University of New Hampshire
University of New Hampshire Scholars' Repository

Faculty Publications

1-1-2016

Development of Distributed Research Center for analysis of regional climatic and environmental changes

E. Gordov
Siberian Branch of the Russian Academy of Sciences

Alexander I. Shiklomanov
University of New Hampshire, Durham, alex.shiklomanov@unh.edu

I. Okladnikov
Siberian Branch of the Russian Academy of Sciences

Alexander A. Prusevich
University of New Hampshire, Durham, alex.proussevitch@unh.edu

A. Titov
Siberian Branch of the Russian Academy of Sciences

Follow this and additional works at: https://scholars.unh.edu/faculty_pubs

Recommended Citation

PAPER • OPEN ACCESS

Development of Distributed Research Center for analysis of regional climatic and environmental changes

To cite this article: E Gordov *et al* 2016 *IOP Conf. Ser.: Earth Environ. Sci.* **48** 012033

View the article online for updates and enhancements.

Development of Distributed Research Center for analysis of regional climatic and environmental changes

E Gordov[1,2,3], A Shiklomanov[4], I Okladnikov[1,3], A Prusevich[4] and A Titov[1,3]

[1]Institute of Monitoring of Climatic and Ecological Systems of the Siberian Branch of the Russian Academy of Sciences, 10/3, Akademichesky pr., Tomsk 634055, Russia
[2]Tomsk State University, 36, Lenina str.,Tomsk 634050, Russia
[3]Institute of Computational Technologies of the Siberian Branch of the Russian Academy of Sciences, Tomsk Branch, 10/4 Akademichesky pr.,Tomsk 634055, Russia
[4]Earth Systems Research Center, Morse Hall, 8 College Road, Durham, NH, USA

Abstract. We present an approach and first results of a collaborative project being carried out by a joint team of researchers from the Institute of Monitoring of Climatic and Ecological Systems, Russia and Earth Systems Research Center UNH, USA. Its main objective is development of a hardware and software platform prototype of a Distributed Research Center (DRC) for monitoring and projecting of regional climatic and environmental changes in the Northern extratropical areas. The DRC should provide the specialists working in climate related sciences and decision-makers with accurate and detailed climatic characteristics for the selected area and reliable and affordable tools for their in-depth statistical analysis and studies of the effects of climate change. Within the framework of the project, new approaches to cloud processing and analysis of large geospatial datasets (big geospatial data) inherent to climate change studies are developed and deployed on technical platforms of both institutions. We discuss here the state of the art in this domain, describe web based information-computational systems developed by the partners, justify the methods chosen to reach the project goal, and briefly list the results obtained so far.

Introduction

The significant increase in the environmental data volumes (up to a petabyte level) and the necessity to store, search, retrieve, process, visualize, and analyze geospatial datasets made the traditional approaches questionable. It required new distributed data access and processing technologies to be created and turned this IT area into a new scientific field for the approaches and tools developed in the recently appeared data intensive application domain [1-5]. The geospatial datasets used in climate and environmental research require the application of a spatial data infrastructure (SDI) approach [6]. This approach, in particular, includes the development of a web portal (geoportal) [7-8] considered as a single access point to SDI resources [9-10]. It provides a functionality of searching the geographic information resources, data sampling and processing, as well as cartographical visualization services along with the corresponding client applications [11]. Such data processing infrastructure should also employ the Web 2.0 standards providing multidisciplinary geographically dispersed research teams with functional capabilities to perform joint investigations [12].

Currently, it is generally accepted that the development of client applications as integrated elements of such infrastructure should be based on the use of modern web and GIS technologies [13-15].

According to the general requirements of the INSPIRE Directive to geospatial data visualization [16], it is necessary to provide such features as data overview, image navigation, scrolling, zooming and graphical overlay, as well as displaying map legends and related meta information. That is, the well-known and widely used basic functionality of a standard GIS should be provided for the potential users. The benefits of the Web-based GIS technologies are quite obvious: theoretical independence of web browsers and operating systems, automatic software location and upgrade, combined use of geographically distributed data sources, and joint use of centralized data warehouses. To perform a reliable analysis of the responses of natural ecosystems and society to the ongoing climate changes, some professional skills and expertise are needed to deal with complex numerical models, modern statistical analysis methods, and programming techniques. Such skills are not very typical for the specialists engaged in related scientific studies. Application of Web-GIS technologies provides environmental scientists and decision makers with reliable comprehensible tools for investigating the climate change impacts using a well-known graphical user interface.

There are several computational and/or information web-systems working with geophysical data linked to environmental science. One of them deals with the analysis of real-time meteorological observational data [17]. A very sophisticated and functional system for analysis and visualization of satellite data was developed at NASA: GES-DISC Interactive Online Visualization ANd aNalysis Infrastructure (GIOVANNI), (http://daac.gsfc.nasa.gov/techlab/giovanni/) [18]. Among those RIMS (http://RIMS.unh.edu/) and CLIMATE (http://climate.climate.scert.ru/) systems are developed by the project partners.

A fully integrated advanced system created at the University of New Hampshire for online analysis of heterogeneous climate, hydrology, and remote sensing data, RIMS was a successful attempt to realize a multi-functional GIS as a web application using MapServer (http://mapserver.org/). RIMS is widely used for various research projects [19-20] including NEESPI (Northern Eurasian Earth Science Partnership Initiative, http://NEESPI.sr.unh.edu/). The main feature of RIMS is that it deals directly with geospatial datasets instead of the corresponding cartographical images, and allows access to the displayed data at each pixel of the map, as well as search and access to data aggregated according to some space and/or time criteria.

