Some Aspects of Development of Virtual Research Environment for Analysis of Climate Change Consequences

© Gordov E.P.[1], © Okladnikov I.G.[1], © Titov A.G.[1], © Fazliev A.Z.[2]

[1] Institute of Monitoring of Climatic and Ecological Systems of the SB of the RAS, Tomsk 634055, Russia

[2] Institute of Atmospheric Optics of the SB of the RAS, Tomsk 634055, Russia
gordov@scert.ru, oig@scert.ru, titov@scert.ru, faz@iao.ru

Abstract

We present general structure, elaborated approach and preliminary results of a project aimed at analyzing climate data and predicting impacts of climate change on the environment. One of the project objectives is provision of 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 the effects of climate change. Its ultimate goal is the development of hardware and software prototype of a virtual research environment (VRE) for the climate and environmental monitoring and analysis of the impact of climate change on socio-economic processes of local and regional scale. This environment will integrate the known and new sets of climate data, software implementations of classic and new methods of statistical analysis of large data sets. It will provide the opportunity to scientists and decision-makers to use different geographically distributed spatially referenced data, processing resources and services through a web browser by integrating distributed systems that store, process and provide information via Geoportal.

The first project results, namely the scheme of large sets of geospatial climate data storage together with supporting metadata database architecture and designed an intuitive graphical user interface, are presented. Also an approach to solution of the reduction problem in quantitative climatology is discussed.

1 Introduction

Understanding of complex mechanisms of the changing climate and its effects on the environment requires generating and analyzing continuously increasing volume of observation and modeling georeferenced data [1]. Increase of diversity and volume of spatial data sets makes it impossible to collect them in a single physical archive to be processed and analyzed on the basis of traditional "working place" approaches [2]. At the same time petabyte level growth of environmental data volumes and necessity to store, search, share, transfer, process, analyze, and visualize those made the area a field for approaches and tools developed in the recently appeared data intensive application domain [3-7]. Examples of data-intensive areas are given in Ref. 6. Since volume of weather and climatic data collected and produced amounts up to petabytes they correspond to the 5V (Volume, Velocity, Variety, Variability, Veracity) model [7] and fall within the definition of "big data". To indicate the presence of geospatial dependence of physical quantities it is more correct in this case to use term "Geospatial Big Data" [8].

For comprehensive usage of large sets of georeferenced meteorological and climatic data it is necessary to create a distributed software infrastructure [9, 10], based on the spatial data infrastructure (SDI) approach [11]. SDI geoportal [12, 13] is considered as a single point that provides functionality of searching geographic information resources, retrieval of the samples according to the specified parameters (data access functionality) as well as processing and cartographical visualization services along with corresponding client applications [14]. Currently, it is generally accepted that the development of client applications as integrated elements of such infrastructure should be based on the usage of modern web and GIS technologies [15, 16, 17, 18]. According to general requirements of the INSPIRE Directive to geospatial data visualization [19], 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 basic functionality of a standard GIS should be provided.

At present there are a number of information systems and services that provide similar functionality. GeoBrain Online Analysis System (GeOnAS) provides access to satellite data (NASA, USGS) via OGC services based on the GRASS GIS open source software, and has a developed web interface based on the DHTMLX library (http://dhtmlx.com/).

ncWMS service [20] is an implementation of OGC Web Map Service (WMS) for geospatial datasets represented in NetCDF format. It is actively used for data visualization within the SDI geoportals, the limiting factor being its minimal support by standard GIS.

Proceedings of the XVIII International Conference «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2016), Ershovo, Russia, October 11 - 14, 2016

Unidata THREDDS (http://www.unidata.ucar.edu/ software / thredds / current / tds / TDS.html) provides access to geospatial data and metadata through OPEnDAP protocols, WMS, Web Coverage Service (WCS) and others. The product also provides data subsetting functionality by using ncWMS for visualization of results.

Open distributed architecture Boundless / OpenGeo is widely used for complex geo-information applications development [21, 22]. It consists of three layers (data, application server and graphical interface), and employs the following open source software:

1. Web mapping software Geoserver + Geowebcache (http://geoserver.org), implementing OGC WMS, WFS, WPS services.
2. OpenLayers JavaScript library (http://openlayers.org/) which provides basic functionality of a thin Web GIS client.
3. GeoExt / ExtJS JavaScript library [23] for development of client web applications with an intuitively clear interface.

