Architecture of a spatial data service system for statistical analysis and visualization of regional climate changes

A G Titov[1,2], I G Okladnikov[1,2] and E P Gordov[1,2]

[1] Institute of Monitoring of Climatic and Ecological Systems SB RAS, Tomsk, Russia
[2] V.E. Zuev Institute of Atmospheric Optics SB RAS, Tomsk, Russia

titov@scert.ru

Abstract. The use of large geospatial datasets in climate change studies requires the development of a set of Spatial Data Infrastructure (SDI) elements, including geoprocessing and cartographical visualization web services. This paper presents the architecture of a geospatial OGC web service system as an integral part of a virtual research environment (VRE) general architecture for statistical processing and visualization of meteorological and climatic data. The architecture is a set of interconnected standalone SDI nodes with corresponding data storage systems. Each node runs a specialized software, such as a geoportal, cartographical web services (WMS/WFS), a metadata catalog, and a MySQL database of technical metadata describing geospatial datasets available for the node. It also contains geospatial data processing services (WPS) based on a modular computing backend realizing statistical processing functionality and, thus, providing analysis of large datasets with the results of visualization and export into files of standard formats (XML, binary, etc.). Some cartographical web services have been developed in a system's prototype to provide capabilities to work with raster and vector geospatial data based on OGC web services. The distributed architecture presented allows easy addition of new nodes, computing and data storage systems, and provides a solid computational infrastructure for regional climate change studies based on modern Web and GIS technologies.

1. Introduction

Dramatically fast global climate changes have been observed during the last 50 years. They set conditions for rapid development of environmental monitoring and modeling technologies. New methods of processing of environmental geospatial datasets have been developed, and most of them are based on modern data science methods. Among them are statistical methods of extracting extreme climate events from meteorological time series presented in the form of geospatial datasets, as well as assessment of their spatio-temporal dynamics. The use of large geospatial datasets in climate change studies requires performing a number of procedures, such as dataset collection, storage, search, exchange, transfer, processing, visualization, and analysis. To achieve this, a set of Spatial Data Infrastructure (SDI, [1]) elements should be created and published, particularly including geophysical data processing (geoprocessing) web services and services of cartographical visualization. It is generally accepted that the development of client applications as integrated elements of such an infrastructure should be based on the use of modern Web and GIS technologies [2, 3, 4, 5]. According to the general requirements of the INSPIRE Directive on geospatial data visualization [6], 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. Also, the OGC specifications define that the basic functions to be performed by geoprocessing services are as follows: geospatial data visualization (WMS), provision in vector (WFS) and raster (WCS) formats, as well as geospatial analytical processing (WPS).

At present there are a number of information systems and services that provide some parts of the functionality required. An advanced example is a fully integrated system created at the University of New Hampshire for online analysis of heterogeneous climate, hydrology, and remote sensing data (RIMS). It is a successful attempt to implement a multi-functional GIS as a web application using MapServer (http://mapserver.org/) software. RIMS have been widely used in various research projects [7, 8] including NEESPI (Northern Eurasian Earth Science Partnership Initiative, http://NEESPI.sr.unh.edu/). A major feature of RIMS is that it deals directly with geospatial datasets instead of corresponding cartographical images, and allows access to the data displayed at each pixel of the map, as well as search and access to data aggregated according to some space and/or time criteria. Its main defect that narrows the system's possible use is that it does not meet the SDI/OGC standards.

This paper presents the architecture of a geospatial OGC web service system as an integral part of a virtual research environment (VRE) general architecture for statistical processing and visualization of meteorological and climatic data aimed at providing decision makers and researchers with accurate and reliable climate change characteristics. The architecture can be considered as a result of an elaborated synergy approach based on RIMS technological solutions and techniques developed by the authors.

