Electronic Earth – network environment of search, integration and analysis of geodata

Yu. M. Arskiy,[1] A. V. Veselovsky,[2] B. G. Gitis,[3] and A. N. Shogin[1]

[1] The architecture of the multi-user distributed geoinformation-analytical environment is considered. It includes informational subject portals, analytical methods and GIS-systems, tools for searching and viewing the information, the distributed system of metadata depositories and, at last, the unique storage of global geodata, providing the cartographical base for the concrete geoinformation projects. This environment makes essentially easy access of the scientists to the geodata and the analytical processing of them in online mode. The concrete example of establishing a multivariate link between gold ore deposits of the Kuril-Kamchatsky volcanic belt and the parameters of geological environment is listed. INDEX TERMS: 0525 Computational Geophysics: Data management; 0530 Computational Geophysics: Data presentation and visualization; 0545 Computational Geophysics: Modeling; KEYWORDS: distributed analytical network systems, GIS technologies, "Electronic Earth" programme.

Introduction

[2] The Russian Academy of Sciences (RAS) Presidium Program "Development of the Fundamentals of the Scientific Distributed Data-processing Environment on the basis of GRID Technologies" has determined the direction of research "Electronic Earth: scientific data resources and information-communication technologies". The RAS branches of geoscience, mathematics and information technologies took part in the work in 2004–2007. By the present time the project "Electronic Earth" has outlined the basic principles of functioning of the systems of data analysis, in the framework of developed and presently operating six web-portals using the modern GIS-, GRID- and WEB-technologies. On the basis of these principles the integrated online information field of a user was elaborated, comprising a set of instruments, analytical methods, descriptions and geophysical data, essential for applied and fundamental research in Earth sciences. It includes thematic data portals, GIS systems, methods of data search and data mapping, distributed system of meta databases and, finally, unique storage of global geophysical data, providing a basic mapping and analytical processing of concrete geographic information projects. For the first time an unparallel scheme of interaction between a researcher and a system of data analysis was designed. The system allows to deal with previously unresolved tasks on modeling and predicting geophysical objects and processes, due to the interdisciplinary character of data resources and scientific methods and thematically mosaic architecture of the network of portals.

[3] The system's experimental exploitation has begun, oriented at qualitatively new level of information support of science. According to the Presidium program, the data-processing segment of GRID in geosciences was formed in the framework of RAS. The project's results were reported at the Russian and international conferences [*Arskiy et al.,* 2007, 2008], and aroused great interest of scientists and experts.

Survey of Existing Distributed GIS Technologies

[4] It was understood long time ago that a scientist or even organization cannot accumulate all geophysical data, necessary for carrying out research, in a local network or in one computer. In particular, it is related to the fact that many geophysical data comprise a strong time component (seismic data etc). However, in reality projects have to be constructed using various sources. Let us examine in this connection the existing technologies using digital data

[1]All-Russia Institute of Scientific and Technical Information, Russian Academy of Sciences, Moscow, Russia
[2]Institute of Geology of Ore Deposits, Petrography, Mineralogy and Geochemistry, Russian Academy of Sciences, Moscow, Russia
[3]A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia
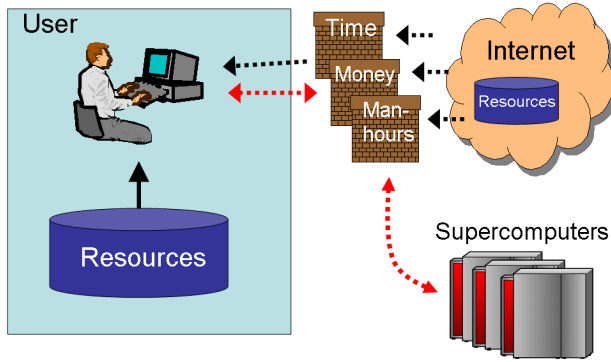
Figure 1. Local Technology.

in the web. The most frequently used technology is based on desktop GIS systems for visual analysis and geocoding (Figure 1). Obviously, in this case a researcher is doomed to i) long-time searching for the initial digital data; ii) possible transformation of obtained data and their adjustment to the required format and iii) purchase of expensive software and paying costly wages to programmers.