In this paper we outline the approach chosen for carrying out the Russian Science Foundation project [24] and specific results obtained during its first phase. In particular, we discuss the developed scheme of large sets of geospatial climate data storage and creation of relevant metadata database, as well as the designed Web GIS client with the intuitive graphical user interface. Also we describe an approach to solution of reduction problem in quantitative climatology as a basis for the planned construction of the knowledge base representing properties of geophysical data to be used in a topic-based decision support system.

2 Project approach and outlines

2.1 Objectives and approach

The proposed project is aimed at provision of specialists working in affiliated sciences focused on impact assessment, adaptation strategies, and other climate related activities and decision-makers with accurate and detailed climatic characteristics and reliable, affordable tool for their in-depth statistical analysis and studies of effects of climate change in the selected area. To reach this objective a hardware and software platform prototype forming topic based VRE for comprehensive study of ongoing and possible future climate characteristic changes and analysis of their impact on regional environment will be developed. It should provide reliable climatic information required for the study of economic, political and social consequences of global climate change at the regional level.

The large project tasks that constitute four work packages (WP) combining sets of particular thematic tasks will be solved to reach this goal.

2.2 WP 1. "Preparation of geo-referenced data sets"

WP 1 is dedicated to the creation of the data archive reflecting the detailed picture of the ongoing climatic processes and their future projections for the period 1960-2100 years and its integration into VRE under development. The archive should comprise detailed data on climatic characteristics of the study area, detailed description of the surface and its changes and form a solid foundation required to study the response of economic, social and political processes on on-going and future climatic processes.

2.3 WP 2 "Improvement of methods of analysis of climate change"

WP 2 is dedicated to the development of new methods of statistical analysis of climatic processes and their implementation into computational VRE research and analysis tools. This task should provide the developed VRE with functionality to perform modern statistical analysis of spatio-temporal climatic characteristics including their extreme manifestations.

2.4 WP 3 "Functionality of the VRE prototype"

WP 3 will be devoted to enhancing the functionality of the VRE prototype in order to create the intuitively clear and reliable tools for investigations of regional social, economic and political consequences of climate change. The result of this task will be development of opportunities for experts in the field of environmental sciences to carry out the study of the impact of climate change on relevant processes in the study region without getting a second education in computer science and their applications.

2.5 WP 4. "Case study of Western Siberia and dissemination of project results"

This task should demonstrate the VRE potential to its end user, namely to local stakeholders, decision makers and experts in the field of economic, social and political sciences. In particular, it will be shown that the elaborated VRE prototype allows one to study the climate change impact on processes in the region without mastering modern statistical analysis of geo-referenced data and advanced programming. To this end a topical DSS using ontology knowledge bases will be developed.

Specific results of this work package are detailed maps of different characteristics of extreme events and trends (with estimates of statistical significance), including estimates of return periods for certain disasters and uncertainty estimates for different types of extreme events. These maps will be available in the VRE providing easy access to interested consumers.

3 The first results

Part of tasks of the work packages WP2 and WP4 has been performed on the first stage of the project carrying out. Approaches used to solve related problems and results obtained are described in this section.

3.1 Metadata database

3.1.1 Data storage scheme

Currently, two major approaches to storing geospatial data are used: geospatial databases and file collections. The database approach utilizes relational and non-relational, local or distributed spatial databases such as

Apache HBase, Esri Geodatabase, Paradigm4, SciDB, etc. This approach requires inserting all data into spatial database before their actual use, which is quite disk space- and time-consuming operation. The file collections approach relies on storing data as a collection of data files in file system directories. Usually, self-describing formats are used for storing geospatial data. It was shown [25] that retrieval of data chunks larger than 40 Mb from a spatial database is less effective than from a simple data files collection. On the other hand, file-based approach provides fast data extraction in most cases, without a redundant preprocessing. However, this approach requires development of additional software layer to provide API adapters for storing and processing of distributed file collections. In our case simplicity and flexibility of file-based data storage approach prevailed over the other one.

Network Common Data Form (netCDF) was chosen as a major file format for most geospatial data in our data archive. This format is formally acknowledged by scientific institutions (including UCAR) and OGC as a standards' candidate for storing geospatial data and stimulating data exchange. 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 a root location of data collections, <*data collection name*> is a name of the directory containing a single data collection, <*spatial domain resolution*> is a name of the directory containing data with the same horizontal resolution, <*time domain resolution*> is a name of the directory containing data with the same time step. All data files (sometimes grouped in subdirectories) are located deeper in the hierarchy. Names of files and subdirectories are not regulated and determined by the individual specifics of a particular data set. Every data file contains one or more multi-dimensional arrays of meteorological parameters.