A Web-GIS platform called CLIMATE developed at the Institute of Monitoring of Climatic and Ecological Systems SB RAS is intended for processing and visualization of meteorological and climate data [21], thus providing an integration of the results of modeling and observations with the functionality of a retrospective analysis. The platform is accessed via a geoportal which serves as an integration point providing web applications, web mapping services, and metadata management. The platform includes: structured geospatial datasets along with relevant metadata; a computational core (an independent set of modules written in GDL–GNU Data Language, http://gnudatalanguage.sourceforge.net) and Python (http://www.python.org)); spatial data services, geoportal and Web-GIS client. It also provides access to complex numerical models and to the calculation results obtained. In addition, the platform provides researchers with the thematic communication environment and professional education support required for virtual training of young scientists. The developed prototype of a hardware and software platform supports integrated regional climate change studies in Siberia within the framework of the NEESPI Program [22]. It combines Web 2.0 capabilities, tools for running climate models, processing and visualization of large geospatial datasets, and joint application development by dispersed research teams. It also supports virtual training courses for Tomsk State University students and postgraduates [23].

To develop specialized DRC supporting interactive processing and analysis of large geospatial datasets using comprehensive computational and visualization tools the above-mentioned CLIMATE and RIMS systems have been chosen as a basis. Below we describe these in detail and present an approach elaborated to reach the project ultimate objective. It will provide the specialists working in climate and related sciences focused on climate change impact assessment, adaptation strategies, and other climate related activities with accurate and detailed climatic characteristics and reliable tools for an in-depth statistical analysis.

1. Web based information-computational systems developed by the partners

1.1. Web-GIS platform CLIMATE

The platform CLIMATE (http://climate.climate.scert.ru/) is aimed to support regional climatic and ecological changes research [24-26]. In particular, it allows the user to perform calculations of statistical characteristics of meteorological parameters, analysis of periodic and non-periodic processes, and estimation of the derivative climatic indices on the basis of observational and modeling datasets. Also, the platform allows running the mathematical models integrated into it and can be used for supporting collaborative scientific research projects and education of high-grade and post-graduate students of the corresponding departments. It incorporates software components elaborated within the framework of such international projects as ATMOS (http://atmos.iao.ru/ and http://atmos.scert.ru/); RISKS (http://climate.risks.scert.ru/) and ENVIROMIS (http://enviromis.scert.ru/en/) aimed at the development of informational-computational systems. Detailed description of those systems is given in [8, 27-31].

Software of the platform was developed on the base of the special software framework [32]. This framework allows development of thematic information-computation web mapping systems according to specifics of application tasks by connecting the new software elements to its basis. These new software elements might include:

- geospatial datasets represented by modeling and observational data grouped by features and placed on remote storage systems;
- server-side external software modules of computational core;
- client-side GUI components;
- the integrated climate model Planet Simulator [33] and the mesoscale meteorological model WRF [34];
- integrated Web 2.0 tools supporting interaction of distributed groups of researchers and software developers (forums, wiki, blogs, and control system versions).

At present the software of the platform is hosted at IMCES SB RAS (Institute of Monitoring of Climatic and Ecological Systems, Siberian Branch of Russian Academy of Sciences, Tomsk, Russia) on a hardware that includes: high performance server HP Proliant DL585 G7 (four 12-core processors AMD Opteron 6172 and 32 Gb RAM) as the computational core; dedicated server (Intel Pentium 4 and 2 Gb RAM) for the geoportal; and dedicated server (two Intel Xeon 5130 and 4 Gb RAM) for Geoserver software. All servers are interconnected using 1 Gb Ethernet. Collections of spatial datasets are prepared for using and stored on three storage systems with a total capacity of 252 Tb. These datasets include several reanalyses and some results of climatic modeling. Also, modeling results of "Planet simulator" and "WRF" models, archives of remote sensing data for LandSat 4-7, Global Land Survey (GLS) and MODIS (http://glovis.tsc.ru/), as well as observational data from meteorological stations located on the Russian Federation territory are included in the platform. All datasets were validated, and their quality was approved by the data origin institution. Each dataset was reformatted and placed in the storage system in a unified manner. The core modules provide functions for data search, retrieval and pre-processing. Additional software modules were developed to provide the platform with the required thematic area data processing and visualization functionality. They were connected to the computational core of the platform using a dedicated API. Graphical user interface of the platform was designed to meet the requirements of the application tasks. It provides the user with intuitive browser-based instrument similar to such widespread desktop GIS applications as uDIG, QuantumGIS, etc. (see Figure 1).
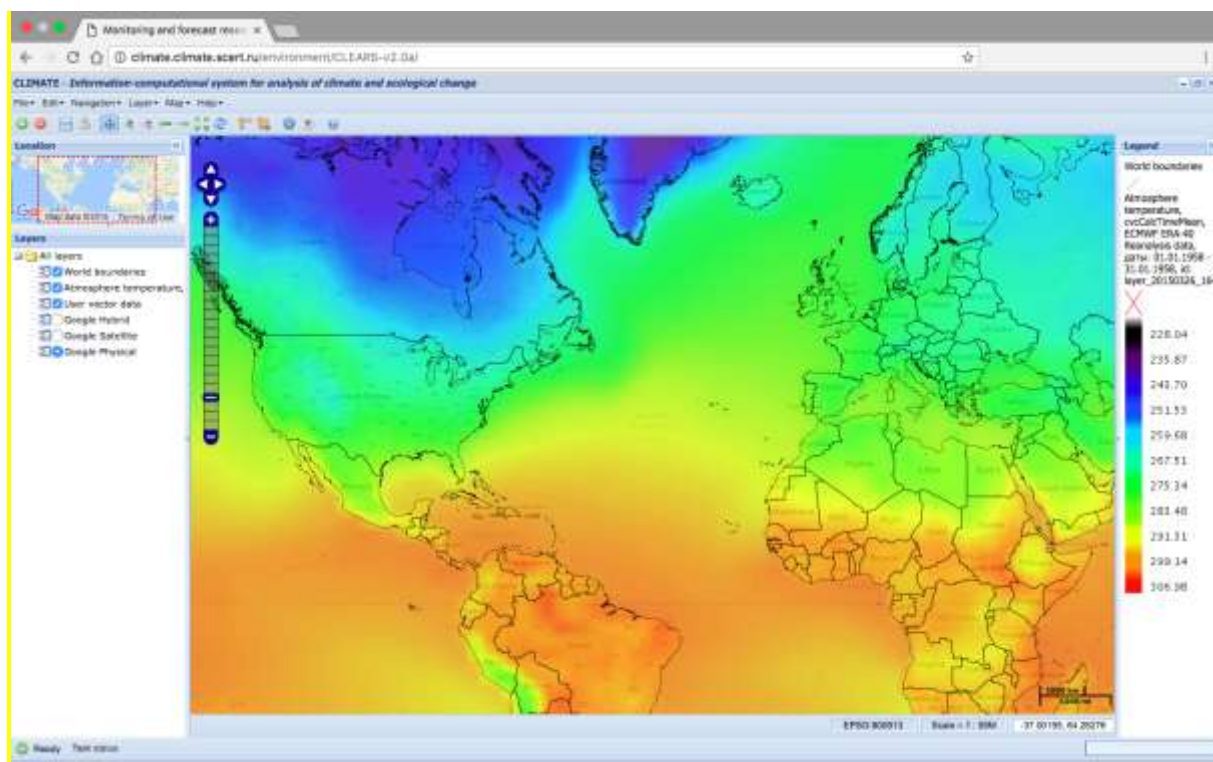
Figure 1. Web-application graphical interface.