[5] Moreover, a use of high-power computers and GRID environment for resource-consuming operations is related to data transformation, preparing of a task and reverse transformation of obtained results for downloading data in a desktop GIS. As a result very little time is left for scientific research. On the other hand, recently in the Internet a number of portals have appeared, the purpose of which leaves no doubt that the problem of access of scientists to distributed data resources in geoscience, their search and usage became one of the most important. The majority of these portals is based on the scheme: OpenGIS servers – portal – OpenGIS client (Figure 2). The obvious shortcoming of this scheme is that it doesn't allow to make any analytically significant processing of data. The only project known to the authors that differs from this scheme is GEON network (www.geongrid.org). The present project uses in theory a GRID component, providing an opportunity to users to apply analytical computing methods. At the present time an

interface dealing with making GIS projects with analytical data processing is being prepared in the system.

Architecture of Electronic Earth Project

[6] Let us examine in detail the project architecture (Figure 3). For the interaction of the portals of "Electronic Earth" at the level of metadata an integrated structure of requests and responses for each component of the system is used. A portal of every system applies its own methods of storing and searching metadata, providing only a unified language of requests and responses.

[7] Whereas metadata is valuable by itself, solution of concrete research tasks requires interaction of portals and a user at a level of concrete geodata and analytical methods of their transformation and processing. Hence the analytical methods imply analytical GIS systems and autonomous computing methods, applied in GRID systems.

[8] The environment provides an opportunity to use a great number of accumulated geodata, distributed among the project portals and in the Internet. A user does not have to make any transformations. The data includes both descriptive data (publications, references etc.) and digital data (maps, databases, ...). Digital data can be downloaded both directly from the project servers (static or rarely changed data) and from proxy servers for quick-changeable data (e.g. operative catalogue of earthquakes). The system also provides the data, corresponding to the standards of OpenGIS consortium – WMS, WFS and WCS.

[9] The systems technological scheme includes a great number of tools, including

a) transfer from meta to geodata and analytical methods;

b) personification of results of search;

c) geodata transformation tools;

d) launch and control of fulfillment of a task in GRID system;

e) design of GIS project, launching of GIS system and maintenance of GIS project;