2. General architecture of the VRE

The architecture (Figure 1) represents a set of interconnected standalone SDI nodes (Virtual Research

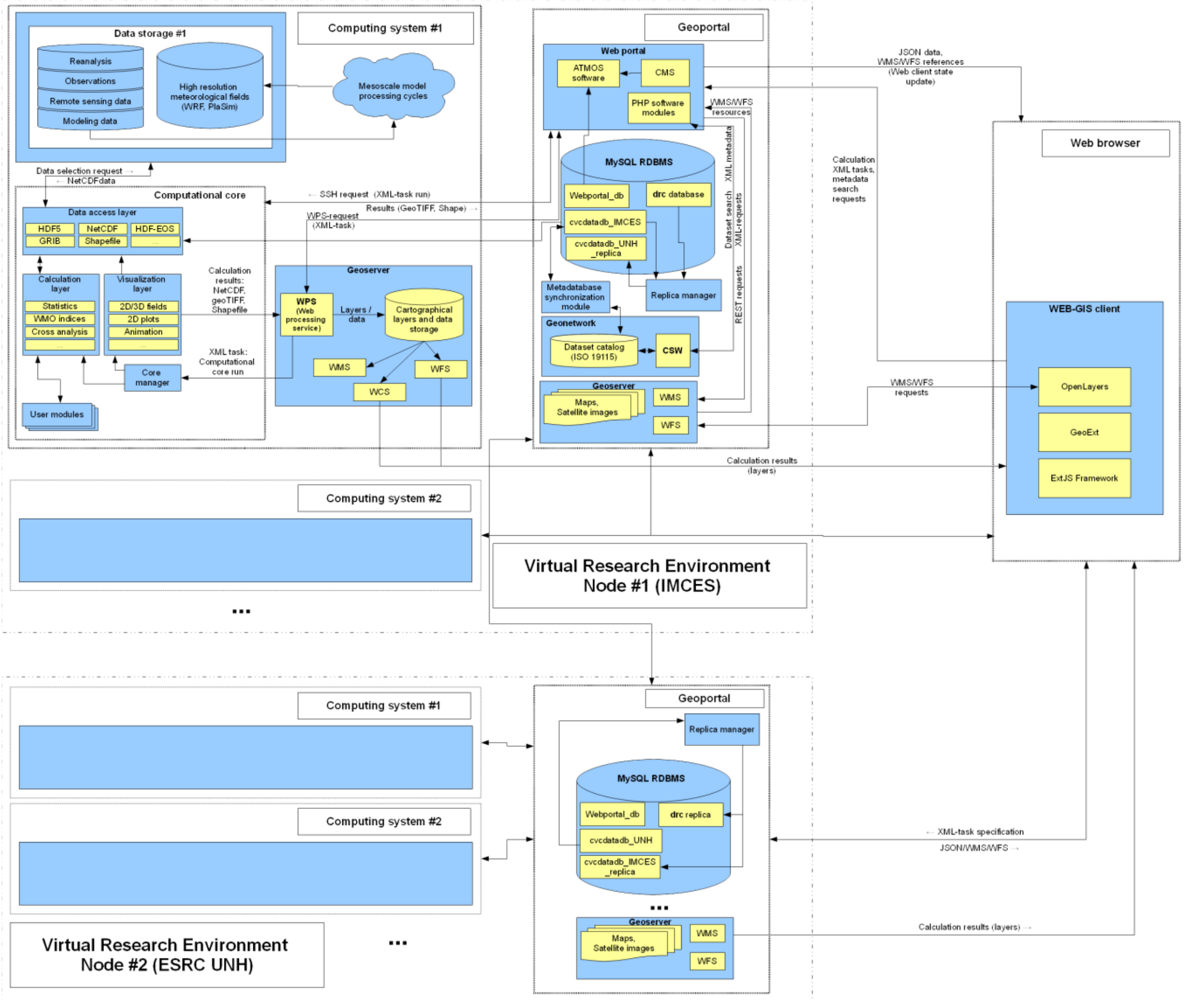

Figure 1. General architecture outline.

Environment Node #1 and Node #2 are displayed as examples) with corresponding computing systems (servers) and data storage systems containing geospatial datasets basically in the NetCDF format, including numerical modeling and reanalysis data, as well as meteorological observation data grouped according to thematic characteristics. Each node runs a specialized software, such as an SDI geoportal, cartographical web services (WMS/WFS) based on Geoserver (http://geoserver.org/), a MySQL database of technical metadata describing the geospatial datasets available for the node, and geospatial data processing services (WPS) based on a modular computational backend realizing statistical processing functionality and thus providing access to, as well as complex processing and analysis of large datasets, results of visualization and export into files of the required formats. Within the general architecture framework, the nodes' MySQL metadata databases and their interconnections are of fundamental importance. The metadata database of the central Node #1 (drc_database) contains descriptions of all the SDI nodes available in the environment as a whole, while each of them, in turn, is connected to a set of computing servers (Computing system #1, ...), data storages and, finally, the available datasets. The SDI node descriptions consist of such fields as name, title, URL, and node's geoportal MySQL database connection parameters. All other secondary nodes contain a central metadatabase replica that allows geoportal applications to check the information about datasets available for processing on other nodes, as well as computing modules for their processing. The central metadata database can be administered by a senior system administrator, while the nodes' administrators have "read-only" access rights for it and its replicas. Due to this approach, the VRE users are provided with information about all the geographically distributed datasets globally accessible, regardless of the availability of direct network connections to the remote databases at any time. Obviously, a standalone node is characterized by the presence of: at least one high-performance computing system with a set of data storage attached; an SDI geoportal; and a computational backend.