The web interface main function is to formalize the requests of a user solving a particular computational task in the field of climatic and ecological research, to compile a task technical description for the computational core, and to present the computational results in digital and graphical formats using modern web GIS technologies in a way similar to the conventional desktop GIS. The environment aimed at organizing the interactions of distributed research groups represents a complex solution for the platform information support and includes capabilities for creating user groups with access rights, public and limited (to particular users and groups) blogs, links to publications (Internet links, presentations, papers, analysis results), and collaborative software and documentation development (version control system).

At present, the platform developed is used for studying the ongoing climatic changes and their consequences in Siberia. The obtained results of an analysis of the surface temperature and precipitation changes are in a good agreement with the results obtained by other researchers, thus approving the reliability and effectiveness of the platform for climate and ecosystems change analysis [35]. The educational components of the platform give an opportunity to use it for education of high-grade and postgraduate students [36]. Since 2012 students of the Meteorology and Climatology Department of Tomsk State University are using the platform to study the basics of climate change analysis and to perform on-line computational labs supporting courses on "Regional climate change analysis" and "Future climate analysis".

1.2. RIMS: Rapid Integrated Mapping and analysis System
RIMS (Rapid Integrated Mapping and analysis System) is developed to test the usability of Web interfaces and services with a wide range of geospatial Earth system and socio-economic information [37]. It is predominantly a data oriented system interfacing the digital content of the datasets rather than their image representation such as data viewing Web tools, e.g. Google Maps/Earth applications. In addition to visualization/mapping, the major Web services of RIMS are specifically designed for interfacing the following digital data content:

- dataset search (Catalog) services. Presently there are over 10000 raster and vector time series and single layer datasets (in 100+ file formats readable by the GDAL library) in our data holdings;
- Data Cube views of time series and/or spatially aggregated data (prototype of Web Aggregation Service: WAS). This is the ability to view different temporal aggregation subsets of the same time series data (e.g. daily, monthly, yearly, daily climatology or long term averages, etc.) as well as their spatial aggregations (e.g. per country, per watershed, etc. integral averages);
- instant access to raw data values for each pixel of client-side displayed maps (WFS/WCS). This is a raw data layer on the client side displayed on the top of the map imagery;
- on-the-fly data production and manipulation tools (a Data Calculator that does algebra, calculus, and built-in server side modeling (e.g. water balance model) over arbitrary sets of data) that utilize concepts of Web Production Service (WPS) extensions;
- tools for easy mounting and management of vector, raster and point/station data and metadata.