f) online analytical GIS systems.
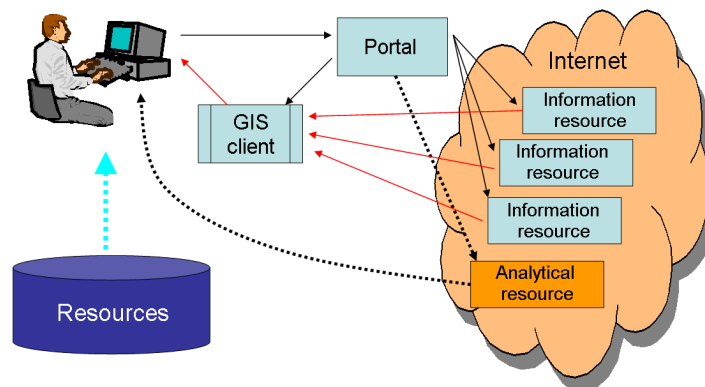
[10] Geoinformation environment "Electronic Earth" pro-
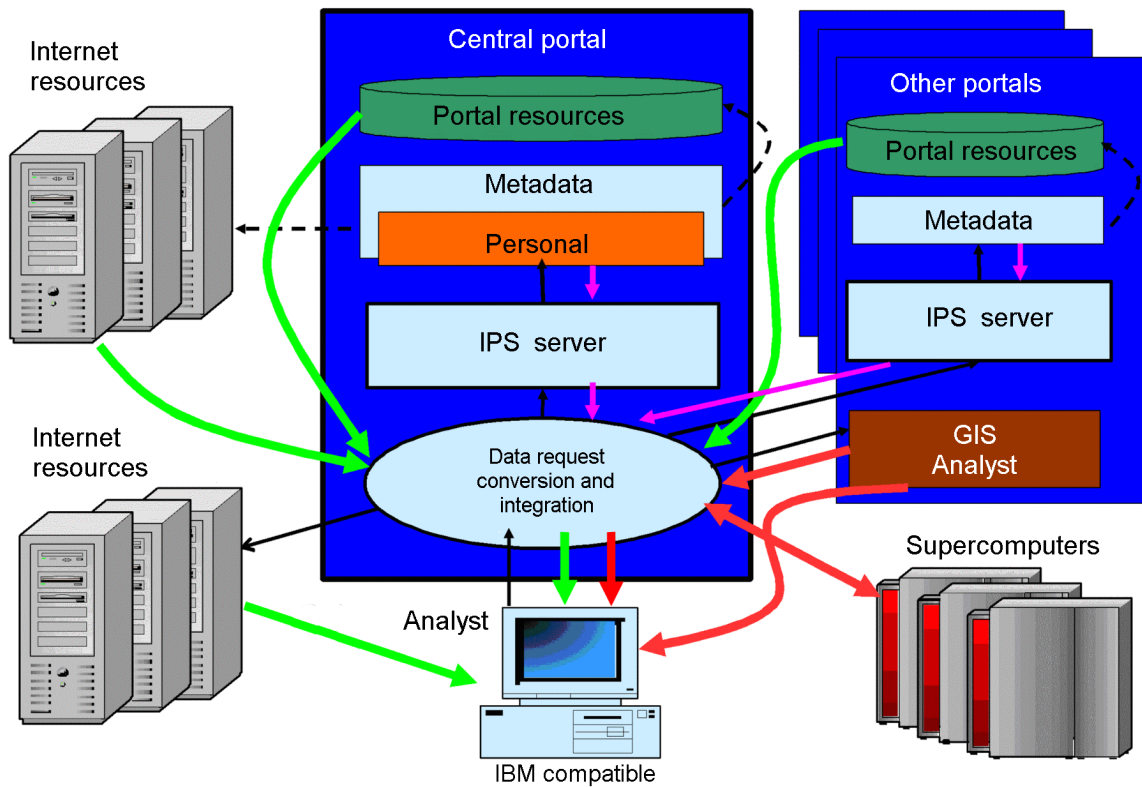

Figure 2. Network technology.

Figure 3. Details of the project architecture.

vides a universal combination of information resources and analytical methods with the help of a bi-component model of metadata. The first component is an "ordinary" metadata together with the type of data and reference to the second component of metadata. The structure of second component is fully determined by a type of data reflecting its parametrical component. An additional advantage of search and selection of data through the central portal is the use of a powerful system of classifiers, including the VINITI and GRNTI classifiers. In the nearest future specialized classifiers in the field of Earth sciences are expected to be located at the portal and the system of their automatic correlation will be introduced as well.

[11] Impressive methods of user's personification are developed, including his authorization, organization of a personal meta database, obtained as a result of distributed search, database of GIS projects and data processing in GRID systems. At that a user can confidentially integrate with the available data and analytical resources his personal data and program modules. These databases together with private geodata and program modules of a user and necessary mechanisms and methods of data integration and transformation comprise his individual information field.

Integral Information Environment of Project

[12] One of the main ideas of the "Electronic Earth"

project was to set up the methods of horizontal and vertical integration of data resources of all types and various ways of their analytical processing (Figure 4). Hence horizontal integration implies the integration of data distributed in the Internet and local resources, and vertical integration – on line application of analytical methods to data resources. The latter represents not only a trivial recalculation of cartographic layers but a multifold transformation of groups of data, their visualization and storage.

[13] Accordingly, the information field of the project "Electronic Earth" is constructed in two directions: horizontal – data location and vertical – type of data.

[14] The vertical component of the field comprises:

1. computing resources – servers, supercomputers, users' PC;

2. system resources – means of search and integration of information and analytical resources;

3. analytical resources – GIS, modeling and computational program systems;

4. diverse information resources – metadata, classifiers, publications, geodata etc.

