatus. *Environmental Research Letters*, 9(6). DOI: <https://doi.org/10.1088/1748-9326/9/6/064023>
- STFC.** 2016. Science and Technology Facilities Council. *The GRIB Discipline Collection*, 5 August 2016. Available at: <http://vocab-test.ceda.ac.uk/collection/grib/Discipline> [Last accessed 30 November 2017].
- WCRP.** 2017. World Climate Research Programme. Coupled Model Intercomparison Project 5 (CMIP5), 29 March 2017. Available at: <https://esgf-node.llnl.gov/projects/cmip5/> [Last accessed 29 November 2017].
- WMO.** 2016. World Meteorological Organization. Climate Services for Supporting Climate Change Adaptation. *Supplement to the Technical Guidelines for The National Adaptation Plan Process*. Geneva, Switzerland: WMO. No. 1170.

How to cite this article: Bart, A, Fazliev, A, Gordov, E, Okladnikov, I, Privezentsev, A and Titov, A. 2018. Virtual Research Environment for Regional Climatic Processes Analysis: Ontological Approach to Spatial Data Systematization. *Data Science Journal*, 17: 14, pp. 1–13, DOI: <https://doi.org/10.5334/dsj-2018-014>

Submitted: 30 November 2017 **Accepted:** 04 June 2018 **Published:** 27 June 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 