The architecture at a single VRE node level may be represented by three basic tiers:

1. Tier of data and calculation procedures providing analysis and corresponding cartographical and visualization services.
2. Tier of middleware representing the geoportal including metadata database and server-side web applications.
3. Tier of client-side web-applications providing the GIS functionality (Web-GIS client, etc.) available to the end user.

2.1. Tier of data and calculation procedures

The tier of data and calculation procedures provides analytical processing of geospatial datasets, as well as cartographical and visualization services. The datasets are located on data storage systems either in the form of collections of netCDF files or PostGIS databases.

2.1.1. Modular computational backend. Within the framework of the data tier, the key software component is a specialized computational backend representing a standalone software toolset. The computational backend of the VRE is based on the GNU Data Language (GDL, http://gnudatalanguage.sourceforge.net/) and Python (http://python.org) providing integral geospatial data mathematical processing and visualization functionality, as well as API to work with NetCDF/HDF, Esri Shapefile data files, and PostGIS databases. The computational backend contains the following key components:

1. Data Input / Output component
2. Geospatial data processing component
3. Graphical visualization component

Depending on the result type required, the backend generates the following file formats: GeoTIFF, ESRI Shapefile, Encapsulated PostScript, CSV, XML, NetCDF, and float GeoTIFF. The computational backend manager controls the general application workflow, thus providing data processing modules calls and their interaction. Computational pipelines are prepared with a special XML file describing a computational task generated by a Web GIS client according to the actions of the end user.

2.1.2. OGC web services. The geospatial data service system is supposed to provide launching of computational procedures and visualization of their results. The cartographical web services to be realized as an integral part of the system implement the differentiation of the user's access rights and provide functional capabilities to work with raster (GeoTIFF) and vector (Shapefile, PostGIS) geospatial data using OGC WMS, WFS, and WPS based on the Geoserver and GeoWebCache software (http://geoserver.org). In general, the cartographical Web processing service provides a standard HTTP interface for remote configuring and launching the data processing modules and presenting the results in binary and graphical formats according to the end user's instructions. Additionally, the task of transforming graphical calculation results along with SLD styles into corresponding Geoserver cartographical layers and styles, as well as their publishing as WMS and WFS resources, should be solved. As successful criteria of the resource publishing, the following might be used:

1. The presence of the WMS/WFS resource descriptions of calculation results in OGC GetCapabilities replies
    a. Obtaining of graphical legends in reply to the GetLegendGraphic WMS request
    b. Correct visualization of the cartographical layer with the required SLD style applied when using any client GIS application (OpenLayers, uDIG, QGIS)
2. Classical differentiation of the Geoserver user access rights for view, creating and modifying of WMS and WFS resources must be provided with all the users, except for the administrator group members able to work with their own cartographical resources.

Thus, the services developed can be used either while working in the standard GIS environment (uDIG, QGIS, etc.) or in specialized web applications deployed within the geoportal framework.

2.2. Tier of middleware

According to the Spatial Data Infrastructure concept, the geoportal serves as a gateway to the geospatial data and derived products, and provides functional capabilities, such as search for geoinformation resources using catalogs, data access, and sampling, as well as management of the geoprocessing services and cartographical processing and visualization applications [9]. The virtual Research Environment geoportal represents a middleware consisting of:

1. A metadata database describing the available geospatial datasets.
2. A metadata catalog based on GeoNetwork to describe the geoinformation resources used in climate research using an ISO 19115 standard.
3. A web portal implementing: web applications as reusable PHP modules and interconnections with the OGC web services, target computing systems, and a metadata database.
4. A central Geoserver repository containing basic cartographical layers, such as state and geographic boundaries, land cover and climate zone maps, river basins, etc.