3.1.2 Architecture of metadata database

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

This database contains spatial and temporal characteristics of available geospatial datasets, their locations, and run options of software components for data analysis. Here the following terminology is used. "Dataset" is 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 simulation or observation conditions (if applicable). It is represented by a collection of netCDF files containing the same set of meteorological parameters. It is necessary to distinguish the term "parameter" ("meteorological parameter") and "variable". Meteorological parameter is the name of some meteorological characteristic: temperature, pressure, humidity, etc. Variable is a unique name of a multidimensional array in a netCDF. Names of meteorological parameters are standardized. In contrast, the names of variables in different datasets could be different, and usually depend on preferences of an institution which produced them. Along with data, netCDF files contain horizontal, vertical and time domain grids.

"Data collection" is a collection of datasets created by the organization within the specific project, but specified on different spatial and/or temporal grids, or for different scenarios. The collection may consist of one dataset.

Tables in MDDB are divided into "technical" and "interface". Technical tables contain data intended for computing software components. Interface tables hold string multilingual content for the graphical user interface.

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. Information about all available for analysis datasets is stored in the first part of MDDB. It is used to locate data files and to provide 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 graphical user interface (GUI). Second part of MDDB contains description of various call sequences and their options. Since some data analysis routines are designed to process only specific meteorological parameters, connections between computing modules and data arrays are set in MDDB.

3.2 Web-GIS client

Boundless/OpenGeo architecture was used as a basis for Web-GIS client development.

A cartographical web application (Web-GIS client) for working with archive of geospatial NetCDF datasets contains 3 basic tiers [26]:

- 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:

1. Objects with the structure conforming to the corresponding metadata database relations, for instance, object of measurement units.
2. 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:

1. Efficiency of filling out graphical user interface interactive forms;
2. 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 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:

1. Loading and updating of metadata JSON objects using AJAX technology
2. Creating, editing, serialization of XML calculation task object
3. Launching and tracking the task execution process located on the remote calculation node
4. Working with WMS/WFS cartographical services: obtaining the list of available layers, presenting layers on the map, export layers into various formats according to user request, obtaining and presenting the layer legend with the selected SLD style applied.

3.2.3 Graphical user interface

The tier is based on the 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). Graphical interface performs two main functions: providing functional capabilities for editing XML task file, and 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 (Fig. 1):

1. Panel displaying user cartographical layers on the map. Google maps are used as a base layer by default, but there's a possibility to set an arbitrary base layer including newly created by the user
2. Layer tree allowing to toggle layer display
3. Layer legends display panel
4. Map information panel (scaling, cartographical projections, cursor geographical coordinates)
5. Application general status panel
6. Overview map panel
7. General application menu
8. Toolbar (adding/removing layer, saving NetCDF data, panning, map refresh, obtaining of information related to given geographical point, etc.)
9. Application context menu
10. 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 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 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 geospatial data cartographical services based on Geoserver software can be used in Web-GIS client considered as well as in standard desktop GIS applications.

3.3 An approach to solution of reduction problem in quantitative climatology

Collected, analyzed, consistent and systematized data of subject domains related to problems demanding application of decision support systems have to be represented both on terminological and conceptual levels. In our project the representation will be implemented in form of two groups of OWL-ontologies related to different subject domains using semantic web technologies. The first group contains ontologies which represent properties of geophysical (including climatological) data. The second group contains OWL-ontologies characterizing a few simple tasks of economic and social domains demanding DSS usage. Different approaches to solution of reduction problem in implementation of quantitative climatology ontologies are considered now.