An illustration of the Web-client application of RIMS is given in Figure 2.
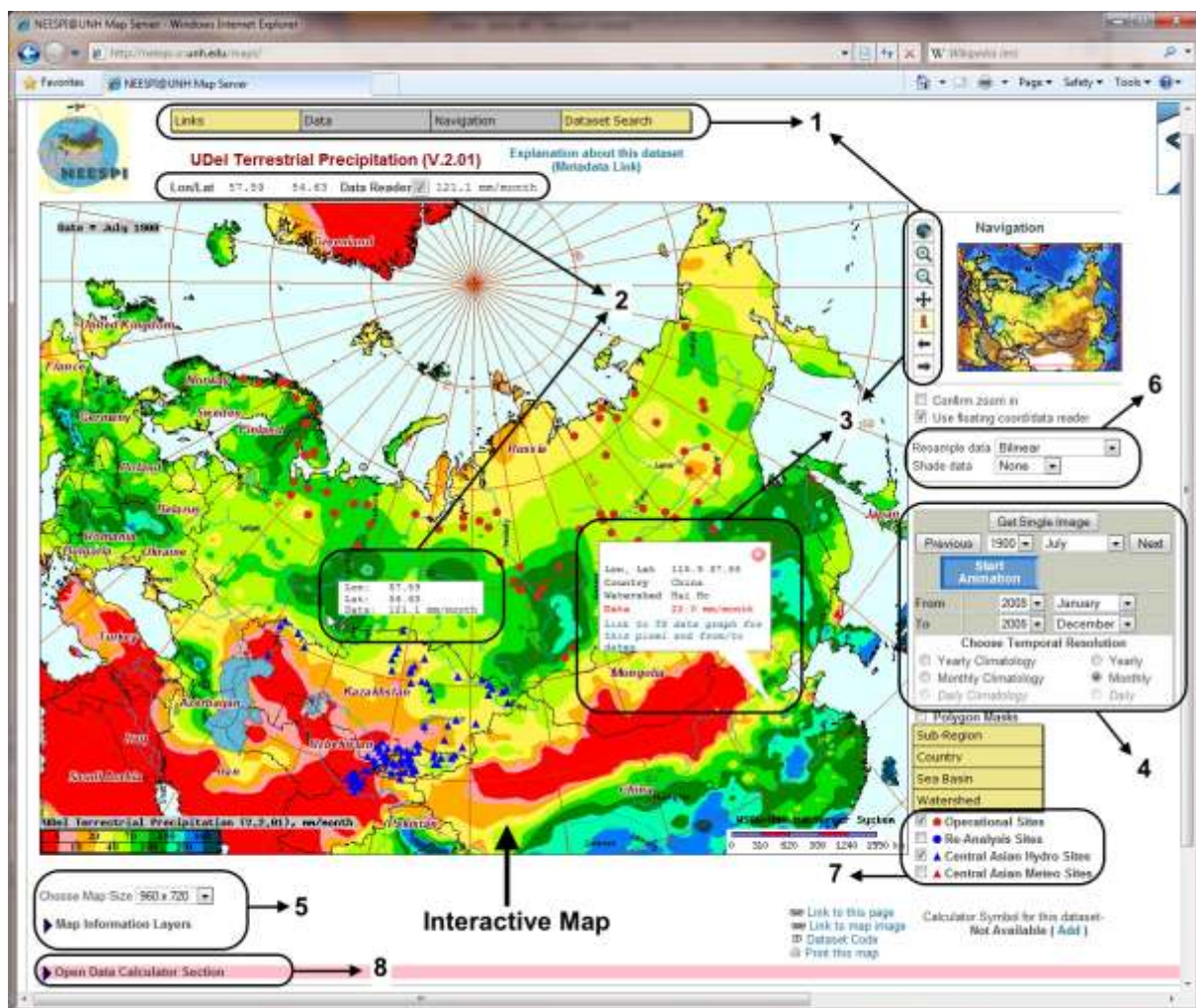

Figure 2. Web-client application of an adaptive projection (North always up) customized for the NEESPI site with adaptive dynamic re-projection (always North up). Keys: (1) basic data and map navigation tools include drop-down menu for data search/selection, sidebar with zoom/pan, back/forward map history tools for spatial/data navigation, metadata link, etc.; (2) interactive

coordinate and map data value reader for the mouse over locations (loads raw map digital data to the browser); (3) pixel query tool (i-tool) gets coordinates, country, watershed, and map data values in a callout frame that has a link to time-series graphs and spreadsheet data for this pixel as well as pre-built polygon aggregations for this pixel location; (4) time series navigation tool allows the user to choose and display maps/animations for a given date(s) and switch temporal resolutions for the selected dataset (see DataCube concept section); (5) map size and base layer choices including a DEM river network option; (6) data interpolation and shading tools allow choices for smoothed/interpolated visual interpretations; (7) point/station data list: selected datasets (stations) are displayed on the map with clickable symbols that open station pages in a separate browser window; (8) fold-out section to run the Data Calculator application to perform mathematical and logical functions over gridded or vector datasets to build custom maps, graphs and spatial queries (see Data Calculator section).

The idea of the DataCube web aggregation service (WAS) is to combine data mapping and query in three virtual dimensions: (a) scaled time aggregation, (b) non-scaled or climatological time aggregation, and (c) polygon aggregation. Only one dataset in the DataCube represents the original core data, while the others are pre-built (predetermined) aggregation derivatives, as illustrated in Figure 3. The DataCube concept allows the user to explore the same dataset in many different ways without the need to run their own aggregation software. For example, a functional Time-Series Navigation tool in RIMS is built on the WAS concept (Figure 2, item #4).
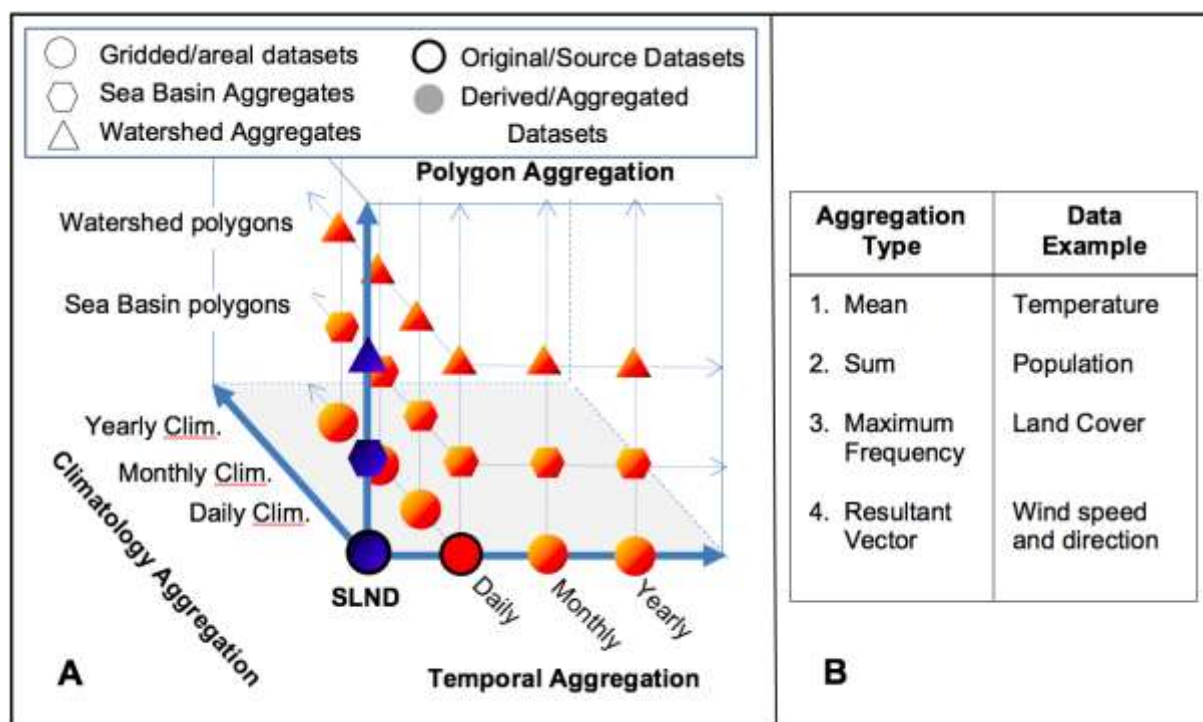