[15] Resources of network environment and instrumental means can be distributed at

1. the Intranet servers and servers of a user's computer;

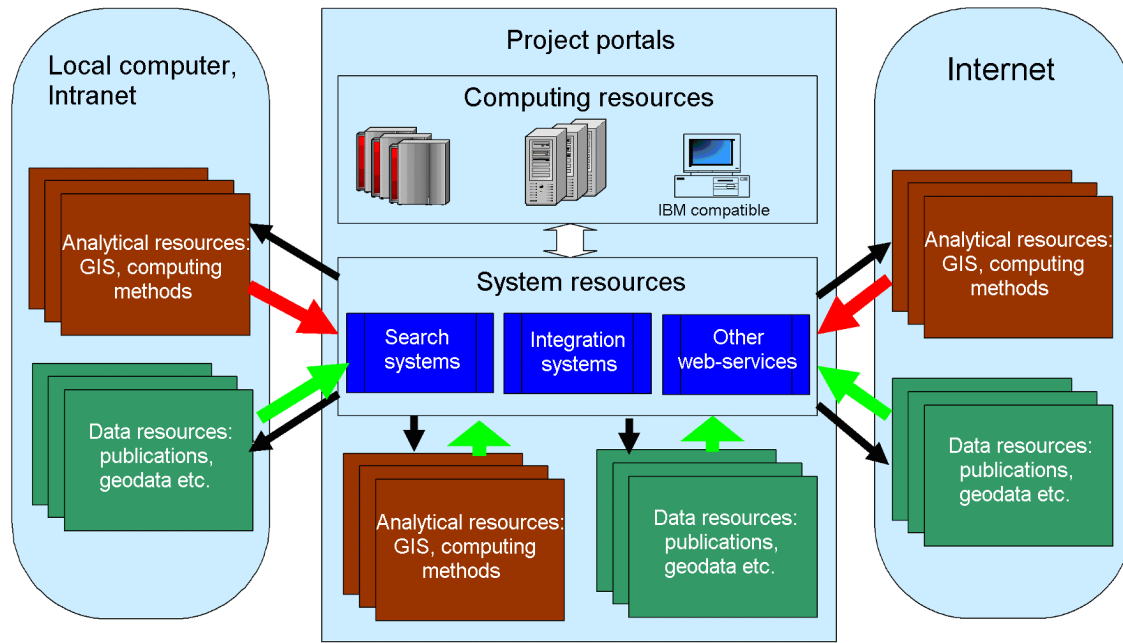2. the servers of the project network of portals;

Figure 4. Integration of data resources.

3. the Internet servers.

[16] The horizontal integration of data implies a creation of personal storage of a user's data together with potential on line transformers of formats.

System Development and Application

[17] At the present time seven portals have appeared in the Internet, supporting compatible protocols of data transmission and about 20–30 sites, operating through the main portals. The environment provides researchers with a great number of accumulated geodata, distributed among the project portals and in the Internet, and a user doesn't have to make any transformations. These data includes descriptive data (publications, references, etc) and digital data: main geographic (model of relief, river, lake etc.), geophysical (magnetic and gravitational anomalies etc) and geological digital data, embracing the whole globe. Digital data is provided to a user both directly from the project servers (static or rarely changeable data) and through proxy-servers for quickly changeable data (operative catalogue of earthquakes). The system allows using of data, corresponding to the standards of OpenGIS consortium – WMS, WFS and WCS. Moreover, a user is provided with various on line methods and algorithms [*Gitis and Yermakov,* 2004], part of which is implemented in a distributed GRID environment. Search and integration of data is implemented by a powerful system of classifiers supplied with programs of data correlation [*Arskiy et al.,* 1999]. Thus the system "Electronic Earth" possesses a universal information-analytical

field in the sphere of geoscience. Due to existing standards for metadata and protocols of exchange of data between the "Electronic Earth" project participants connecting to new portals and separate computers becomes an easy and inexpensive task for any organizations, dealing with research in the field of Earth Sciences.

[18] Let us discuss a concrete example of complex analysis, implemented in the environment "Electronic Earth" according to the data of the integral data bank of IGEM RAS.