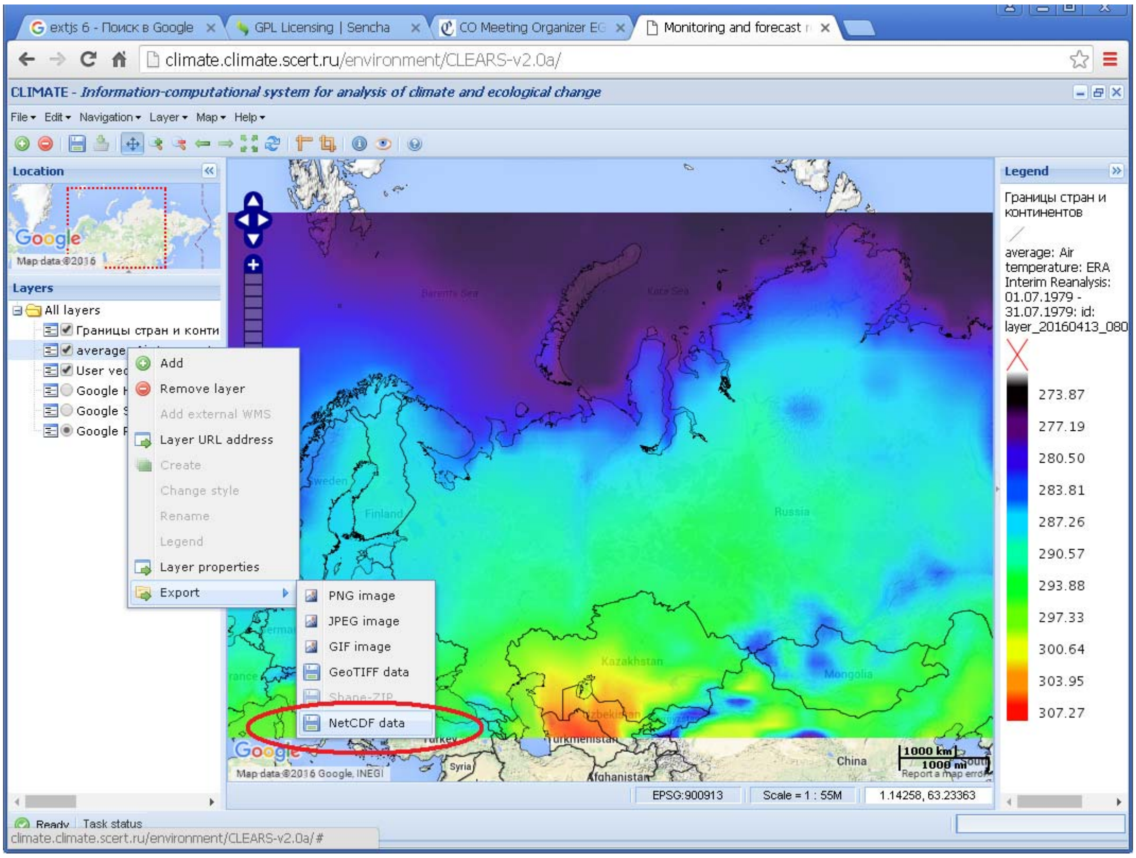

Figure 1 Graphical user interface of the web-GIS client. Exporting layer into NetCDF format is shown

4 Conclusion

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. Designed metadata database solves three main tasks: 1) provides content for the graphical user interface; 2) provides to geoportal information needed to generate the correct task file for the computational core; 3) contains information on the structure and arrangement of the data sets allowing computational core to read and process them efficiently. The use of this database organizes information on available data sets, facilitates the automatic retrieval of data files and improves the scalability and flexibility of computations.

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

First application of the developed metadata database and user interface showed that their combined usage facilitates expanding of a set of data archives available for analysis and adding new statistical processing procedures [27].

Results obtained show that the developed VRE and tools would be useful for decision makers and specialists working in affiliated 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 reliable climate related characteristics required for studies of economic, political and social consequences of global climate change at the regional level.

Acknowledgements

The authors thank the Russian Science Foundation for the support of this work under the grants 16-19-10257, 16-07-01028.

References

[1] Lykosov V.N., Glazunov A.V., Kulyamin D.V., Mortikov E.V., Stepanenko V.M. Supercomputing Modeling in Physics of Climatic System. Moscow State University Publishing House, 2012, 402 p.

[2] Gordov E.P., Kabanov M.V., Lykosov V.N. Information-Computational Technologies foe Environmental Science: Preparation of Young Researchers. Computational Technologies, 2006. V.11, Special Issue 1, p. 3-15.

[3] MIKE 2.0, Big Data Definition. (http://mike2.openmethodology.org/wiki/Big_Data_Definition)

[4] Dan Kusnetzky. What is "Big Data?". ZDNet. (Электронный ресурс: http://www.zdnet.com/blog/virtualization/what-is-big-data/1708)

[5] Ashley Vance. Start-Up Goes After Big Data With Hadoop Helper. New York Times Blog. (Электронный ресурс: http://bits.blogs.nytimes.com/ 2010/04/22/start-up-goes-after-big-data-with-hadoop-helper/?dbk).

[6] Kalinichenko L.A., et al. Problems of access to data in Data Intensive Domain in Russia. Informatika i ee primeneniya, v. 10, p. 3-23, 2016.