Figure 3. Concept of data aggregation scheme used in RIMS. Original Daily dataset (e.g. NCEP daily temperature at 2 m) can be aggregated along the temporal scale to monthly and yearly derivative datasets, and along the climatology scale to daily, monthly and yearly climatology (long-term averages) derivative sub-datasets. In turn, each of these can be aggregated by any number of polygon sets (on the polygon aggregation scale) to polygon averages or cumulative (e.g. average temperature per country). Single layer non-dated datasets (e.g. elevation) can be aggregated only along the polygon aggregation scale (e.g. average elevation of a watershed). Aggregation method can be one of the following types: (1) average, e.g. temperature; (2) cumulative, e.g. population; (3) max frequency, e.g. land cover; (4) vector average, e.g. wind.

The Data Calculator concept for the Web Processing Service (WPS) extension originated from the idea to allow the user to perform interactive on-the-fly arbitrary mathematical, statistical, custom functions, and/or logical statements/queries on a pixel or at the polygon (area integral) level. In the existing prototype it works with any number of datasets and their time series data layers. This service will be provided by a dedicated Data Production Server, which first loads data hyperslabs from the WCS or directly from the primitive data DB and then validates and evaluates the submitted equations and/or logical expressions over the data. The WPS returns the calculation results in any combination of the following formats: (1) maps; (2) data value frequency histograms; (3) georeferenced data files (e.g. NetCDF, AAIGrid or other GDAL formats); (4) polygon time series data (spreadsheet format); (5) graph of polygon data; and (6) statistical moments of the resulting data.

2. The approach

The project includes several major directions of research (Tasks) listed and described below. Existing geospatial datasets used in climate and environmental research are very heterogeneous in storage structure and file formats; they cover different spatial and temporal scales, consist of different sets of climatic and environmental characteristics, and often have a huge size that significantly complicates their analysis. Thus, the first step is pre-processing of these geospatial datasets unifying their storage formats and access tools that allow searching and the retrieval of multidimensional subsets for specific research needs of various users. We shall develop an approach that will unify access to distributed archives of environmental data hosted at ESRC and IMCES and extract (generate) principal metadata for these datasets.

Network Common Data Form (netCDF) was chosen as the major file format for most geospatial data in our data archive. This format is formally acknowledged by scientific institutions (including UCAR - University Corporation for Atmospheric Research, http://www2.ucar.edu/) and OGC (Open Geospatial Consortium, http://www.opengeospatial.org/) as a standards' candidate for storing geospatial data and stimulating data exchange in binary form. It is perfectly suited for storing geospatial data and supported by a wide range of commercial and open source software.

The data are stored on data storage systems as collections of netCDF files and arranged in a strict hierarchy of directories:

/<data root directory>/
   <data collection name>/
      <spatial domain resolution>/
         <time domain resolution>/
            <files and directories with data>

Here, *<data root directory>* is the root location of data collections, *<data collection name>* is the name of a directory containing a single data collection, *<spatial domain resolution>* is the name of a directory containing data with the same horizontal resolution, *<time domain resolution>* is the name of a directory containing data with the same time step. All data files (sometimes grouped in subdirectories) are located deeper in the hierarchy. The names of files and subdirectories are not regulated and determined by the specifics of a particular data set. Each data file contains one or more multi-dimensional arrays of meteorological parameters. The data are stored on working nodes or on connected data storage systems (Local data) according to a "data awareness" principle. Data awareness is some knowledge about all of the data available within each working node and where those data are located [38]. According to this information, the datasets are processed locally, on the nodes containing them. A detailed description of the data storage and processing model is represented in [39].

Due to the huge volume of geospatial datasets, additional measures are taken to increase the speed of data searching, viewing, and sampling. Within the framework of the project, a dedicated database of metadata describing the available at ESRC and IMCES multidimensional arrays of climatic, environmental, demographic, and socio-economic characteristics will be developed on the base of the

DBMS MySQL. This database will facilitate and accelerate searching of required information on storage systems.

The rapidly increasing volumes of information during the past decade significantly decreased the effectiveness of single-server Web GISs, like a thematic Web GIS for cloud analysis of multidimensional climatic characteristics recently developed at IMCES SB RAS. This project seeks to develop an online Distributed Research Center with information and computing software package for distributed processing of large datasets. A paradigm of "share nothing" provides the best approach to solving this problem. This technology, when the data are residing where they are and only upfront software handles the information needs of the customers, will be used for the development of a new computing software complex, i.e., the Center itself. All efforts to minimize the typical shortcomings of this paradigm will be made within the framework of the project.

More than 90% of queries for climatic and environmental data services are quite simple. The users request spatial pattern of one or more variables, their temporal dynamics and a suite of their basic statistics. They may also ask for specific analyses (e.g., "what if") made on-line. Scientists prefer to do their own assessments and require original (but preprocessed) data and comprehensive information about their specifics and quality (metadata). Only in rare situations an in-depth analysis for specific issues is required from the Data Services. The proposed DRC will respond to all three types of these queries. The first two types will be handled on-line, and the third type of requests will be processed in-house by several dedicated software packages on an order-response basis. The DRC staff will collect (or project) typical requests for the third, most demanding type of queries and develop a suite of software packages to allow expediently addressing them.