[19] The task consisted of (A) establishing a multivariate link between gold ore deposits of the Kuril-Kamchatsky volcanic belt and the parameters of geological environment and (B) application of the obtained empirical dependency for predicting new gold ore deposits.

[20] The following data resources of the "Electronic Earth" environment and object-oriented resources of IGEM RAS were applied as initial data:

- Digital model of ground surface height in grid $2' \times 2'$;

- Types of rocks of the Cainozoe era;

- Types of rocks of the quaternary period;

- Types of rocks of the Neogene;

- Types of rocks of the Paleogene;

- Sub-volcanic intrusions of the Kamchatka link of the Cainozoe Kuril-Kamchatsky volcanic belt;

- Geological faults;

- Borders of volcanic belts;
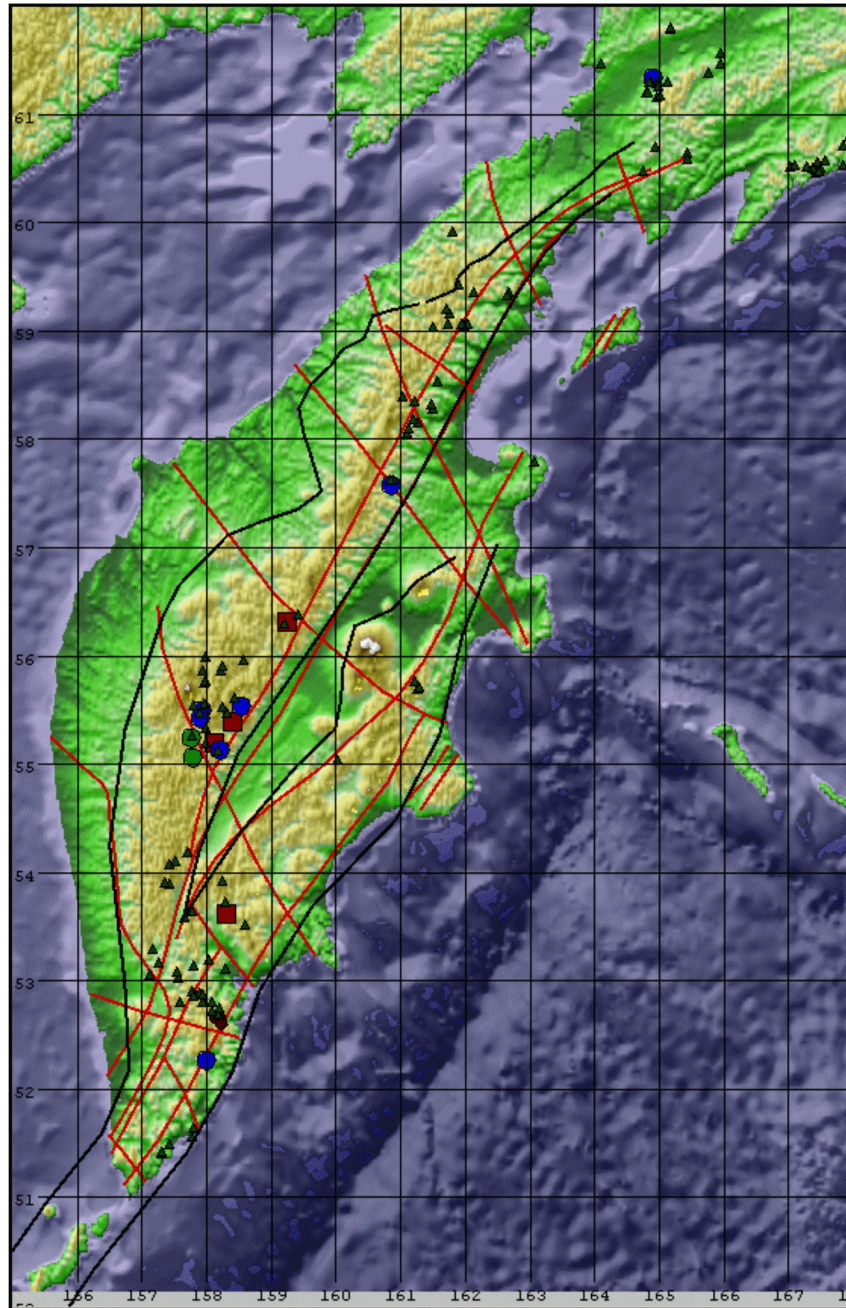
- Holocene volcanic constructions;

Figure 5. Map of the examined region. Red squares – gold deposits, green circles – polymetallic gold-silver with lead and zinc prevailing over copper, blue circles – gold-silver, dark-red squares – gold-quartz, black triangles – manifestations, not divided into formations.