2.2.1. Metadata database. At a single VRE node, its metadata database [10] presents multilingual descriptions of climatic geospatial datasets, with processing routines in the form of dedicated software modules. Thus, there are two major parts of the MDDB:

1. One part contains information about all datasets available for analysis. It is used to locate data files and to provide metadata on computational backend request. This part of the MDDB describes the geospatial datasets available for the node: their spatio-temporal characteristics, lists of meteorological parameters and physical locations, which provides fast location of data files on storage systems by the computational backend.

2. The other part contains descriptions of processing routines represented as various pipelined call sequences of dedicated computational backend modules and their run options. Since some data analysis routines are designed to process only specific meteorological parameters, the connections between the computing modules and data arrays are set in the MDDB. It also contains descriptions of the corresponding GUI elements required to provide the module calculation.

In Figure 1 the metadata database for a particular node of the VRE has such name like cvcdatadb_<# of the VRE node> (cvcdatadb_1 for the VRE node # 1, etc.), and except for the relations describing datasets and the corresponding routines, it has the following additional tables:

– computing_system_type: types of the available computing systems (local, remote, OGC WPS);
– computing_system: list of descriptions of the computing systems available for the VRE node.

All the other VRE nodes contain replicas of the considered metadatabase named cvcdatadb_<# of the VRE node>_replica, which provides the VRE users with the information about all geographically distributed datasets available within SDI regardless of the presence of Internet connection. Actualization of the replica is performed by the geoportal "Replica manager" component. The node's metadatabase can be administered by the local system administrator; the administrators of other nodes have only read access to it.

2.2.2. Web portal. A web portal complying with the general concepts and standards of development of an open source software providing OGC cartographical web services serves as a connection point between the SDI elements. It implements basic functionality, such as user's authentication, RDBMS connectivity, HTML templates usage, multilingual support, a content management system, and some other features [11]. The core part of the web portal was developed aiming at computational solution of scientific tasks in meteorology and climatology. Its main feature provides a unified API for web applications.

2.3. Tier of client-side web applications providing GIS functionality

The geoportal web applications comply with the conventional Boundless / OpenGeo architecture [12]. The architecture consists of three basic tiers (data, application server, and graphical user interface) and implies the use of the following open source software: PostGIS, Geoserver, OpenLayers, and GeoExt/ExtJS.

2.3.1. Web GIS client. The Web GIS client, as the main software instrument of the VRE desktop, consists of two basic software parts:

1. The server-side part as a set of web portal PHP applications implementing an open API for browser-based client software part and realizing the functionality of interaction with:
   a. GDL / Python backend in the form of a standalone computing core;
   b. WMS/WFS/WPS cartographical services based on the Geoserver software.

   This part has a secondary meaning, since it does not contain an application business logic, while providing a limited set of GUI manipulation procedures via the standard HTTP interface.
2. The front-end part representing web mapping client developed according to a "single page application" technology based on JavaScript OpenLayers (http://openlayers.org/) and GeoExt (http://geoext.org/) libraries and providing a user-friendly interface similar to one of the well-known desktop applications, such as uDIG, QGIS, and so on.

There are two main server-side web applications:

1. Manager of cartographical layers to administer Geoserver raster and vector layers via the standard REST (Representational State Transfer) interface. It also includes a processing unit for handling the computational task obtained from the user in XML format and the computational backend launching and tracking module.

2. Metadata manager working with geospatial dataset formalized descriptions developed to interact with the MySQL metadatabase and providing a presentation of the NetCDF metadata in JSON format optimized for processing by the browser based GIS client. The technique applied allows the client side structured data caching and updating of the graphical user interface elements.

The cartographical front-end web application (web mapping client) for work with the archive of geospatial NetCDF datasets contains 3 basic tiers [13]:

1. Tier of NetCDF metadata retrieved from MySQL database and represented in JSON format
    a. Middleware tier of JavaScript objects implementing methods to work with:
    b. JSON metadata
    c. XML file containing configuration of selected processing routines, input and output formats (XML task)
2. OGC WMS/WFS cartographical services
3. Graphical user interface tier representing JavaScript objects realizing general application business logic

Web GIS client must comply with the general INSPIRE requirements on geospatial data visualization and provide computational processing services launching to support the solving of tasks in 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.