Summarizing the approach we chose the following features for the DRC under development:
- geoportal with user-friendly and intuitive clear GUI as a single access point to distributed archives of spatial data and tools for their processing;
- cloud data analysis and visualization;
- flexible modular computational engine with verified data processing routines;
- export of processing results and metadata as image (PNG, JPEG, GIF, GeoTIFF), as NetCDF files as well as access via WMS and WFS.

To meet the above requirements, the following tasks are carried out now:
- development of an architecture and defining the major hardware and software components of the DRC for monitoring and projecting the regional environmental changes;
- development of an information database and computing software suite for distributed processing and analysis of large geospatial data hosted at ESRC and IMCES;
- development of geoportal, thematic web client and web services providing international research teams with an access to cloud computing resources at DRC; two options will be executed: access through a basic graphical web browser and using geographic information systems.

Using an output of these tasks, compilation of the DRC prototype, its validation, and testing the DRC feasibility for analyses of the recent regional environmental changes over Northern Eurasia and North America and public services is planned to be done this fall.

3. First results

3.1. Metadata database

To describe geospatial datasets and their processing routines and provide effective DRC functioning, a dedicated metadata database (MDDB) was developed. This database contains spatial and temporal characteristics of available geospatial datasets, their locations, and run options of software components for data analysis. The tables in MDDB are divided into "technical" and "interface" ones. The technical tables contain data intended for computing software components. The interface tables hold textual multilingual content for a graphical user interface.

There are two major parts of MDDB providing description of climate datasets and description of data processing software components (computing modules).

Each climate dataset in MDDB is uniquely identified by its four major characteristics: name of the data collection, resolution of the horizontal grid, resolution of the time grid and name of the modeling scenario (if applicable). And each dataset includes one or several data arrays containing values of various meteorological parameters given on spatial and temporal grids. The information about all datasets available for analysis is stored in the first part of MDDB. It is used to locate the data files and to provide the metadata on request.

Geospatial data processing is performed by a set of dedicated computing modules. These modules are run in accordance to a pipelined call sequence. This sequence is prepared by the web portal on the basis of user interactions with the graphical user interface. Second part of MDDB contains a description of various call sequences and their options. Since some data analysis routines are designed to process only specific meteorological parameters, the connections between computing modules and data arrays are set in MDDB.

3.2. Web-GIS client

The boundless/OpenGeo architecture was used as the basis for a Web-GIS client development. The cartographical web application (Web-GIS client) for working with archive of geospatial NetCDF datasets contains 3 basic tiers [24]:

- tier of NetCDF metadata in JSON format;
- middleware tier of JavaScript objects implementing methods to work with:
  - NetCDF metadata;
  - XML file of selected calculations configuration (XML task);
  - WMS/WFS cartographical services;
- graphical user interface tier representing JavaScript objects realizing general application business logic.

3.2.1. NetCDF metadata tier.
Web-GIS client metadata tier represents a set of interconnected JSON objects, created on the base of MySQL metadata relations, and presenting NetCDF datasets information (spatial and temporal resolutions, meteorological parameters available, acceptable processing procedures, etc.). Generally, there are two kinds of objects:

- objects with the structure conforming to the corresponding metadata database relations, for instance, object of measurement units;
- objects based on complex SQL queries to metadata relation sets that allow fast retrieving of necessary information using MySQL indices as associative array keys.

The structure of JSON objects was chosen according to the following criteria:

- efficiency of filling out graphical user interface interactive forms;
- optimization of process of creating and editing of XML file of selected calculations configuration (XML task).

It might be concluded that by virtue of the approach chosen above the processes of interaction between user and metadata database via Web-GIS graphical interface are optimized.

3.2.2. Middleware tier of JavaScript objects.
This tier implements methods to work with NetCDF metadata, XML task file and WMS/WFS cartographical services, and appears to be a middleware, which connects JSON metadata and graphical user interface tiers. The methods include such procedures as:

- loading and updating of metadata JSON objects using AJAX technology;
- creating, editing, serialization of XML calculation task object;
- launching and tracking the task execution process located on the remote calculation node;

- working with WMS/WFS cartographical services: obtaining a list of available layers, presenting layers on a map, export layers into various formats according to a user request, obtaining and presenting a layer legend with the selected SLD style applied.

3.2.3. Graphical user interface. The tier is based on a conjunction of JavaScript libraries, such as OpenLayers, GeoExt and ExtJS, and represents a set of software components either standalone (information panels, buttons, list of layers, etc.), or implementing general application business logic (menu, toolbars, wizards, mouse and keyboard event handlers, and so on). The graphical interface performs two main functions: providing functional capabilities for editing XML task file, and a visual presentation of cartographical information for the end user. It is similar to the interfaces of such popular classic GIS applications as uDig, QuantumGIS, etc. The basic elements of the graphical user interface include (figure 1):

- panel displaying user cartographical layers on the map. Google maps are used as the base layer by default, but there's a possibility to set an arbitrary base layer including newly created by the user;
- layer tree allowing to toggle layer display;
- layer legends display panel;
- map information panel (scaling, cartographical projections, cursor geographical coordinates)
- application general status panel;
- overview map panel;
- general application menu;
- toolbar (adding/removing layer, saving NetCDF data, panning, map refresh, obtaining of information related to given geographical point, etc.);
- application context menu;
- wizard creating cartographical layers based on results of computational processing of geospatial datasets available to the system.

The toolbar, application, and context menus contain mouse and keyboard event handlers which uniquely define Web-GIS behavior depending on the user actions with the execution context applied. Web-GIS client complies with the general INSPIRE standard requirements and provides computational processing services launching to support solving tasks in the area of environmental monitoring, as well as presenting the calculation results in the form of WMS/WFS cartographical layers in raster (PNG, JPG, GeoTIFF), vector (KML, GML, Shape), and binary (NetCDF) formats.