- Pliocene volcanic constructions;

- Gold-silver deposits and Kamchatka manifestations:

  - Gold;

  - Pilymetallic-gold-silver with lead and zinc prevailing over copper;

  - Gold-silver;

  - Gold-quartz;

  - Manifestations, not divided into formations.

[21] A part of these data was available at a user's computer, the other part was taken from the project "Electronic Earth". As a result of these data loading into the project and launching of GIS applet "GeoProcessor" the data could be processed analytically. Figure 5 shows the map of the ex-
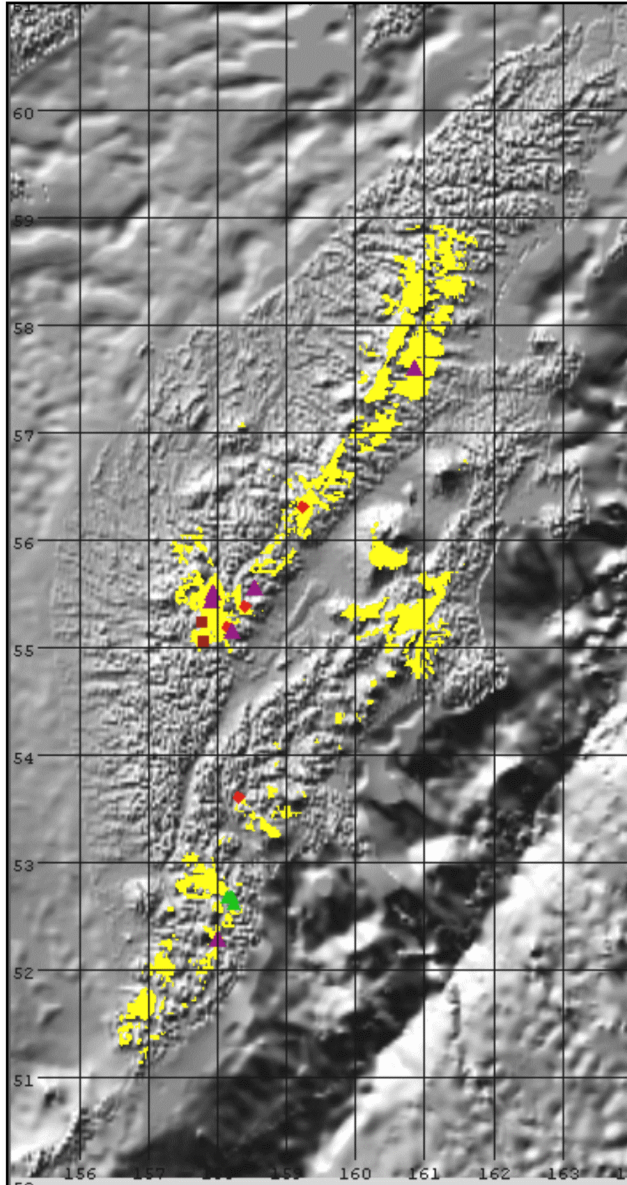
Figure 6. Prognostic regions of gold ore deposits.

amined region with a part of the data: Red squares are gold deposits, green circles – polymetallic gold-silver with lead and zinc prevailing over copper, blue circles – gold-silver, dark-red squares – gold-quartz, black triangles – manifestations, not divided into formations.