3. The prototype system

An element developed for the Virtual Research Environment i**s** supported by two standalone nodes interconnected by telecommunication channels, which form a hardware complex providing distributed access, analytical processing, and visualization of large geospatial datasets using cloud computing techniques. An access to the VRE geospatial data services is provided via the standard web browser from any workstation with an Internet connection.

Within the framework of Web GIS client OGC WMS/WFS services the support is realized as an integrated component. Its goal is the interaction with the system end user during the digital map creation process, Geoserver connectivity, and accurate presentation of the calculation results using modern Web and GIS technologies. The component provides such basic elements as:

1. Window displaying lists of the available raster and vector layers accessible via WMS/WFS, as well as their preview
2. Module handling requests adding new layers to the map
3. Layer legend display panel
4. Module to work with WFS layers including element selection and attribute table retrieval
5. Window displaying results of the user GetFeatureInfo request (geographic position, parameter values for the point selected)

A general algorithm implemented by the component supporting OGC web services is presented below:

a) While creating a thematic map using Web GIS client the user opens a dialog window listing the WMS/WFS layers available
b) The user selects the layers to be added to the map
c) An HTTP request is formed and sent to the corresponding cartographical web service

d) According to the HTTP reply obtained, the web interface graphical components are modified and selected layers are displayed on the map in the correct order
e) The user analyzes the map and attribute tables of the point of interest
f) The user gets either a raster (GeoTIFF) or vector (GML, JSON) presentation of the selected cartographical layer and corresponding source data in NetCDF format.

The Web GIS client of the prototype was developed using DHTML, PHP, and JavaScript languages. The JavaScript library implementing Web GIS client interface components is based on OpenLayers, GeoExt and ExtJS libraries, and OGC services provided by Geoserver. Graphical interface performs two main functions: providing functional capabilities for editing calculation task XML file, and a visual presentation of the cartographical information for the end user. The basic elements of the graphical user interface are (Figure 2):