It should be noted that the geospatial data cartographical services based on Geoserver software can be used in the Web-GIS client considered as well as in the standard desktop GIS applications.

4. Specific results

To date, there is no formalized description of the metadata database for large sets of geospatial meteorological and climatic data. The architecture presented above is the first attempt to address this crucial for data intensive domain problem. The designed metadata database solves three main tasks: 1) provides a content for the graphical user interface; 2) provides the geoportal with the information needed to generate a correct task file for the computational core; 3) contains information on the structure and arrangement of data sets allowing the computational core to read and process them efficiently. The use of this database organizes the information on the available data sets, facilitates an automatic retrieval of the data files, and improves the scalability and flexibility of the computations.

The developed GIS Web client is based on the architecture Boundless / OpenGeo. The first version of the GUI uses interrelated JavaScript libraries, OpenLayers, GeoExt and ExtJS and represents a set of software components including independent components (dashboards, buttons, layer lists), and those implementing a general logic of the software application (menus, toolbars, wizards, mouse and keyboard event handlers, etc.).

The first application of the above-developed metadata database and user interface showed that their combined usage facilitates the expanding of the set of data archives available for analysis and adding new statistical processing procedures [40].

The results obtained show that the thus developed VRE and tools will be useful for the decision makers and specialists working in the related sciences, with the focus of their work as socio-economic impact assessment, ecological impact assessment, adaptation strategies, science policy administration, and other climate related activities. On this basis they will get the reliable climate related characteristics required for studies of the economic, political, and social consequences of global climate change at a regional level.

5. Conclusions

The obtained and anticipated project results will create a pathway for the development and deployment of thematic international virtual research centers focused on interdisciplinary studies by international research teams.

The software developed in this project will be utilized for processing the partners datasets and other datasets including observations and reanalysis products in order to perform a space-time analysis of the recent and possible future dynamics of climate and all of its components with a special emphasis on the extreme climatic phenomena in the Northern extratropics.

The following main goals will be achieved upon the completion of the project:

- data analysis products made available to researchers and stakeholders in the form of digital atlases and data arrays in electronic form through the project website and other recognized media;
- interactive web tools developed for a comprehensive analysis of climate variables and their derivatives provided by the thematic geoportal;
- dialog with stakeholders established for further improvement of the developed technology.

The DRC will provide an opportunity for scientists, professionals, and decision-makers to use various geographically distributed and georeferenced resources and processing services via a web browser by integrating the geoportal distributed storage, processing, and delivery of information. In particular, it will allow carrying out a simultaneous analysis of several georeferenced thematic data sets using modern statistics and thus identify the responses of ecological processes to climate change.

References

[1] Big Data Definition *MIKE 2.0* http://mike2.openmethodology.org/wiki/Big_Data_Definition
[2] Kusnetzky D 16 February 2010 What is "Big Data?" *ZDNet* http://www.zdnet.com/article/what-is-big-data/
[3] Vance A 22 April 2010 Start-Up Goes After Big Data With Hadoop Helper *New York Times Blog* http://bits.blogs.nytimes.com/2010/04/22/start-up-goes-after-big-data-with-hadoop-helper/?dbk
[4] Kalinichenko L A, Volnova A A, Gordov E P, Kiseleva N N, Kovaleva D A, Malkov O Yu, Okladnikov I G, Podkolodny N L, Pozanenko A S, Ponomareva N V, *et al* 2016 Problems of access to data in Data Intensive Domain in Russia *Informatics and Applications* **10** 3–23
[5] Hilbert M 2016 Big Data for Development: A Review of Promises and Challenges. *Development Policy Review* **34** (1) 135–74 doi: http://doi.org/10.1111/dpr.12142
[6] Steiniger S and Hunter A J S 2012 Free and open source GIS software for building a spatial data infrastructure *Geospatial Free and Open Source Software in the 21st Century* ed Bocher E, Neteler M (Heidelberg: LNGC, Springer) pp 247–61
[7] Gordov E P and Fazliev A Z 2004 *J. Computational Technologies (Spec. Iss.)* **9** 123–26