[22] The primary data includes examples of gold ore deposits but contain no information about the researches territories, where these deposits are absent. It impedes carrying out of a complex analysis using classical methods of image identification. An alternative decision implies the construction of decisive rule, designed as a cover of objects of a study sampling of one class. Given some assumptions the cover can be constructed as a function of certitude in the presence of the deposit [*Gitis and Yermakov,* 2004].

[23] The following items were selected for the problem solution:

1. Digital model of ground surface height in grid $2' \times 2'$;

2. Zones of faults, calculated as a summarized faults length in circle $R = 30$ km;

3. Volcanogenic zones, calculated as the proximity function to volcanic constructions of the Pliocene-Miocene time at $R = 100$ km;

4. Rocks of the Neogene period according to geological chart.

[24] It was assumed that an increase of values of each parameter at other equal conditions doesn't exclude a possibility of the presence of gold ore deposit. In this case the function of certitude in the presence of the deposit due to the selected prognostic parameters correlates with the function of empirical distribution. The inductive decision rule requires the study sampling, including all the present deposits.

[25] Let us denote the sampling of precedents

$$x(n) = \left(x^{(1)}, \ x^{(2)}, \ldots, \ x^{(I)} \right) .$$

Then the decision rule appears to be:

$$f(x) = \frac{1}{N} \sum_{n=1}^{N} v\left(x^{(n)} \right) ,$$

where

$$v\left(x^{(n)} \right) = \left\{ \begin{array}{ll} 1, \ x_i > x_i^{(n)} & \forall : i = 1, 2, \ldots, I \\ 0, \ x_i \leq x_i^{(n)} & \exists : i = 1, 2, \ldots, I. \end{array} \right.$$

[26] The result is shown in Figure 6.

[27] Let us explain the inductive conclusion using the terms of the subject-matter:

IF ground surface heights >500 m;

AND the distance to Pliocene volcanic constructions is less than 60 km;

AND the summarized length of faults in circle $R = 30$ km exceeds 50 km;

AND the volcanic rock of the Neogene period is present;

THAN gold, gold-silver, polymetallic gold- with lead and zinc prevailing over copper or gold-quartz deposits.

[28] Thus by this example the efficiency of the "Electronic Earth" system for solving concrete tasks in the field of Earth sciences is shown.

Conclusions

[29] The most important result of the "Electronic Earth" project is the fact that for the first time we could turn from GIS to a multi-user distributed geographical information analytical environment.

[30] The most significant results for users are the following:

(1) the integral information field was developed for carrying out research and solution of tasks;

(2) the technology of complex analysis was elaborated, available for an outsider in the field of IT.

[31] Significance of the project results is confirmed by the fact that remote access, search, exchange and integration of the interdisciplinary distributed resources of the Earth sciences and their complex analysis have become the basis of such projects and the Electronic Geophysical Year (eGY) and the global system of Earth monitoring (GEOSS) etc.

[32] To date the main directions of the future work on the project are: development of a universal storage of geodata, development of analytical methods of geodata processing in GRID environment and encouraging scientists to active application of the system "Electronic Earth".

References

Arskiy, Yu., V. Gitis, and A. Shogin (2008), Electronic Earth – GRID Network of Search, Integration and Analysis of Geodata, in *Smirnovsky Collection – 2007,* p. 117, PIK VINITI, Moscow.

Arskiy, Yu., V. Gitis, A. Shogin, and A. Weinstock (2007), Network geoinformation environment for analysis of spatial and spatio-temporal data, Abstracts, IUGG XXIV General Assembly, Perugia, Italy.

Arskiy, Yu., et al. (1999), *Rubricator of Information Editions of VINITI,* 31 pp., VINITI, Moscow.

Gitis, V., and B. Yermakov (2004), *Basics of Spatio-Temporal Prediction in Geoinformatics,* 256 pp., FIZMATLIT, Moscow.

Yu. M. Arskiy, A. N. Shogin, All-Russia Institute of Scientific and Technical Information, Russian Academy of Sciences, Moscow (alex@viniti.ru)

B. G. Gitis, A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow

A. V. Veselovsky, Institute of Geology of Ore Deposits, Petrography, Mineralogy and Geochemistry, Russian Academy of Sciences, Moscow