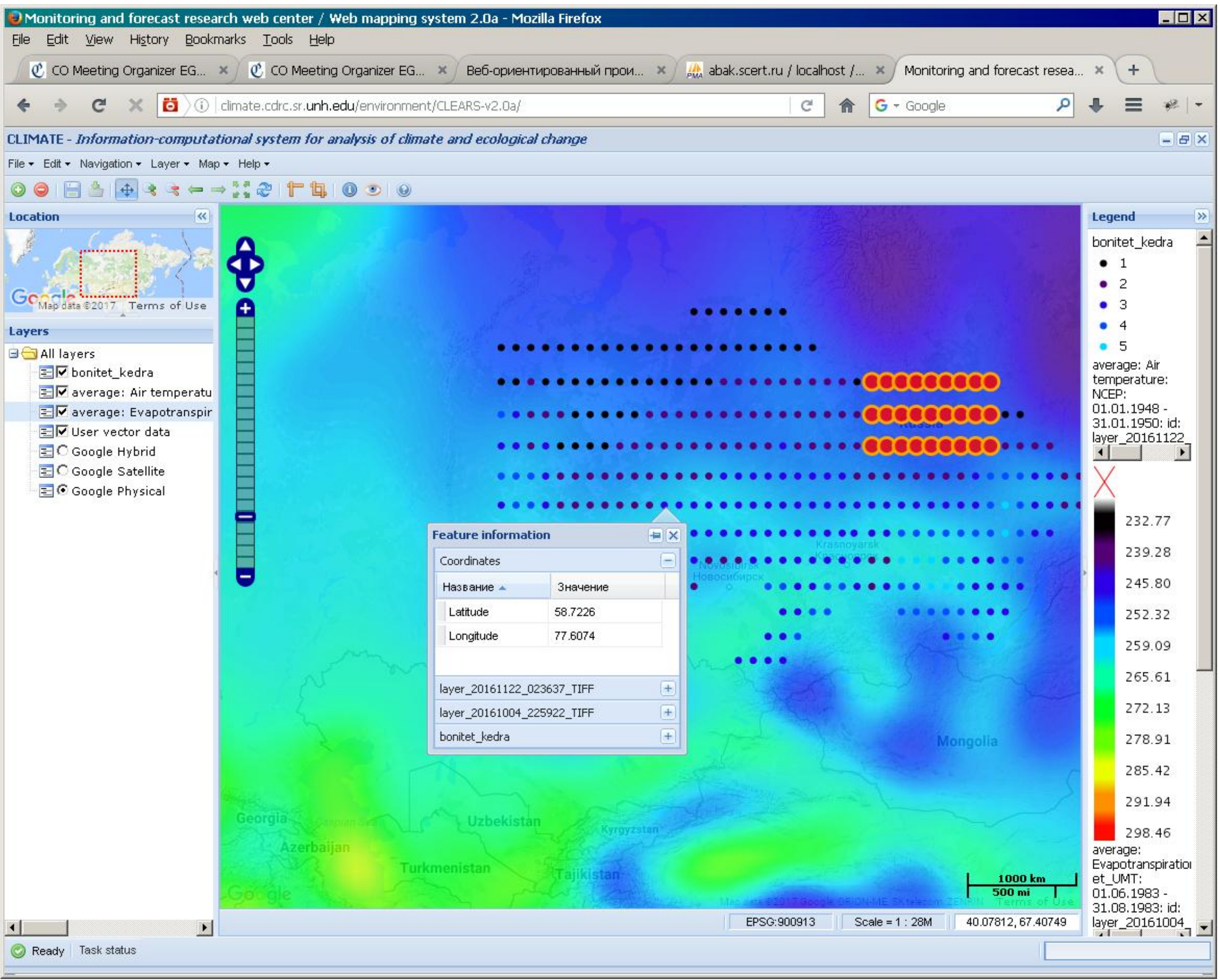

Figure 2. Web GIS client interface.

- Panel displaying user cartographical layers on the map, Google maps being used as the base layer by default, but there is a possibility to set an arbitrary base layer including a new one 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 (Figure 3) based on the results of computational processing of geospatial datasets available to the system