[8]   Gordov E P, Lykosov V N and Fazliev A Z 2006 *Advances in Geosciences* **8** 33–38
[9]   Koshkarev A V, Ryakhovskii A V and Serebryakov V A 2010 Infrastructure of distributed environment of storage, search and transformation of geospatial data *Open Education* **5** 61–73
[10]   Krasnopeev S M 2001 Experience of deployment of key elements of special data infrastructure on the base of web-services *Proc. XIV All-Russia Conf. on Internet and Modern Society* (November 20-23, 2001, St. Petersburg, Russia) pp 92–9
[11]   Koshkarev A V 2008 Geoportal as a tool to control spatial data and services *Spatial data* **2** 6–14
[12]   Allan R J 2009 *Virtual Research Environments: from Portals to Science Gateways* (Chandos: Oxford Publishing)
[13]   Yakubailik O E 2007 Geoformation geoportal *J. Computational Technologies (Special Issue 3)* **12** 116–25
[14]   Dragicevic S, Balram S and Lewis J 2000 The role of Web GIS tools in the environmental modeling and decision-making process *4th Int. Conf. on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs* (September 2-8, 2000, Banff, Alberta, Canada) http://www.srcosmos.gr/srcosmos/showpub.aspx?aa=5823
[15]   Frans J M van der Wel 2005 Spatial data infrastructure for meteorological and climatic data *Meteorol. Appl.* **12** 7–8
[16]   Vatsavai R R, Thomas E B, Wilson B T and Shekhar S 2000 A Web-based browsing and spatial analysis system for regional natural resource analysis and mapping *Proc. of the 8th ACM Int. symp. on Advances in geographic information systems* (Noveber 6-11, 2000, Washington, D.C., US) pp 95–101
[17]   Samet R and Tural S 2010 Web based real-time meteorological data analysis and mapping information system *WSEAS Transactions on Information Science and Applications* **7**(9) 1115–25
[18]   Berrick S W, Leptoukh G, Farley J D and Rui H 2009 GIOVANNI: A Web service workflow-based data visualization and analysis system *IEEE Trans. Geosci. Remote Sens.* **47**(1) 106–13.
[19]   Welp L R, Randerson J T, Finlay J C, Davydov S P, Zimova G M, Davydova A I and Zimov S A 2005 A high-resolution time series of oxygen isotopes from the Kolyma River: Implications for the seasonal dynamics of discharge and basin-scale water use *Geophys. Res. Lett.* **32** L14401 doi: 10.1029/2005GL022857.
[20]   Shiklomanov A I 2012 Regional Integrated Monitoring System for the Pan-Arctic (RIMS) *Proc. of the Northern Eurasian Earth Science Partnership Initiative (NEESPI) Regional Science Team Meeting devoted to the High Latitudes* (June 2-6, 2012, Helsinki, Finland) pp 128–30
[21]   Gordov E P, Lykosov V N, Krupchatnikov V N, *et al* 2013 *Computational-Information Technologies for Monitoring and Modeling of Climate Change and its Consequences* (Novosibirsk: Nauka)
[22]   Gordov E, Bryant K, Bulygina O, Csiszar I, Eberle J, Fritz S, Gerasimov I, Gerlach R, Hese S and Kraxner F et al 2013 Development of Information-Computational Infrastructure for Environmental research in Siberia as a baseline component of the Northern Eurasia Earth Science Partnership Initiative (NEESPI) Studies *Regional Environmental Changes in Siberia and Their Global Consequences (Series: Springer Environmental Science and Engineering)* ed P Ya Groisman and G Gutman **XII** pp 19–55
[23]   Gordova Yu E, Genina E Yu, Gorbatenko V P, Gordov E P, Kuzhevskaya I V, Martynova Yu V, Okladnikov I G, Titov A G, Shulgina T M and Barashkova N K 2013 Support of the educational process in the field of modern climatology based on Web GIS "Climate" platform *Open Distance Education* **1**(49) 14–9
[24]   Okladnikov I G, Gordov E P and Titov A G 2012 Hardware-software platform «Climate» as a basis for local spatial data infrastructure geoportal *Vestnik NSU (Series: Information Technologies)* **10** (4) 105–11

[25]   Okladnikov I G, Gordov E P, Titov A G, Bogomolov V Yu, Shulgina T M and Genina E Yu 2012 Geoinformation system for investigation of regional climatic changes and first results obtained *Atmos.Ocean. Opt.* **25** (2) 137–43

[26]   Gordov E P, Okladnikov I G and Titov A G 2016 Information and Computing Web-System for Interactive Analysis of Georeferenced Climatological Data Sets *Vestnik NSU (Series: Information Technologies)* **14** (1) 13–22

[27]   Baklanov A and Gordov E P 2006 Man-induced environmental risks: monitoring, management and remediation of man-made changes in Siberia *J. Computational technologies (Special issue)* **11** 162–71

[28]   Gordov E P and Lykossov V N 2007 Development of information-computational infrastructure for integrated study of Siberia environment *J. Computational technologies (Special issue)* **12** 19–30

[29]   Gordov E P, Okladnikov I G and Titov A G 2007 Development of elements of a web-based information-computational system for studies of regional environment processes *J. Computational technologies (Special issue 3)* **12** 20–8

[30]   Gordov E P, Okladnikov I G and Titov A. G 2011 Application of web mapping technologies for development of information-computational systems for georeferenced data analysis *Vestnik NSU (Series: Information Technologies)* **9** (4) 94–102

[31]   Shulgina T M, Gordov E P, Okladnikov I G, Titov A G, Genina E Y, Gorbatenko V P, Kuzhevskaya I V and Akhmetshina A S Computational module for regional climate change analysis *Vestnik NSU (Series: Information Technologies)* **11** (1) 124–31

[32]   Titov A G and Okladnikov I G 2014 Architecture of Web-GIS for climate monitoring based on geospatial data services *Vestnik NSU (Series: Information Technologies)* **12** (1) 79–88

[33]   Fraedrich K, Jansen H, Kirk E *et al* 2005 The Planet Simulator: Towards a user friendly model *Meteorologische Zeitschrift* **14** 299–304

[34]   Skamarock W C, Klemp J B, Dudhia J, Gill D O, Barker D M, Duda M G, Huang X Y, Wang W and Powers J G 2008 *A Description of the Advanced Research WRF Version 3* (Technical note   NCAR/TN-475+STR,   June   2008) http://www2.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf

[35]   Shulgina T M, Genina E Yu and Gordov E P 2011 *Environmental Research Letters* **6** 045210

[36]   Gordova Yu E, Martynova Yu V and Shulgina T M 2014 *Bulletin of Irkutsk State Univ. (Series: Earth's Sciences)* **9** 55–68

[37]   Prusevich A A, Shiklomanov A I and Lammers R B 2012 RIMS: An Integrated Mapping and Analysis System with Application to Siberia *Regional Environmental Changes in Siberia and Their Global Consequences* ed P Ya Groisman and G Gutman (New York: Elsevier) pp 46–50

[38]   Brackett M 2012 *Data Resource Integration: Understanding and Resolving a Disparate Data Resource* (Technics Publications, LLC publishing) p 540

[39]   Okladnikov I G, Gordov E P and Titov A G 2016 Development of data storage and processing model *IOP Conf. Ser.:Earth and Environmental Science* In press

[40]   Ryazanova A A, Voropay N N and Okladnikov I G 2016 Application of information and computing web system «Climate» for estimation of aridity of South Siberia *Proc. of Int. Conf. and Early Career Scientists School on Environmental Observations, Modeling and Information Systems ENVIROMIS-2016 (*July 11-16, 2016, Tomsk, Russia) pp 358–62