[7] Hilbert, Martin. "Big Data for Development: A Review of Promises and Challenges. Development Policy Review". martinhilbert.net. Retrieved 2015-10-07

[8] Shekhar S. Spatial Big Data // Proc. AAG-NIH Symp. on Enabling a National Geospatial Cyberinfrastructure for Health Research. July 2012. Minneapolis. USA

[9] Gordov E.P., Lykosov V.N. Development of information-computational infrastructure for integrated study of Siberia environment. Computational Technologies, 2007. V.12, Special Issue 2, p. 19-30.

[10] Stefano Nativi, Mohan Ramamurthy, Bernd Ritschel. EGU-ESSI Position Paper. [Digital resource]. — Access : http://scert.ru/files/EGU-PositionPaper-final.pdf

[11] Steiniger S., Hunter A.J.S. Free and open source GIS software for building a spatial data infrastructure. / In: Bocher E., Neteler M., (eds.), Geospatial Free and Open SourceSoftware in the 21st Century, LNGC, Heidelberg, Springer, 2012a, p. 247-261.

[12] Koshkarev A.V., Ryakhovskii A.V., Serebryakov V.A. Infrastructure of distributed environment of storage, search and transformation of geospatial data. Open Education, 2010, No 5, p. 61-73.

[13] Krasnopeev S.M. Experience of deployment of key elements of special data infrastructure on the base of web-services. Proceedings of XIV All-Russia Conference “Internet and Modern Society”, Sankt Peterburg, Russia, 2001, p. 92-99.

[14] Koshkarev A.V. Geoportal as a tool to control spatial data and services. Spatial data, 2008, No 2, p. 6-14.

[15] Yakubailik O.E. Geoformation geoportal, Computational Technologies, 2007. V.12, Special Issue 3, p. 116-125.

[16] Dragicevic, S., Balram, S., Lewis, J. The role of Web GIS tools in the environmental modeling and decision-making process // 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs. Banff, Alberta, Canada, September 2 - 8, 2000.

[17] Frans J. M. van der Wel. Spatial data infrastructure for meteorological and climatic data // Meteorol. Appl. 12, 2005. Pp. 7-8. DOI:10.1017/S1350482704001471.

[18] Vatsavai, Ranga Raju, Thomas E. Burk, B. Tyler Wilson, Shashi Shekhar. 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. 2000. Washington, D.C., US., p. 95-101.

[19] Katleen Janssen. The Availability of Spatial and Environmental Data in the European Union: At the Crossroads Between Public and Economic Interests / Kluwer Law International, 2010, ISBN 9041132872, 9789041132871, 617 p.

[20] J.D. Blower, A.L. Gemmell, G.H. Griffiths, K. Haines, A. Santokhee, X. Yang. A Web Map Service implementation for the visualization of multidimensional gridded environmental data // Environmental Modelling & Software, Volume 47, September 2013, p. 218–224. doi:10.1016/j.envsoft.2013.04.002

[21] L. Becirspahic and A. Karabegovic. Web portals for visualizing and searching spatial data // Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on, Opatija, 2015, pp. 305-311. doi: 10.1109/MIPRO.2015.7160284

[22] I.G. Okladnikov, E.P. Gordov, A.G. Titov, T.M. Shulgina. Information-computational System for Online Analysis of Georeferenced Climatological Data // Selected Papers of the XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015), October 13 – 16, 2015, Obninsk, Russia / Edited by Leonid Kalinichenko and Sergey Starkov. CEUR Workshop Proceedings. Vol. 1536. P. 76-80.

[23] Shea Frederick, Colin Ramsay, and Steve Cutter Blades. Learning Ext JS. / Packt Publishing, 2008.– 299pp.

[24] RSF Project №16-19-10257 “Virtual computational information environment for analysis, evaluation and prediction of the impacts of global climate change on the environment and climate of a selected region” (http://rscf.ru/en/enprjcard?rid=16-19-10257).

[25] Santokhee, J. Blower, K. Haines. Storing and Manipulating Gridded Data In Spatial Databases // Reading E-science Center, University of Reading

[26] Titov A.G., Okladnikov I.G., Gordov E.P. Development of web-GIS for climatic change analysis and projections on the basis of services of spatial data processing and visualization. Proceedings of XXI Baikal All-Russia Conference "Information and mathematical technologies in science and control".

[27] Ryazanova A.A., Voropay N.N., Okladnikov I.G. Application of information and computing web system «Climate» for estimationn of aridity of South Siberia. // Proc. of International Conference and Early Career Scientists School on Environmental Observations, Modeling and Information Systems ENVIROMIS-2016, July 11-16, 2016, Tomsk, Russia. Pp. 358-362.