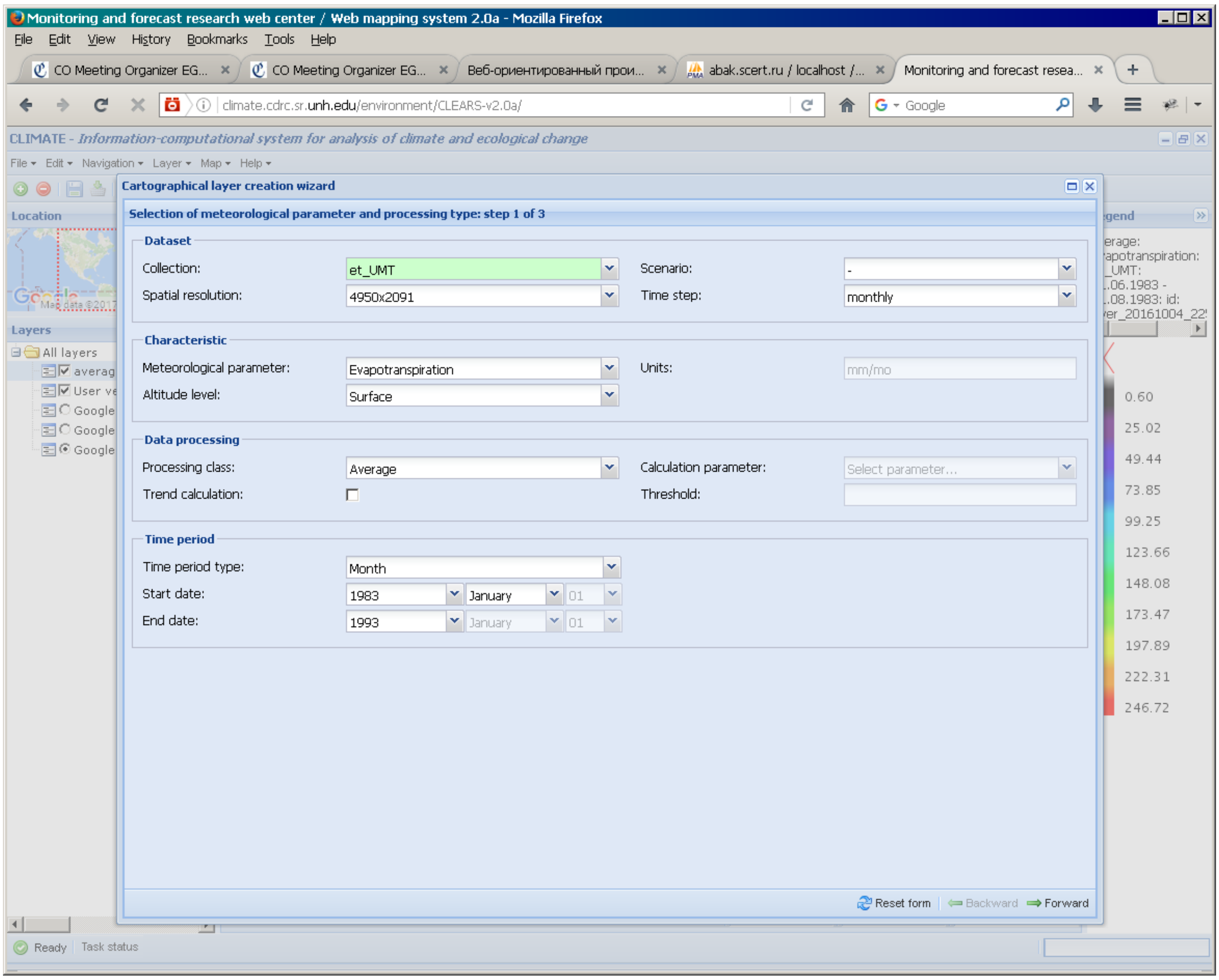

Figure 3. Layer creation wizard.

A cartographical layer creation wizard was developed based on the following criteria:

1. Efficiency of filling out the graphical user interface interactive forms;
2. Optimization of the process of creating and editing of an XML file describing the selected calculation configuration.

The toolbar, application and context menus contain mouse and keyboard event handlers which uniquely determine the Web-GIS behavior depending on the user's actions with the execution context applied. It should be noted that the geospatial data cartographical services based on the Geoserver software can be used in the Web-GIS client considered, as well as in the standard desktop GIS applications.

4. Conclusions

The VRE distributed architecture presented allows easy addition of new nodes, computing and data storage systems. It provides a solid computational infrastructure for regional climate change studies based on modern Web and GIS technologies. The metadata database has improved the system functional capabilities in terms of extending the geospatial dataset archives and the statistical processing routines, as well as providing computational resources as web services. The front-end developed complies with the general GIS requirements, while the computing backend technical realization allows a unified geospatial dataset processing regardless of the client application type. The working prototype presented has been shown to be effective in the process of solving real climate change research problems and disseminating the investigation results in cartographical form. It has been developed for decision makers and specialists working in affiliated sciences, and provides accurate environmental characteristics required in the studies of economic and social consequences of global climate change at the regional level.

Acknowledgments

The authors thank the Russian Science Foundation for the support of this work under grant # 16-19-10257.

References

[1] 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 E Bocher and M Neteler (Heidelberg: LNGC, Springer) pp 247–61
[2] Yakubailik O E 2007 Geoformation geoportal *J. Computational Technologies (Special Issue 3)* **12** 116–25
[3] 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
[4] Frans J M van der Wel 2005 Spatial data infrastructure for meteorological and climatic data *Meteorol. Appl.* **12** 7–8
[5] 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) 95–101
[6] Janssen K 2010 The Availability of Spatial and Environmental Data in the European Union: At the Crossroads Between Public and Economic Interests *Kluwer Law International*, p 617
[7] 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.
[8] 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) 128–30
[9] Koshkarev A V 2008 Geoportal as a tool to control spatial data and services *Spatial data* **2** 6–14
[10] Okladnikov I G, Gordov E P and Titov A G 2016 Development of climate data storage and processing model *IOP Conf. Series: Earth Environ. Sci* **48** 012030
[11] Gordov E P Lykosov V N and Fazliev A Z 2006 Web portal on environmental sciences "ATMOS" *Adv. Geosci.* **8** 33-38

[12] Becirspahic L and Karabegovic A 2015 Web portals for visualizing and searching spatial data *Information and Communication Technology, Electronics and Microelectronics (MIPRO)* 305-11 doi: 10.1109/MIPRO.2015.7160284
[13] E Gordov, A Shiklomanov, I Okladnikov, A Prusevich and A Titov 2016 Development of Distributed Research Center for analysis of regional climatic and environmental changes *IOP Conf. Series: Earth Environ. Sci* **48** 012033